Fundamental Theories of Physics 176

Kurt Sundermeyer

Symmetries in Fundamental Physics

Springer

Fundamental Theories of Physics

Volume 176

For further volumes:
http://www.springer.com/series/6001

Kurt Sundermeyer

Symmetries in Fundamental Physics

Second Edition

Springer

Kurt Sundermeyer
Fachbereich Physik
Freie Universität Berlin
Berlin
Germany

ISBN 978-3-319-06580-9 ISBN 978-3-319-06581-6 (eBook)
DOI 10.1007/978-3-319-06581-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014939845

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

I dedicate this book to our grandson Jakob

Preface

Magische Symmetrie—Die Ästhetik in der modernen Physik

I like sailing. On many of my sailing trips I join my former colleague and friend Frieder[1], who is an electrical engineer by education. You can imagine that especially if the wind has slowed down we have plenty of time to chat about on-board short-circuits (which is Frieder's task) and the influence of relativity theories on our navigational GPS (which is supposed to be my task, being a theoretical physicist by education). Both non-working batteries and well-working GPS naturally lead to conversations about the laws of nature–how physicists formulate and understand them, and how engineers are utilizing them for our daily life. And I can tell you: these are worlds apart. When I tried to explain to Frieder why and how I'm giving lectures on "Symmetries in Physics" he was completely confused, because symmetry for him is not a category of pure and applied science, but–if at all–maybe for art. As a matter of fact, it is not easy to convey a feeling to a person outside of physics (and mathematics) how arguments of elegance, beauty, aesthetics, ... can be leitmotifs in such rational disciplines.

Let me give Frieder another chance to get my message through: Isn't it surprising that we can do physics at all (and engineering, in order to suit him)? Only because certain patterns in Nature occur here and there (and today and yesterday, etc.) in a comparable manner, can we dare to establish certain "laws of nature".[2] And strangely enough the laws of physics can be expressed in the language of mathematics. During the course of time, mankind discovered more and more phenomena of nature. But physics did not simply grow in the same way; now and then it was realized that certain phenomena, previously not known to be related, astoundingly do have common roots. As a matter of fact, physics today is simpler than the physics at the beginning of the last century. And one of the reasons for this are symmetries of Nature–or to be more precise: "symmetries of fundamental physics".

[1] Frieder and I worked for more or less twenty years in an industrial research institute on robotics and artificial intelligence.

[2] Here we have the first disparity between laws of nature and batteries: a battery may have worked yesterday, but it does not work today, and low and behold, if one just waits until tomorrow, it will work again.

After some struggling with myself I decided to use the wording “fundamental physics” in the title of this book. This is a risk in the sense that every scientist more or less claims to do fundamental research. And, as explained also in this book, the notion of “fundamental” received a change in attitude with the very idea of effective field theories. Apart from that, I had to indicate that this is not a book on symmetries in “physics”, because for instance symmetries in solid state and atomic physics are left out completely. Instead of “fundamental physics” I could have chosen the more apt wording “elementary particle physics and relativity theories”. Actually my lectures at the Freie Universität Berlin, from which this book originates, were once announced this way, but I thought it too clumsy to be used as the title of a book.

The theme “symmetries in elementary particle physics and relativity theories” is also dealt with in the nice book by A. Zee[3] [578], which very emphatically brings the symmetry arguments of today’s fundamental physics to the layman–in the sense that no mathematical formula is used. The present book is neither meant for layman nor for experts, but for readers with some background in theoretical physics. Nothing of the material is really new. Nearly everything can be found in textbooks or review articles. However, I know of no other attempt to treat fundamental physics entirely under the aspect of symmetries. You may take this monograph as a “mathematically enriched” version of Zee’s book. And although it is meant for physicists, it may also be of interest to philosophers and mathematicians (even though it may not completely satisfy their standards). Hopefully also Frieder can benefit from browsing through the text.

The very idea of this book is to explore established physics and introduce modern particle and relativity physics following the golden thread of symmetries. I asked myself how much of physics can be deduced from symmetries. To be sure, the symmetry arguments need to be fed and supported by experimental results. But it seems that the interaction between the theoretician and the experimentalist can be channelled by symmetries. And therefore in this book I start from very well established notions of analytical mechanics and follow the symmetry thread to highly topical issues such as, for example, the entropy of black holes.

Only at the beginning of the last century with the advent of the relativity theories and of quantum physics it was realized that symmetries and the related notion of invariance are a kind of principle of nature. For instance E. P. Wigner, one of the main symmetry players, coined the phrase “... if we knew all the laws of nature, or the ultimate Law of nature, the invariance properties of these laws would not furnish us new information.” Today two types of symmetry play a prominent role, namely space-time symmetries–as visible in the relativity theories and in relativistic field theories–and symmetries in abstract mathematical spaces, so called internal symmetries. Symmetries imply conservation laws, they allow us to classify particles and field variants, and they make sure that infinities otherwise arising in a quantum field theory cancel each other. In particle physics it became

[3] The exceedingly well-chosen German title is the motto at the beginning of this preface.

apparent that the basic interactions are based on local internal symmetries (or gauge symmetries). And in a sense to be described, also general relativity is a gauge theory. Since the symmetry transformations constitute a group in the mathematical sense, group theory plays a prominent rule in fundamental physics. Predominantly we are dealing with continuous groups, named after Sophus Lie, their algebras and–as a consequence of quantum-mechanical principles–the irreducible unitary representations of these groups. Some of the symmetries seem to be exact within the present experimental boundaries (like the symmetry with respect to Lorentz transformations), others are more or less broken. Examples are the spontaneous breaking of a symmetry in the electroweak theory, giving rise to the Higgs boson, or the violation of space reflection and of time reversal in weak interactions. Given the success of symmetry arguments for constructing the standard model in particle physics, it is no surprise that further symmetries (as for instance supersymmetries) are envisaged. Indeed, at present, the drawers of particle and gravity theorists can hardly hold the huge amount of speculations. All of these are symmetry inspired, but need either to be falsified or supported by experimental results such as are expected from the Large Hadron Collider, the most powerful microscope available today–and probably for years to come.

It might sound strange and it does not shed light on the imagination of an author, if he must concede in his preface that the headings of the chapters in his book do not cover what they pretend to contain. You may read each heading "XXX" as having the real meaning of "Symmetries and XXX". In order to make this clear, each chapter starts with a short preamble of what "XXX" adds[4] to our understanding of symmetries.

Instead of treating the topic directly and top-down from a modern perspective, I preferred to follow the historical course in unraveling symmetries in physics.

- Therefore I start with "Classical Mechanics" (Chap. 2), its conservation laws and the Galilei group as its symmetry group. Although hardly mentioned in textbooks and lectures, the so-called first Noether theorem is hidden behind the scene.
- Chapter 3 on "Electrodynamics and Special Relativity" is the motivation for discussing the Poincaré group and the two Noether theorems. This became a rather long chapter: Together with the Poincaré symmetry we can consider not only its limit to Galilei symmetry, but also understand the Poincaré symmetry as a limit of de Sitter symmetry, and as a specific restriction of conformal symmetries. The latter show up for instance as approximate symmetries in the high-energy limit of quantum chromodynamics, or as essential symmetries of string models. The pivotal work by E. Noether on variational symmetries remained outside the focus of physicists for more than half a century (and was re-derived by them several times). In the context of present-day theoretical

[4] … or added–in an historical point of view.

research it experienced an extension from finite and from closed algebras to infinite-dimensional and to open algebras.

- A book on symmetries in fundamental physics must reflect that fundamental physics is quantized (at least particle physics, but also gravity of black holes or near the big bang). The notion of symmetry in "Quantum Mechanics" (Chap. 4) gives rise to discuss both Wigner's theorem about symmetry generators and unitary (respectively anti-unitary) operators, and the ensuing important relation between symmetries and ray representations.
- In "Relativistic Field Theory" (Chap. 5) one requires for the defining properties of particles the irreducible unitary representations of the Poincaré group, and the representation of the Lorentz group for classifying fields. Notice that this chapter is headed "Relativistic Field Theory" and not "Quantum Field Theory". Certainly, today's theories of particle physics are quantum field theories. However, going into details here would take us far beyond the context of this book. Nevertheless, I felt obliged to have a subsection on effective theories including the idea of the renormalization group. For one thing, the notion of running couplings is needed to fully understand the standard model of particle physics and its possible extensions, and for another thing, going from one effective field theory to a neighboring one is related to the introduction or deletion of symmetry-related terms in the Lagrangians of the theories. There is a further section on spontaneous symmetry breaking, arising if the symmetry of the action is not a symmetry of the vacuum state. This phenomenon is essential in the formulation of the electroweak sector of the standard model. Another section deals with the discrete symmetries related to charge conjugation, space reflection and time reversal.
- After having established in Chap. 5 the symmetry (and the dimensional renormalization) principle for constructing actions for field theories the technical apparatus for understanding the gauge-symmetry-based aspects of "Particle Physics" (Chap. 6) is at our disposal. Chapter 6 also contains material on 'anomalies'. In a QFT-based presentation of symmetries, anomalies could be worth a separate chapter, since these represent situations where symmetries in the classical laws are spoiled or even completely destroyed by radiative quantum corrections.
- The next chapter (Chap. 7) deals with symmetries in "General Relativity and Gravitation". You doubtless know that general relativity (GR) deals with 'curved' spacetime, and therefore to understand gravitodynamics there is a section on Riemann geometry and more general geometries. The symmetry group of GR is related to general coordinate transformations which constitute an infinite-dimensional Lie group. It is an open issue in which sense general relativity (respectively possible extensions) is a gauge theory. There are common features but from its basics, GR is different from Yang-Mills type theories. This is specifically true considering the point of energy-momentum conservation. Theories of gravity also opened the understanding of boundary terms in action

functionals and their relation to topological invariants. Although without doubt GR is the best theory of gravity, symmetries allow many ways to extend Einstein's theory. Among these are alternative geometries (manifolds with torsion) and/or dynamics (actions quadratic in curvature and torsion).

- The chapter on "Unified Field Theories" (Chap. 8) extends the scope of currently observed gauge symmetries (grand unified theories), reflects about general relativity extended to higher dimensions (Kaluza–Klein models), to a symmetry comprising bosons and fermions (supersymmetry and supergravity). I also added some material on superstrings since these–although not amenable to the world as we know it–seem to be a testing ground for all kind of down-to-earth, exotic, and still unknown symmetries.
- In the "Conclusion" (Chap. 9) I present the symmetries of the 'World' action in a condensed form. Further I sketch how far symmetries serve for unifying physics, and I also reflect about some philosophically inspired issues around the topic of symmetries, like their possible origin, or their role in a basket of 'principles'.

As for the appendices: In order to fully understand symmetries some knowledge in "Group Theory" (Appendix A), especially on representation of groups and Lie algebras is indispensable. Anti-commuting entities enter physics in the context of "Spinors, $\mathbb{Z}_2$-gradings, and Supersymmetry" (Appendix B). Since all theories in fundamental physics are invariant with respect to local symmetry transformations, they necessarily are singular systems, and as such give rise to constraints, dealt with in Appendix C on "Symmetries and Constrained Dynamics". The theorems of E. Noether play a key role in the symmetries dealt with in this book. These assume a different form in classical mechanics, quantum mechanics, and in relativistic field theories. Since the latter are conceptually most often formulated in the language of path integrals, an appropriate Appendix D "Symmetries in Path Integral and BRST Quantization" is added. In a top-down approach to the theme of symmetry, the path integral and specifically the Faddeev-Popov recipe for quantizing Yang-Mills and generally covariant theories would probably constitute the first chapter; immediately followed by the BRST symmetries and there role in field quantization. The basic structures of fundamental physics (especially with regard to symmetries) can best be understood in the mathematical language of differential manifolds and bundles. It is in the language of modern "Differential Geometry" (Appendix E) where one can see for example the commonalities and the distinctions of Yang-Mills theories and general relativity. Especially in terms of differential forms (Appendix F "Symmetries in Terms of Differential Forms"), symmetries and their consequences can be described in a very elegant and compact way; as a matter of fact invariance with respect to general coordinate transformations is guaranteed by the calculus per se. In some contexts (e.g. for Poincaré gauge theories of gravity and in their generalization to supermanifolds) they are far superior to an index notation.

As may be already clear, this book is written by a physicist and it addresses itself to physicists in the first instance.[5] The text originates from lectures, and my audience was made up of advanced undergraduate students, graduate students, and curious colleagues from science departments. From the students I only required knowledge in analytical mechanics, Maxwell's form of electrodynamics, and quantum mechanics. But I expected no prior knowledge of the standard model or of general relativity. The lectures were not part of a curriculum with some final examinations on the topic, but were considered by the students as a "tip of cream" on top of their required lectures. I think, that the theme of the lecture serves to satisfy the curiosity which drives young people to study physics anyhow.

By far not all of the material in this book was covered in my 32 h of lectures. So I did not have the time to talk about symmetries in unified field theories and all those other topics that are marked with an asterisk in the Contents. I also did not cover completely what is here the Appendix A on group theory. However, since this material is absolutely essential, I gave it as a handout to the students. Together with this appendix and the part on spinors in Appendix B, the main text is completely self-contained. All other appendices are additional material going deeper either into field theory or into a modern differential geometry notation. For a first reading these appendices are not needed.

As to the literature I not only refer to articles and books which are on the same level as the lectures, but also more advanced material, including topical articles. And, given my relation to the Max Planck Institute for the History of Science, I aim to be as correct as possible in citing historical sources. Everybody has a chance to follow the discussion on symmetries, their scope and their applicability in more detail. Everyone is invited to start on his/her level of knowledge, pick up the golden symmetry thread, and hopefully will be led to areas of new knowledge.

[5] If you prefer a book on the topic written by a mathematician you may consult [467], a book (in German) which regrettably is out of print.

Acknowledgments

This book could only be written due to the influence of colleagues with whom I discussed symmetry-related topics all over my scientific life. I would like to specifically mention Donald C. Salisbury and Josep M. Pons who shared my particular interest on Noether theorems, constrained dynamics and phase-space symmetry generators. Several colleagues and friends read parts of earlier versions of the book and gave me a feedback. Particularly I thank Milutin Blagojević for helpful clarifications concerning Poincaré gauge theory approaches of general relativity, and Alexander Blum for critically reading and commenting the chapters on relativistic field theory and on particle physics. I appreciate that Friedrich Hehl pointed out to me some recent work concerning the formulation of energy-momentum and angular-momentum in field theories. I'm much obliged to William Brewer who scanned most of the text in order to polish the (American-)English. Thanks go to Claus Kiefer for encouraging and supporting this project. I also would like to express my gratitude for the hospitality at the Max Planck Institute for the History of Science, and specifically to Jürgen Renn for giving me the opportunity to use resources of the institute.

Notation and Conventions

X_D	D-dimensional differentiable manifold
U_D	Riemann Cartan space
V_D	Riemann space
M_D	Minkowski space with metric diag $(\eta_{\mu\nu}) = (+1, -1, \ldots, -1)$
E_D	Euclidian space
dS/AdS	de Sitter/Anti-de Sitter space

Indices

$i, j, k, \ldots$	indices in E_3 running over 1, 2, 3 or for finite dimensional systems enumerating the degrees of freedom
$\mu, \nu, \lambda, \ldots$	coordinate indices in X_4
$I, J, K, \ldots$	local Lorentz (tangent space) indices in X_4, used e.g. in tetrad formulations of gravitational theories
$A, B, \ldots$	coordinate indices in $X_D (D > 4)$
$a, b, c \ldots$	(gauge) group indices running from 1 to dim **G**
$\alpha, \beta, \ldots$	generic indices–sometimes standing for a collection of the previous indices
$r, s, \ldots$	indices numbering symmetry parameters

In most of the text the summation convention introduced by Einstein is used: If two indices in an expression are the same a summation is to be understood. I sometimes also follow the notation introduced by Bryce DeWitt [120] according to which there is to be understood an additional integration with respect to coordinates in case of functions, e.g.

$$U^{\alpha\beta}\phi_\beta = U^{\alpha\alpha'}\phi_{\alpha'} = \sum_\beta \int dx' U^{\alpha\beta}(x, x')\phi_\beta(x').$$

Fields

All fields considered are tensor or spinor fields in Minkowski spaces

Q or Φ generic field transforming in a specific way under the Lorentz group $\delta Q^\alpha = -\frac{i}{2}\epsilon^{\mu\nu}\left(\Sigma_{\mu\nu}\right)^\alpha_\beta Q^\beta$, where $\Sigma_{\mu\nu}$ is the spin matrix

φ scalar field

ξ_α, $\bar{\xi}_{\dot\alpha}$ Weyl fields; $\bar{\xi}_{\dot\alpha} = (\xi_\alpha)^*$

A_μ vector field in M_4

ψ, $\bar\psi$ Dirac or Majorana fields; $\bar\psi = \psi^\dagger\gamma^0$

Ψ_μ Rarita-Schwinger field

Spinorial Objects

$\varepsilon^{\alpha\beta}$ "metric" for Weyl spinors: $(\varepsilon^{\alpha\beta}) := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, as matrix identical to $(\varepsilon^{\dot\alpha\dot\beta})$; $\quad \xi^\alpha = \varepsilon^{\alpha\beta}\xi_\beta \quad \bar\xi^{\dot\alpha} = \varepsilon^{\dot\alpha\dot\beta}\bar\xi_{\dot\beta}$; $\xi\eta := \xi^\alpha\eta_\alpha$ and $\bar\xi\bar\eta := \bar\xi_{\dot\alpha}\bar\eta^{\dot\alpha}$, i.e. undotted/dotted indices are descending/ascending

γ^μ Dirac matrices with $\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$
$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$
$\gamma^{\mu\nu} = i/2\,[\gamma^\mu, \gamma^\nu] = i/2(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)$
$\not{v} := v_\mu\gamma^\mu$

σ^k Pauli matrices $\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ $\sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

Variational Objects

$S[\Phi]$ action functional

$\mathcal{L}(\Phi, \partial\Phi, \ldots)$ Lagrangian density depending on fields and their derivatives. In this text at most second derivatives are assumed to occur in $\mathcal{L}$ $S[\Phi] = \int_M d^Dx\mathcal{L}(\Phi, \partial\Phi, \ldots)$

$[\mathcal{L}]^\alpha$ Euler derivatives $[\mathcal{L}]^\alpha = \frac{\delta S}{\delta\Phi\alpha} = \frac{\partial\mathcal{L}}{\partial\Phi\alpha} - \partial_\mu\frac{\partial\mathcal{L}}{\partial\Phi_{\alpha,\mu}} + \partial_\mu\partial_\nu\frac{\partial\mathcal{L}}{\partial\Phi_{\alpha,\mu\nu}} + \ldots$

δF generic variation of a function

$\delta_S F$ infinitesimal symmetry variation, more specifically detailed to δ_ξ for general coordinate trafos, δ_ε for translations, δ_λ for Lorentz transformations, δ_Θ for (internal) gauge group transformations

$\bar\delta F(x)$ "active" change of the function $F : \delta F := F'(x) - F(x)$. The "passive" one is $\bar\delta F := F'(x') - F(x)$. For infinitesimal coordinate transformations δx^μ they are related by $\bar\delta F = \delta F - F_{,\mu}\,\delta x^\mu = -£_\xi F$

Gauge (Group) Objects

Given a gauge Lie group **G** resp. its Lie algebra $\mathfrak{g}$

X^a	generators of $\mathfrak{g}$ group elements (near the identity) are of the form $g(\xi) = \exp(i\xi^a X^a)$
f^{abc}	structure coefficients: $\left[X^a, X^b\right] = if^{abc}X^c$
T^a_ρ	ρ-representation (matrix) of $\mathfrak{g}$; specifically fundamental representation T^a_f, adjoined representation $\hat{T}^a$
A^a_μ	gauge potentials $A_\mu = A^a_\mu T_a$
$D_\mu(A)$	gauge covariant derivative

- acting on "matter" fields Φ as $D_\mu \Phi_\alpha = \left(\partial_\mu + igA^a_\mu T_a\right)_\alpha^\beta \Phi_\beta$
- acting on Lie-algebra valued objects G^a as $D_\mu G^a = (\partial_\mu + igA^b_\mu \hat{T}^b)^{ac} G^c$ where $\hat{T}$ is the adjoint representation of the algebra, and g has the meaning of a coupling constant

$F^a_{\mu\nu}$	field strength $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - gA^b_\mu A^c_\nu f_{bca}$

Geometric Objects

∂_μ, e_I	coordinate resp. frame base of the tangent space related by: $\partial_\mu = e^I_\mu e_I$, $e_I = e^\mu_I \partial_\mu$ where e^I_μ are the D-beins (specifically tetrads) and $e^{\ \mu}_I$ their inverses
dx^μ, ϑ^I	coordinate resp. frame base of the cotangent space related by $dx^\mu = e^I_\mu \vartheta_I$, $\vartheta_I = e^{\ \mu}_I dx^\mu$
g	metric tensor with components $g_{\mu\nu} = g\left(\partial_\mu \partial_\nu\right)$, $\eta_{IJ} = g(e_I , e_J)$
ω^{IJ}_μ	spin connection
$\nabla_\mu = D_\mu(\omega)$	covariant derivative with respect to ω : $\nabla_\mu u^I = \partial_\mu u^I + \omega^I_{J\mu} u^J$
$T^I_{\mu\nu}$	torsion of the connection ω: e.g. $T^I{}_{\mu\nu} = \nabla_{[\mu} e^I{}_{\nu]}$
$\tilde{R}^{IJ}_{\mu\nu}$	curvature of the connection ω: $$\tilde{R}^{IJ}_{\mu\nu} = (\partial_\mu \omega^{IJ}_\nu + \omega^{KJ}_\mu \omega^I_{k\nu}) - (\mu \leftrightarrow \nu)$$
$\Gamma^\lambda_{\mu\nu}$	linear (or affine) connection
$D_\mu = D_\mu(\Gamma)$	covariant derivative with respect to Γ: $D_\mu(\Gamma)u^\nu = \partial_\mu u^\nu + \Gamma^\lambda_{\mu\nu} u^\nu$

$T^{\lambda}_{\mu\nu}(\Gamma)$	torsion of the connection Γ : $T^{\lambda}_{\mu\nu} = \Gamma^{\ \lambda}_{\mu\cdot\nu} - \Gamma^{\ \lambda}_{\nu\cdot\mu}$
	contortion $K^{\lambda}_{\mu\nu} := \frac{1}{2}\left(T^{\lambda}_{.\mu\nu} - T^{\ \lambda}_{\mu.\nu} - T^{\ \lambda}_{\nu.\mu}\right)$
$\tilde{R}^{\rho}_{\sigma\mu\nu}(\Gamma)$	curvature of the connection Γ:
	$\tilde{R}^{\rho}_{\sigma\mu\nu}(\Gamma) := \partial_\mu \Gamma^{\rho}_{\nu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - (\mu \leftrightarrow \nu)$
$\left\{{}^{\mu}_{\lambda\nu}\right\}$	Christoffel symbol–or–V_4 connection:
	$\left\{{}^{\mu}_{\lambda\nu}\right\} = \frac{1}{2} g^{\rho\mu}\left(g_{\lambda\rho,\nu} + g_{\nu\rho,\lambda} - g_{\lambda\nu,\rho}\right)$
$F_{;\mu} = D_\mu(\{\})$	covariant derivative of F with respect to the Riemann connection
$R^{\mu}_{\nu\lambda\rho}(\{\})$	curvature in Riemann geometry V_4
	Ricci tensor $R_{\sigma\nu} := R^{\rho}_{\sigma\rho\nu}$

Further Symbols

$*$ and $\dagger$	complex conjugation (also denoted as c.c) resp. Hermitean conjugate for operators (also denoted as h.c.)
$\epsilon^{I_1 I_2 \ldots I_D}$	totally antisymmetric object in D dimensions $\epsilon^{012\ldots(D-1)} = 1$ $\epsilon_{012\ldots(D-1)} = -1, \epsilon_{I_2\ldots I_D} := \epsilon_{0I_2\ldots I_D}$
J^{μ}_{ν}	Jacobi matrix: $J^{\mu}_{\nu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}}$ for a coordinate transformation $x \longrightarrow x'(x)$. The inverse is denoted as $K^{\mu}_{\nu} = \frac{\partial x^{\mu}}{\partial x'^{\nu}}$
Λ^{μ}_{ν}	Lorentz transformations
ξ^{μ}, ϵ^{μ}, λ^{μ}_{ν}, θ_a	infinitesimal coordinate transformations, translations, Lorentz transformations and Lie group transformations
$F_{,\mu}(x)$	partial derivative $F_{,\mu} = \partial_\mu F = \frac{\partial F}{\partial x^{\mu}}$
$a := b$	a is defined by the expression b
$\equiv$	mathematical identity
$\sim$	approximate identity
$\doteq$	"on-shell equality": indicates that an equality holds on the solutions of the field-equations
$\simeq$	equivalence with respect to an equivalence relation
$\approx$	"weak equality": indicates that an equality holds on the surface of phase-space constraints; which surface is meant for the time being should become clear from the context
$\cong$	isomorphism of algebraic structures
$\ltimes$	semi-direct product, with the understanding that $\mathbf{G}\ltimes\mathbf{H}$ if $\mathbf{H}$ is a normal/invariant subgroup
$[F, G]$	commutator of F and G: $[F,\ G] = FG - GF$
$F_{(\alpha\beta)}$, $F_{[\alpha\beta]}$	symmetrization $F_{(\alpha\beta)} = \frac{1}{2}(F_{\alpha\beta} + F_{\alpha\beta})$ and anti-symmetrization $F_{[\alpha\beta]} = \frac{1}{2}(F_{\alpha\beta} - F_{\alpha\beta})$
$[F]$	mass-dimension of an object; can be derived for any field in D-dimensions from $[\mathcal{L}] = [\partial_\mu] = +1$
$\mathcal{D}\Phi$	measure in a functional integral

Constants

Mostly (unless an essential argument makes it necessary to keep them) units are chosen such $c = \hbar = 1$ where c is the velocity of light and $\hbar = \frac{1}{2\pi} h$ in terms of Planck's constant h. At other places Planck units are employed:

$L_{Pl} = \sqrt{\frac{\hbar G}{c^3}}$	Planck length: $L_{Pl} \approx 1.6 \times 10^{-33}\,\mathrm{cm}$
$E_{Pl} = \sqrt{\frac{\hbar c^5}{G}}$	Planck energy: $E_{Pl} \approx 1.2 \times 10^{19}\,\mathrm{GeV}$
$M_{Pl} = \sqrt{\frac{\hbar c}{G}}$	Planck mass: $M_{Pl} \approx 2.2 \times 10^{-5}\,\mathrm{g}$

in terms of Newton' constant $G \approx 6.7 \times 10^{-8}\,\mathrm{cm}^3 g^{-1} s^{-2}$ which in turn is related to Einstein's constant κ by $\kappa = 8\pi G/c^2$

Abbreviations and Acronyms

PB/DB	Poisson Bracket/Dirac Bracket
SRT	Special Relativity Theory
GR	General Relativity
QFT	Quantum Field Theory
ED	Electrodynamics
YM	Yang-Mills
GSW	Glashow-Salam-Weinberg (model of electroweak sector in the SM)
SM	Standard Model (for Elementary Particles)
GUT	Grand Unified Theories
SuSy/SuGra	SuperSymmetry/SuperGravity
KK	Kaluza-Klein
ADM	Arnowitt-Deser-Misner
EC, PGT	Einstein-Cartan (theory), Poincaré Gauge Theory
c.c. resp. h.c.	complex conjugate resp. Hermitean conjugate
b.t.	boundary term
FC, SC, PFC	first class, second class, primary first class (constraint)

Contents

Chapter 1
Introduction

Symmetry, as wide or as narrow you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty and perfection.

1.1 Symmetry: Argument, Principle, and Leitmotif

During the last century it was more and more recognized that fundamental physics can be based on symmetries and invariance principles. This is reflected by statements such as

- . . . if we knew all the laws of nature, or the ultimate Law of nature, the invariance properties of these laws would not furnish us new information. (E.Wigner [558])
- As far as I see, all a priori statements in physics have their origin in symmetry. (H. Weyl [552])
- The most important lesson that we have learned in this century is that the secret of nature is symmetry. (D. Gross [244])
- Today we realize that symmetry principles . . . dictate the form of the laws of nature. (D. Gross [247])
- Symmetry dictates interaction[1].
- Symmetry principles have moved to a new level of importance in this century and especially in the last few decades: there are symmetry principles that dictate the very existence of all the known forces of nature. (S. Weinberg[2].).
- To a remarkable degree, our present detailed theories of elementary particle interaction can be understood deductively, as consequence of symmetry principles . . .. (S. Weinberg [534])

Hermann Weyl in his reflections about symmetry [552].

---

[1] This is a slogan attributed to C.S.Yang, which I was not able to find. It is stated in [247]. Interestingly enough many people jumped on it, as one can track in looking for this slogan in an Internet search machine.

[2] This is cited at many places, for instance in [57] without mentioning the source.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_1,

- . . . profound guiding principles are statements of symmetry. (F. Wilzcek [560])
- "If you can identify Nature's complete symmetry group, you will know everything" is what became a pivotal dogma. (G.'t Hooft [504])

1.2 Operations and Invariants

What is meant by the aforementioned authors and their enthusiasm about the notion of symmetry in modern physics? At this place in the introduction it suffices to explain a symmetry as follows ([552]): Take an object and perform some operation on it. If after this operation the object looks the same–that is, it is invariant under this operation–the object is symmetric. This preliminary definition contains the essential ingredients of a symmetry, namely the *operation* on an object and its *invariance* under the operation. The definition leaves open the kind of object you are considering, but it fits well with a common sense understanding of a symmetry.

Take for instance as the object a square. This square looks the same–its form is an invariant–if you rotate it clockwise by 90°, or by 120°, 270°, or 360°. The same applies for counter-clockwise rotations. (Off course the square still looks the same if you leave it untouched.) And there is more you can do to the square and leave its form invariant, namely by using the operations of reflection.

This example already reveals a very important fact about symmetry operations: They form a *group* in the mathematical sense: (1) If two operations are performed one after another this yields again a symmetry operation. (2) There is a neutral operation of doing nothing. (3) Each operation has an inverse. (4) If three operations are considered, the operations are associative: performing the third operation after the combined action of the two others has the same effect as doing the first operation and then the operation combined from the second and the third. This observation gives a first hint, why group theory is so important in symmetry considerations, and thus for fundamental physics.

Symmetries are abundant in nature. Let me tentatively list several types of symmetries in physics:

- Geometric symmetries
  akin to the example of the square. Instead of the square, one can of course consider other geometric objects and classify their symmetries, as is done for instance in crystallography.
- Space-time symmetries
  relate to invariance properties of "laws of nature" (this notion being made more precise below) under space and time operations/transformations. They are identifiable already in classical mechanics. But more important to fundamental physics, the spacetime symmetries underlying the special theory of relativity together with quantum-mechanical postulates largely fix the form of the dynamical equations of the constituent fields (e.g. quarks and leptons) and the mediating fields (e.g. photon, gluons) in elementary particle physics.

- Gauge symmetries
  originate from the freedom of locally choosing phases in wave functions. These symmetries fix the interaction between constituent particles and/or those four fields that mediate the interactions (photon, W- and Z-bosons, gluons, graviton).
- Dynamical symmetries[3]
  are the symmetries prominently dealt with in quantum physics (and thus the most influential for atomic, nuclear and sub-nuclear physics) since they determine the spectral properties of these quantum systems. Already known by W. Pauli and other pioneers in quantum mechanics this became a research in its own right due to the work of A. O. Barut and A. Böhm [32].
- Permutation symmetries
  relate to the invariance properties of wave functions of many-particle systems with respect to the exchange of particles. These symmetries immediately lead to the two distinguished basic entities in fundamental physics, namely bosons and fermions.
- Duality symmetries
  mapping two theories or two different descriptions of a theory one-to-one to the other. The prime example is self-duality for the free Maxwell theory: you can mutually exchange the electric and the magnetic field, and this replacement leaves Maxwell's equations invariant. Another example is quantum physics either described by Schrödinger's wave equation or with Heisenberg's matrix mechanics. Today, dualities are experiencing a boom, triggered by so called string theory[4]. The five families of string models exhibit astoundingly many dualities, if one introduces branes; for a non-expert this is nicely described in [433]. Another very influential development was the discovery of a duality between anti-de Sitter gravity and a conformal field theory, sometimes named after its inventor as Maldacena conjecture[5].

Of course there is a plethora of further symmetries within specific physical systems. By this I have in mind the Chladni figures originating from the symmetry of a drum, say, or symmetries in a molecular arrangement and its influence on the pattern of energy levels, or symmetries of the universe as a whole, if it is assumed to be homogeneous and isotropic. These symmetries are indeed very helpful for easing calculations, but they are specific to the system in question.

The topics of this book are essentially the space-time and the gauge symmetries. These are characterized by operations on and invariants of "laws of nature", specifically in fundamental physics. Here I have used quotation marks because it has to be specified what is meant in this context by laws of nature and their symmetry operations.

---

[3] The term 'dynamic' is by some used in the sense of what is called in this book 'Lie symmetries', by others in the sense of 'local symmetries'.

[4] My "so called" refers to "theory", since–although undoubtably highly developed–the string picture is still on the level of a possible model of nature.

[5] I do not expect the reader to understand this jargon here in the introduction, but I promise that the catchwords 'de Sitter' and 'conformal' will be explained later. Some more words will be spent on this "gravity/gauge theory conjecture" in Sect. 8.4.

1.3 "Symmetries" in "Fundamental Physics"

In this section I will analyze each of the words in the title of this book–except for the word "in". That is, I will lay down, what is meant by "fundamental physics", and especially which aspects of "physics", and to which of these aspects the "symmetry" considerations apply.

1.3.1 What is Meant by "Fundamental Physics"?

Of course many physicist claim to work on fundamental topics. But, I suppose that thinking of the basics, foundations and deeper roots in physics none of them has three[6] Ohm's laws in mind, namely (1) $I = U/R$, (2) $R = U/I, \ldots$ No doubt, Ohm's law is a law of nature, but not a fundamental one (it even has a restricted range of applicability), since it can be deduced from the laws of electrodynamics and from material properties, the latter in turn being derivable from the molecular and atomic structure of the material governed by quantum mechanical laws.

In considering fundamental physics in these lectures, I have in mind the two frame theories or pillars of modern physics, namely special relativity and quantum theory, and the models for the four basic interactions of nature (gravitational, electromagnetic, weak and strong interactions). Special Relativity is a frame theory since every other theory in physics (in principle) has to obey its postulates–according to the current experimental results concerning the Poincaré invariance of our world. And since today we know that the world has a quantum nature (classical physics only being an approximation), again in principle every theory in physics must obey its postulates. As elucidated in later chapters, both special relativity and quantum theory have very drastic consequences in terms of symmetries. In particular, their postulates and symmetry arguments lead inevitably to the existence of mass and spin of (elementary) particles, as well as to the existence of anti-particles.

The postulates of special relativity and of quantum theory are obeyed in quantum field theories, quantum electrodynamics (QED) being the prime example. Today, QED is part of the so-called standard model of particle physics, as it was established both theoretically and experimentally in the 1970's. In retrospect, one realizes that this became possible by using symmetry arguments. In a terminology to be explained later, the electromagnetic, the weak and the strong interactions are each described by a Yang-Mills theory with a specific gauge symmetry group.

The other well known basic interaction, namely gravity, has today only a classical field theoretic expression, which is general relativity. General relativity also has its very specific symmetry group, loosely speaking the group of all non-singular coordinate transformations.

To conclude, "Fundamental Physics" may be taken as a synonym for "Elementary Particle Theory and Relativity Theories". I think that here I am in good company

[6] No joke: It the army I was told that there are three laws.

with D. Gross and his short but excellent paper entitled "The role of symmetry in fundamental physics" [247]. And at the same time it should be clear, what this book is NOT about, namely symmetries in nuclear, atomic, molecular and solid state physics.

1.3.2 "Physics" on Which Level of Description?

Symmetries may be investigated for various different physical entities, and here I list a few relating to specific configurations/states, to generic solutions of dynamical equations, to the dynamical equations themselves, to the Lagrangian or Hamiltonian description, and to the action. I will exemplify the different levels by the example of the Chladni[7] figures. These refer to patterns exhibiting the various modes of vibration on a mechanical surface. In high school, the type of vibrations are exemplified by patterns shown by sand on drumheads, and which are remarkably symmetric: The sand aggregates at those places ('nodes') where no vibrations actually occur. The "physics" is controlled by the equation

$$\frac{\partial^2 u}{\partial t^2} = v^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) \tag{1.1}$$

where the vibrational amplitude in the z- direction $u(x, y, t)$ depends on the membrane (x, y)-coordinates and on time, and where v carries the meaning of the speed of sound on the membrane. This equation can–if at all–only be solved analytically by fixing boundary conditions. These are related of course to the shape of the membrane (e.g. as a square or a circular plate) and how it is fixed. Let's take the case of a circular plate that is fixed on its boundary R, so that $u(R, t) = 0$. In this case a set of solutions of (1.1) is given in terms of polar coordinates (r, θ) as

$$u_{nm}(r, \theta, t) = (A \cos v\lambda_{nm} t + B \sin v\lambda_{nm} t) J_m(\lambda_{nm} r)(C \cos m\theta + D \sin m\theta) \tag{1.2}$$

where $m = 0, 1, \ldots; n = 1, \ldots,\ \lambda_{nm} = \alpha_{nm}/R$ and α_{nm} is the n'th positive root of the Bessel function J_m[8].

- Configuration/State
  A state of the membrane is a specific pattern, given by a specific choice of (n, m) together with fixed constants in the solution. Only the states u_{0m} exhibit a radial symmetry, while those for $n \neq 0$ do not. This can easily be seen by exhibiting the pattern which sand forms on the line of nodes $u(t, x, y(x)) = 0 = \dot{u}(t, x, y(x))$. Generically a configuration in classical mechanics is described by a solution of the

[7] named after Ernst Florens Friedrich Chladni (German 1756–1827), who in the encyclopedias is called either a German, a Hungarian, or a Slovak. From this confusion you may realize how unstable–and perhaps unimportant–the status of nationality was at his times.

[8] It is not in the scope of this book to derive these solutions. You either believe me or cheque it yourself…

equations of motion together with initial conditions. In electrodynamics a configuration is a solution of the field equations together with some boundary conditions. In quantum physics one refers to definite states in a Hilbert space, realized by e.g. a solution of the Schrödinger equation. In any case one has solutions of dynamical equations with some additional conditions leading to a unique configuration.

- Solutions of the dynamical equations
  One may define a symmetry as an invertible map which sends solutions of the equations of motion or the field equations into other solutions. Generically one can write the dynamic equations as

$$\mathfrak{O}\Lambda(x) = 0, \tag{1.3}$$

  where the functions $\Lambda(x)$ denote e.g. a column with a finite or infinite number of components or a matrix characterizing the classical or quantum state, and $\mathfrak{O}$ denotes a (differential) operator. The expression (1.3) could for example be the set of Maxwell's equations, a Schrödinger equation etc. In case of the vibrating membrane (1.3) obviously corresponds to

$$\left(\frac{\partial^2}{\partial t^2} - v^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right) u(t, x, y) = 0.$$

  If there are operators $\mathcal{S}^a$ with the property

$$\mathfrak{O}\left(\mathcal{S}^a \Lambda(x)\right) = 0$$

  solutions are transformed into solutions, and $\mathcal{S}^a$ may be called symmetry operators.
- Dynamical equations
  Instead of investigating their solutions one could look at the dynamical equations directly. Classically these are the equations of motion or the field equations; in quantum mechanics we deal for instance with the Schrödinger equation or with expressions for transition matrices. For the example of vibrating membranes, this amounts to investigate symmetries of (1.1). These are invariant with respect to arbitrary rotations, reflections, and translations of the x-y-plane. We saw already that the solutions of the dynamical equations in general do not exhibit this full symmetry.
  Indeed mathematicians have investigated extensively symmetries of differential equations in general [399] and of dynamical equations of fundamental physics (Klein-Gordon, Dirac, Maxwell, Einstein equations) [207]. Here symmetry is understood in the broadest sense, including non-linear and non-local transformations in the dependent and independent variables. Although mathematicians sometimes use other names, in this book I will term all those transformations which map solutions to solutions as *Lie symmetries*, in order to distinguish these from those transformations that leave a Lagrange function invariant (or quasi-invariant, that is invariant up to a divergence.). The latter are called *variational symmetries*; in some situations also Noether-Bessel-Hagen symmetries

- Lagrange Function/Hamilton Operator
  You know from analytic mechanics that the dynamical equations are very often encoded in a Lagrange function, in the sense that they are obtained from demanding the vanishing of the variational–or–Euler derivative of the Lagrangian. For instance, if the Lagrange function $L(q^k, \dot{q}^k)$ depends on some coordinates and their first derivatives with respect to time, the Euler-Lagrange equations are

$$[L]_k \ := \frac{\partial L}{\partial q^k} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^k} \ = 0. \tag{1.4}$$

  In the example of the Chladni figures the Euler-Lagrange equations derived from

$$L = \frac{1}{2}\left(\left(\frac{\partial u}{\partial t}\right)^2 - v^2\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right)$$

  are identical to (1.1).
  As it turned out, the field equations of all four basic interactions can be stated in terms of Lagrangians. These will be derived largely by symmetry arguments in Chap. 6 and Chap. 7. A Lagrange function yields not only the classical field equations but–in case of renormalizable quantum theories–also computational methods for calculating transition amplitudes in the form of Feynman graphs. In the canonical formulation of quantum physics, it is not the Lagrangian but the Hamiltonian operator which entails the dynamical evolution. A first step to construct this operator is to find the classical Hamiltonian. As known from analytical mechanics one is led from a Lagrangian (formulated in configuration-velocity space) to a Hamiltonian (formulated in phase space) by a Legendre transformation. Typically, in case of continuous symmetries this transformation has its subtleties: The phase space becomes restricted because of the existence of constraints (i.e. relations among the coordinates and the momenta). Appendix C is completely devoted to this aspect of symmetries.
- Action S
  Behind the procedure of how to derive from a Lagrangian the dynamical equations lurks of course another principle of nature (allow me to formulate it this way, I will try to give it a better flair in Chap. 9.): The principle of least action. Since field theories are treated in extenso in later chapters it suffices to present the arguments here for the action of classical mechanics

$$S\{q(t_1, t_2)\} = \int_{t_1}^{t_2} L(q, \dot{q}, t)dt.$$

  As a matter of fact the action is a fundamental entity both classically in that–provided certain boundary conditions are satisfied–

$$\delta S = 0 \qquad \Longleftrightarrow \qquad \text{Euler}-\text{Lagrange equations},$$

and quantum mechanically in that the transition amplitude $K(a,b)$ for a particle moving from a point a to a point b is proportional to

$$K(a,b) \sim \sum_{path} \exp\left\{\frac{i}{\hbar} S_{\text{path}}\right\}$$

where the sum runs over all paths connecting a and b. The fundamental role of the action is also visible in representing it as a total differential

$$dS = p^k dq^k - Hdt,$$

from which the concepts of momentum and energy are derived as[9]

$$\frac{\partial S}{\partial q^k} = p^k \qquad\qquad \frac{\partial S}{\partial t} = -H$$

and thus–for reasons of consistency–the equations of motion with the concept of force:

$$\dot{p}^k = \frac{\partial^2 S}{\partial q^k \partial t} = -\frac{\partial H}{\partial q^k}.$$

As shown later, the invariance of an action with respect to an infinitesimal transformation yields a conservation law (or, strictly speaking, a conserved current). But it is not sufficient for the existence of a conservation law that the equations of motion be invariant. In classical mechanics, for instance, the equations of motion with a frictional force are invariant under time displacement but the Lagrangian is not, and no conserved energy exists. Hence the invariance of the action integral is to be considered more basic. The invariance of the equations of motion are an immediate consequence of the invariance of the action. Since the interactions in fundamental physics can be described by Lagrangians, the symmetry considerations in this book will turn to be most fruitful if investigated on the level of the action. And indeed aside from the discrete C, P, T-symmetries everything in this book is an epos about variational symmetries.

- Path integral
  Classical physics is only an approximation to quantum physics. Quantum physics and classical physics are for instance related in that the Green's functions for the fundamental interactions can be derived from the so-called generating functional

$$Z[J] = \int \mathcal{D}\Phi \, \exp\frac{i}{\hbar}\left\{S[\Phi] + \int d^4x J\Phi\right\}.$$

[9] Strictly speaking the following applies to Hamilton's function $\bar{S}$ which relates to the action as $S = \bar{S}(t_2)$-$\bar{S}(t_1)$; see Sect. 2.1.4

Here Φ stand for a collection of fields, $S[\Phi]$ is the classical action, and $\mathcal{D}\Phi$ denotes the measure in the functional integral. For $Z[J]$ to be invariant it is not sufficient that the action is invariant, but that also the measure (and also the boundaries of integration) must be invariant under symmetry transformations. Although for the plethora of physically interesting cases the measure turns out to be invariant, there are exceptions. For historical reasons these are termed "anomalies". Generically an "anomaly" designates a situation in which the symmetry of a classical action is not a symmetry of the path integral.

1.3.3 Which Kind of "Symmetry"?

As stated before, symmetries can be defined by operations on and invariants of objects. Now I will be more precise and specify which objects, invariants and operations are (mainly) dealt with in this book.

Objects

The question "which objects?" was contemplated in the previous subsection, where it was argued that the action or–given the remark before–the path integral seems to be the most pivotal (and compact) description of a theory. In this book we are mostly interested in the invariance of an action under a symmetry transformation. The action is a functional of entities depending on coordinates, that is dependent and independent variables. More generically, according to [221] it is indispensable to distinguish externally specified background structures Σ and dynamical structures Φ. For example

- Classical mechanics and electrodynamics

$$\Phi = \{q^k(t)\} \qquad\qquad \Sigma = \{t, \phi(t, \mathbf{x})\}$$

  where t is the time variable and ϕ is a given external field; like for instance $\phi = \{\mathbf{E}, \mathbf{B}\}$ with an electric and a magnetic field. However, you may also decide to incorporate the dynamical equations for the electric and magnetic fields and thus transport these fields from the background structure to the dynamical structure:

$$\Phi = \{q^k(t), \mathbf{E}(\mathbf{x}, t)), \mathbf{B}(\mathbf{x}, t)\} \qquad\qquad \Sigma = \{t, \mathbf{x})\}.$$

- Relativistic field theory

$$\Phi = \{\varphi^\alpha(x^\mu)\} \qquad\qquad \Sigma = \{x^\mu, g_{\mu\nu}\}$$

  where the fields φ^α depend on the spacetime co-ordinates x^μ, and $g_{\mu\nu}$ is the metric tensor of the spacetime in which the fields φ^α live. This makes sense for a flat or otherwise fixed background. If the metric itself is a dynamical entity we have the situation

$$\Phi = \{\varphi^\alpha(x), g_{\mu\nu}(x)\} \qquad \Sigma = \{x^\mu\}$$

typical for generally relativistic theories.

The point to be made here, is that it depends on the context which objects one includes within the symmetry considerations. This needs to be dealt with in distinguishing covariance and invariance.

Operations

The most general operations on the background and the dynamical structures are

$$\Sigma \Rightarrow \widehat{\Sigma}(\Sigma, \Phi, D^{(\sigma)}\Phi) \qquad \Phi \Rightarrow \widehat{\Phi}(\Sigma, \Phi, D^{(\sigma)}\Phi),$$

where $D^{(\sigma)}\Phi$ stands for derivatives of the dynamical structures Φ. In our symmetry considerations for fundamental physics we are mainly interested in those transformations that include the identity mapping, and which can be investigated near the identity: $\widehat{\Sigma} = \Sigma + \delta\Sigma$, $\widehat{\Phi} = \Phi + \delta\Phi$. The most widely considered symmetry transformations are so called infinitesimal point transformations:

$$\widehat{\Sigma} = \Sigma + \delta\Sigma(\Sigma, \Phi) \qquad \widehat{\Phi} = \Phi + \delta\Phi(\Sigma, \Phi).$$

Transformations which contain derivatives $D^{(\sigma)}\Phi$ seem not to be of relevance for fundamental physics (although, as will be discussed in Sect. 2.2.4, they arise with the Kepler problem). But they were investigated by mathematicians and go under a variety of names like generalized symmetries or Lie-Bäcklund transformations. Specific cases are contact transformations and dynamical transformations. The Lie-Bäcklund and generalized transformations are of relevance in the context of symmetry invariances of differential equations. They allow one to derive from a solution further solutions by symmetry arguments ([284], [399], [485]). As argued above our scope is not the symmetries of the equations of motion, but the invariance of actions. Actions that are invariant under any infinitesimal point or generalized transformation are termed to exhibit *variational symmetries*. If the symmetry transformations constitute a Lie group also the term *Noether transformations* will be used. Even if the transformation group behind a variational symmetry is not a Lie group, these transformations are physically relevant, since for instance it sometimes suffices that they form a group on-shell, that is on the solutions of the dynamical equations.

Invariants

In this section the example concerns point transformations in classical mechanics

$$\hat{t} = t + \delta t(t), \qquad \hat{q}^k(\hat{t}) = q^k(t) + \delta q^k(q^i, t). \tag{1.5}$$

Let us investigate first what it means to have an invariant action, and what consequences there are to the Lagrangian, the field equations and their solutions.

- The action transforms as

$$S\{q\} = \int_{\hat{t}(t_1)}^{\hat{t}(t_2)} d\hat{t} \, \left(\frac{dt}{d\hat{t}}\right) L\{q(\hat{q}, \hat{t})\} := \hat{S}\{\hat{q}\}.$$

  We have an invariance with respect to the transformation (1.5) iff $S\{\hat{q}\} \equiv S\{q\}$, signaling a variational symmetry.
- The Lagrange function transforms as

$$\hat{L}\{\hat{q}\} := \left(\frac{dt}{d\hat{t}}\right) \, L\{q(\hat{q}, \hat{t})\}.$$

  In case of a variational symmetry this enforces

$$\left(\frac{d\hat{t}}{dt}\right) L\{\hat{q}\} \stackrel{!}{=} L\{q\} + \frac{d}{dt}\Omega(q, t),$$

  where Ω is an arbitrary function.
- Because the Euler-Lagrange equations transform as

$$[\hat{L}\{\hat{q}\}]_k \equiv \left(\frac{dt}{d\hat{t}}\right) \frac{\partial q^j}{\partial \hat{q}^k} [L\{q\}]_j$$

  it is always true that

$$[L\{q\}]_j = 0 \Leftrightarrow [\widehat{L}\{\hat{q}\}]_k = 0$$

  but the equations are not form-invariant (or "covariant") in general[10]. However, as we will see, the dynamical equations are covariant if the transformation belongs to a Noether symmetry.

1.4 The Scope of Symmetries

1.4.1 Ontology of Symmetries

One may distinguish different kinds of symmetries such as

---

[10] The definitions of invariance and covariance and their mutual relation are not really obvious, and depend on the dynamical and the background structure. More about this in Sect. 7.5.3.

- Discrete vs. continuous
  The rotation of a square by multiples of $90°$ is an example of a discrete symmetry, whereas the rotation of a circle by an angle φ leads to a continuous symmetry, since the angle can take on arbitrary values.
  In this book, the only discrete symmetries considered are time inversion T, space reflection P, and charge conjugation C. The main continuous symmetries are those with respect to translations and rotations in 3D, the Poincaré transformations in 4D, the $\mathbf{SU}(\mathbf{N})$-groups in particle physics, and those related to general coordinate transformations.
- Space-time vs. internal
  (sometimes also "external" vs. "internal"). The space-time symmetries refer to symmetries with respect to coordinate transformations, as is the case in the relativity theories. The internal symmetries refer to degrees of freedom in the gauge groups of the standard model of particle physics.
  This distinction becomes obsolete in light of Kaluza-Klein type theories, where space-time symmetries in higher dimensions can mimic internal symmetries, and in light of advanced notions like fibre spaces, where space-time symmetries and internal symmetries are described by comparable notions.
- Global vs. local
  Both properties refer to continuous groups (e.g. Lie groups) only. The elements of a Lie group depend on a finite number of parameters. (For instance the group elements describing rotations in 3D depend on two angles.) The symmetry is called global if the group parameters are constant, and it is called local, if the parameters are allowed to depend on the location in space and time. The theories of the four fundamental interactions are characterized by local symmetries.
- Bosonic vs. fermionic
  This is the topic of supersymmetry, where actions are formulated that are invariant with respect to transformations depending on even and odd parameters in a Grassmann algebra and representing bosonic and fermionic elements.
- Exact vs. non-exact
  If the symmetries in nature were not as exact as our observation of them, we would not be able to observe them. This may sound like a tautology. But perhaps one would find that none of the symmetries is really exact if our observational/experimental capabilities were sufficiently improved. Not even the Lorentz symmetry on which–as will become manifest in the next sections–a large part of our current understanding of particles and fields relies, is sacred. Thus one should not be surprised to see articles in which possible violations of Lorentz symmetry are discussed; see also Sect. 3.6.
  The current picture of cosmology is depicted by some as a story of the successive breaking of symmetries from a large symmetry to the symmetries of the four interactions as they are manifest today.
  Non-exactness of a symmetry can have different qualities
  1. Approximate symmetry
     Take for example the almost exact symmetry of exchanging a proton and a neutron. They are of course not the same, because of their different charges and

their tiny difference in mass[11]. However from nuclear physics one knows for instance that nuclei with the same number of nucleons behave in the same way. This led W. Heisenberg to the concept of isotopic spin. The same idea came up again in the early quark model, where the symmetry between the up- and the down-quark, and their difference in mass, would be responsible for the mass difference between a proton and a neutron. In the standard model of particle physics this concept survived in form of the weak isospin; see Sect. 6.3.

2. Explicitly broken symmetries
   Neither parity nor charge conjugation are exact symmetries in certain processes of weak interactions.
3. Spontaneously broken symmetries
   In contrast to the previous case, the theory itself can provide a mechanism by which a symmetry can be broken. The prime example is the Higgs mechanism (see Sect. 5.4), playing an eminent role in the breaking of the electroweak interaction into its weak and its electromagnetic components. As will be shown, the relevant action is invariant under a transformation, but the ground state (the vacuum) is not.

- Active vs. passive symmetry transformations
  An "active" symmetry transformation relates different configurations or states of the physical system. A "passive" symmetry transformation describes the same configuration/state in terms of different references (frames, calibrations, gauges). A related meaning is as follows: Instead of performing an operation on objects and investigating the invariants, one may as well leave the objects untouched and change the "position" of the observer.
- Executable vs. non-executable transformations
  A system can for instance be shifted in space. This is an example of an executable symmetry transformation. However, it is not possible to construct a physical mirror image of the system.

1.4.2 Symmetry Groups in Fundamental Physics

Group theory is the appropriate mathematics for investigating symmetries. Amazingly only few groups seem to be important in physics. These groups are dealt with later in this book either in the corresponding chapters or in the appendix on group theory. Here the most common groups are characterized in a listing following the sequence of chapters in this book. Don't worry about the overabundance of technical terms in this subsection. They will all be defined and explained later in the book.

- Classical Mechanics
  The symmetry group of classical mechanics is the Galilei group with spatial translations and rotations, time translations and Galilei boosts. The groups which properly characterize the canonical structure of the phase space are the symplectic groups $\mathbf{Sp}(\mathbf{2N})$.

[11] The quantitative relation is $m_n = 1.00135 m_p$.

- Electrodynamics and Special Relativity
  Here is the origin of the Poincaré group–also called the inhomogeneous Lorentz group. This group is essential for formulating (relativistic) field theories–such as the standard model of particle physics. The Galilei group is an appropriate limiting case (in group theory jargon: group contraction) for non-relativistic mechanics. There are other group contractions of the Poincaré algebra, and the Poincaré algebra itself is the contraction of still another algebra, namely the de Sitter algebra. Electrodynamics has an internal $\mathbf{U(1)}$ gauge symmetry. This internal Abelian symmetry is generalized in particle physics to non-Abelian Yang-Mills theories. Electrodynamics gives also rise to the conformal group, a group always coming into play if massless particles exist within a theory.
- Quantum Mechanics and Relativistic Field Theory
  The Hilbert space foundation of quantum mechanics reveals that it is not the Poincaré group–or as made more precise later: its universal covering group–itself, but its irreducible unitary representations which matters. These yield both the particle properties and a distinction of fields (with different integer and half-integer spin) and their dynamical equations. As a matter of fact, symmetry arguments largely dictate[12] the types of fields and particles. Additionally one needs to investigate discrete groups because (1) the operations charge conjugation C, space reflection P, and time reversal T are part of the full Lorentz group, and because (2) operations related to exchange symmetry give rise to the distinction of bosons and fermions.
- Particle Physics
  After a long interplay between experimental and theoretical physicists from the 30's to the 70's of the last century the standard model of particle physics was established, and since then successfully tested with an astounding precision. The standard model is a Yang-Mills gauge theory with the gauge group $\mathbf{SU_C(3)} \times \mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ (with an appropriate assignment of color C, weak isospin I, and weak hypercharge Y).
- Gravitation and/or General Relativity
  The symmetry of GR is encoded by the group of diffeomorphisms in four spacetime dimensions, also known as general coordinate transformations. Attempts have been made to formulate gravitational theories as gauge theories of the Poincaré group, necessarily leading to manifolds with and without curvature and torsion.
- Unified Field Theories
  Grand unified theories aim to establish Yang-Mills theories with a gauge group $\mathbf{GUT} \supset \mathbf{SU(3)} \times \mathbf{SU(2)} \times \mathbf{U(1)}$. Favorites for $\mathbf{GUT}$ are $\mathbf{SU(5)}$ and $\mathbf{SO(10)}$, although the former is currently ruled out because the proton is more stable than predicted by an $\mathbf{SU(5)}$ Yang-Mills theory.

[12] Observe that the word "dictates" shows up in many of the quotes of Sect. 1.1.

Kaluza-Klein theories start from a Hilbert-Einstein Lagrangian in $(4+N)$-dimensions. They break up into 4D diffeomorphisms and a rank-N gauge group symmetry, the resulting group depending on how these N dimensions are compactified.
Supersymmetric theories (establishing a symmetry between bosons and fermions) formulated in terms of superalgebras which are graded extensions of the Poincaré algebra. Supergravity theories can be formulated as gauge theories of supergroups.

1.4.3 The Use of Symmetries

Symmetries are not only aesthetically pleasing, but also rather practical and serviceable:

- Conservation
  By Noether's first theorem, every global continuous symmetry entails a conserved quantity and a Lie algebra of these quantities (s. Chap. 3). The most well-known ones are derived from the symmetries due to isotropy and homogeneity of space-time: energy, momentum, and angular momentum conservation.
- Classification
  As obvious from crystallography, symmetries can serve for classification. But to emphasize it again, this is not a topic treated in this book. Examples more appropriate in our context (and dealt with in Chap. 5) are the Wigner characterizations of particles (in terms of masses and spin) due to the postulates of special relativity and quantum physics, the distinction of bosons and fermions due to the postulates of quantum physics, and the distinction of scalar, spinor, vector ... fields due to the representation theory of the Lorentz group.
- Unification
  The unifying symmetry of Maxwell's theory of electromagnetism and relativistic mechanics is given by the Poincaré group. Further examples are the electroweak unification with $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ (Chap. 6) and the attempts at a GUT (Grand Unified Theory).
- Prediction
  During the past century on various occasions symmetry and associated group theory considerations led to the prediction of particles which later were found. Thus for instance, although not immediately interpreted that way, Dirac's relativistic equation for the electron pointed to the necessary existence of the positron. Another example is the original (up,down,strange)-quark model (the "Eightfold Way") for which Gell-Mann, led by properties of the product of $\mathbf{SU(3)}$ representations predicted the mass, the strangeness, and the charge of the Ω-particle. An asymmetry in the participation of the quarks and the leptons in certain processes of weak interactions gave rise to the prediction of the c quark, later experimentally established as the charmonium resonance, a bound state of the c with its antiparticle $\overline{c}$. Weinberg and Salam predicted the existence and masses of the electroweak gauge

bosons, the $W^{\pm}$ and the Z, found in the 70's. Another remarkable case is the prediction of still a further family of quarks (the top and bottom quarks discovered later) in order to conceive a mechanism for CP violation in weak interactions.

- Guidance
  After the importance and role of symmetries were gradually understood, starting with the work of Einstein at the beginning of the last century, symmetries became a means of guiding research in fundamental physics. A prime example is Einstein's search for a theory of gravitation, and his observation that essentially because of the equivalence of inertial and gravitational mass this theory must be invariant with respect to general coordinate transformations (Chap. 7). Another extremely successful route was found for particle theory when it was realized that the electromagnetic, weak, and strong interaction can be formulated as Yang-Mills theories based on specific gauge (symmetry) groups. Because of these successes theoreticians are convinced that any further development in the fundamental theories will be based on symmetries (Chap. 8). These developments are to be seen for instance in GUTs, where one is looking at a genuine symmetry group, which is broken to the symmetry groups of the standard model. Another route is supersymmetry, where the distinction between bosons and fermions ceases away. Theories in the sense of Kaluza and Klein have the attractive feature that the external and internal symmetries are nothing but coordinate symmetries in a higher-dimensional space-time.
- Simplification
  To an outsider, physics seems to be rather complex. However, both classical and modern physics can be simplified by the analysis of their symmetries. As will be understandable at the end of the book, the "world" action can already be written astoundingly compact–essentially due to symmetries. In this sense, fundamental physics is looking simpler today than fifty or one hundred years ago. However there are still different sectors in the action which are structurally different.
  Some researchers hope to find an expression where the whole "world" can be described very elegantly on a tiny piece of paper. In the mid-1980's theoreticians got excited about the possibility of formulating fundamental physics in terms of strings; but this hope more or less diminished in a cloud of technical details and too many "world"-options, called multiverses.
- Comprehension
  We do not know why, but it seems to be the fact that physics has some few basic principles. Symmetry is one of those, together with the principle of least action[13]. Both of them lead us to the hope to discovering why doing physics is possible at all. These philosophically-rooted questions will be dealt with in Chap. 9.

[13] . . . and maybe a quest of renormalizability.

1.5 Bibliographical Notes

For those who are somewhat impatient and want to have a quick qualitative overview, I recommend the introductory chapter in [57][14], the article by D. Gross in Physics Today [246] or his colloquium paper [247]. A very enthusiastic presentation meant for high school teachers is [336]. Last but not least I recommend again the enjoyable book by A. Zee [578].

[14] Also to be found in the Stanford Encyclopedia of Philosophy; http://plato.stanford.edu/entries/symmetry-breaking/.

Chapter 2
Classical Mechanics

Une intelligence qui, à un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la compose embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome; rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux.

This chapter is meant to provoke curiosity on the topic of symmetries merely by reflecting on conservation laws in good old classical mechanics. Just by asking the deeper question why these laws hold, we arrive at a first understanding of how properties of space and time relate to invariances. And invariances in turn entail symmetry groups, which in case of classical mechanics is the Galilei group. Thus the more appropriate caption of this chapter could be "Galilei Group".

All of us started to learn the concepts of physics along the notions of classical mechanics. It's qualitative formulation began with Galileo Galilei (1564–1642), found its powerful formulation by Isaac Newton (1643–1727) a century later, and got mathematically refined as analytical mechanics by Jean L. Lagrange (1736–1813), William R. Hamilton (1805–1865) and Gustav J. Jacobi (1804–1851) in the 19th century. It received a further mathematical refinement in the 20th century; see e.g. [14]. Classical mechanics was considered as the model of science up to the end of the 19th century. Today we believe that classical mechanics is a threefold limiting case of our world: It is the limit of low velocities–or–a world with an infinite velocity of light, a world with vanishing Planck constant, and a world with weak gravitational fields–meaning that the Schwarzschild radius of an object is much larger than the typical length scale of the object. Nevertheless, this book, although aimed at fundamental physics, starts with a chapter on classical mechanics. It may come as a surprise that classical mechanics is conceptually more complicated than modern quantum field theory–the reason being found in symmetries.

Pierre-Simon Laplace, Essai philosophique sur les probabilités, Paris, Christian Bourgeois, 1986, pages 32 et 33.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_2,

2.1 Newtonian and Analytical Mechanics

The basic laws of mechanics can be expressed either in terms of differential equations (e.g. Newton's, Lagrange's, Hamilton's) or as integral variational principles (e.g. those of Maupertuis or Hamilton). Let us recapitulate the formulation by differential equations first.

- Newton
  The foundation of qualitative and quantitative mechanics in terms of mass points, velocities, accelerations, forces, ... and their mutual dependencies is due to I.Newton and is laid out in the three so-called "Newton's laws". From today's point of view one recognizes a circular definition in these laws: There are (inertial) systems in which Newton's laws are valid.
- Lagrange
  Although from the historical perspective the roles of Lagrange, Hamilton and others are more subtle, every student in physics associates with Lagrange the concept of generalized coordinates, the Lagrange function and the Euler-Lagrange equations derived from the Lagrange function. The Lagrange function may be taken as the starting point to understand and to investigate symmetries in terms of operations and invariants.
- Hamilton
  Notions such as phase space, canonical transformations, Poisson brackets, ... are associated with the name of Hamilton. The phase space approach on the one hand side reveals the "furniture" of classical mechanics and on the other hand paves a canonical way towards quantization.
- Hamilton-Jacobi
  This is a formulation of classical mechanics in terms of action-angle variables, a formulation that directly uncovers conserved quantities. The Hamilton-Jacobi form of physics is most appropriate for describing chaotic systems and for geometrical optics. It also may serve as a bridge to the Schrödinger equation of quantum mechanics.

2.1.1 Newtonian Mechanics

Newton bequeathed us with three laws [385]

N1: Every body continues in its state of rest, or of uniform motion in a right line unless it is compelled to change that state of forces impressed upon it.

N2: The change of motion is proportional to the motive force impressed, and is made in the direction of the right line in which that force is impressed.

N3: To every action there is always opposed an equal reaction.

Newton points out that his laws do hold in an ever-existing absolute space and in an external absolute time flow[1]. Those frames in which N1 holds are called inertial

[1] A comment about this assumption is made in the concluding remarks of this chapter.

frames. These are at rest in absolute space or move uniformly along straight lines and are interrelated by the transformations

$$x^i \to x'^i = x^i + v^i t + a^i \qquad t \to t' = t + \tau \qquad \text{with constants} \qquad v^i, a^i, \tau, \quad (2.1)$$

also called Galilei translations.

In our modern notation: For a system with K mass points m_α $(\alpha = 1, .., K)$ at (Cartesian) co-ordinates $\vec{x}_\alpha(t)$ and with velocities $d\vec{x}_\alpha/dt =: \dot{\vec{x}}_\alpha$ one defines

$$\begin{array}{ll}
\text{momenta :} & \vec{p}_\alpha = m_\alpha \dot{\vec{x}}_\alpha \\
\text{kinetic energy:} & T = \sum_\alpha \frac{m_\alpha}{2} \dot{\vec{x}}_\alpha^2 \\
\text{total momentum:} & \vec{P} = \sum_\alpha \vec{p}_\alpha \\
\text{total angular momentum:} & \vec{J} = \sum_\alpha \vec{x}_\alpha \times \vec{p}_\alpha \\
\text{center of mass:} & \vec{R} = \frac{1}{M} \sum_\alpha m_\alpha \vec{x}_\alpha \qquad \text{with} \qquad M := \sum_\alpha m_\alpha .
\end{array}$$

The force $\vec{F}_\alpha(\vec{x}, \dot{\vec{x}}, t)$ exerted on the mass point m_α is the sum of an external force and internal forces,

$$\vec{F}_\alpha = F_\alpha^{(ext)} + \sum_{\beta \neq \alpha} \vec{F}_{\alpha\beta}$$

where according to Newton's third law $\vec{F}_{\alpha\beta} = -\vec{F}_{\beta\alpha}$. The equations of motion according to Newton's second law are

$$\vec{N}_\alpha := m_\alpha \ddot{\vec{x}}_\alpha - \vec{F}_\alpha \equiv \dot{\vec{p}}_\alpha + \vec{\nabla}_\alpha V = 0, \qquad (2.2)$$

where the last identity holds for conservative forces, namely those derivable from a potential function as $\vec{F}_\alpha = -\vec{\nabla}_\alpha V$. In this case (which will be assumed in the sequel) the total energy of the system is defined as $E = T + V$. The Newtonian equations of motion are 3N differential equations of second order for the positions $\vec{x}_\alpha$ as functions of time. It is remarkable that not only the equations of motion of classical mechanics, but all dynamical equations of fundamental physics are of second order. As a matter of fact, higher order differential equations tend to have instabilities as shown by a theorem by Ostrogradski[2].

[2] I suppose that aside from a historian of science, no one will really read the original article in Mem. Ac. St. Petersbourgh VI, 4, 385 (1850); the year is not a misprint!

2.1.2 Lagrange Form of Mechanics

The crucial step from Newtonian to analytical mechanics is the recognition that the equations of motion can be derived from a Lagrange function

$$L(\vec{x}_\alpha, \dot{\vec{x}}_\alpha, t) \;=\; T \;-\; V(\vec{x}_\alpha, t)$$

with the kinetic energy T and potential energy V. The Lagrange function can also be written in "generalized" coordinates q^k (e.g. curvilinear coordinates) and velocities $\dot{q}^k$:

$$L(q, \dot{q}, t) \;=\; \sum_{i,k}^{N} a_{ik}(q,t)\,\dot{q}^i \dot{q}^k \;-\; V(q,t), \tag{2.3}$$

where N is the number of degrees of freedom. One easily convinces oneself that the set of Euler-Lagrange equations

$$[L]_k \;:=\; \frac{\partial L}{\partial q^k} - \frac{d}{dt}\,\frac{\partial L}{\partial \dot{q}^k} \;=0 \tag{2.4}$$

is equivalent to the equations of motion in the Newtonian form (2.2). The Lagrangian is not unique, since adding to the Lagrangian a total derivative $\frac{d}{dt}B(q,t)$ does not change the equations of motion.

It is comprehensible that because of the freedom in choosing the generalized coordinates, the Euler-Lagrange equations remain "structurally" invariant after an (at least locally) invertible coordinate transformation $q \to \hat{q}(q)$. Explicitly

$$\dot{\hat{q}}^k = \frac{\partial \hat{q}^k}{\partial q^l}\,\dot{q}^l \qquad \text{from which} \qquad \frac{\partial \dot{\hat{q}}^k}{\partial \dot{q}^j} = \frac{\partial \hat{q}^k}{\partial q^j} \qquad \text{and} \qquad \frac{d}{dt}\frac{\partial \hat{q}^k}{\partial q^l} = \frac{\partial \dot{\hat{q}}^k}{\partial q^l},$$

and similar expressions, where the hatted and the un-hatted variables are exchanged. Now, with $\hat{L}(\hat{q}, \dot{\hat{q}}, t) := L(q, q, t)$ we derive

$$\begin{aligned}
\frac{d}{dt}\frac{\partial \hat{L}}{\partial \dot{\hat{q}}^k} &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^l}\,\frac{\partial \dot{q}^l}{\partial \dot{\hat{q}}^k}\right) = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^l}\,\frac{\partial q^l}{\partial \hat{q}^k}\right) = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^l}\right)\frac{\partial q^l}{\partial \hat{q}^k} + \frac{\partial L}{\partial \dot{q}^l}\,\frac{d}{dt}\,\frac{\partial q^l}{\partial \hat{q}^k} \\
&= \left(\frac{\partial L}{\partial q^l} - [L]_l\right)\frac{\partial q^l}{\partial \hat{q}^k} + \frac{\partial L}{\partial \dot{q}^l}\,\frac{\partial \dot{q}^l}{\partial \hat{q}^k} = \frac{\partial \hat{L}}{\partial \hat{q}^k} - [L]_l\,\frac{\partial q^l}{\partial \hat{q}^k}.
\end{aligned}$$

This can be rewritten as

$$[\hat{L}]_k \;=\; [L]_l\,\frac{\partial q^l}{\partial \hat{q}^k}, \tag{2.5}$$

and reveals what is meant by structural invariance of the Euler-Lagrange equations under coordinate transformations[3]: Together with the equations $[L]_l = 0$ also

[3] Equation (2.5) expresses that the Euler-Lagrange derivative $[L]_k$ transforms like a covariant vector with respect to the coordinates q^i.

$[\hat{L}]_k = 0$. Structural invariance is different from form invariance (or covariance) with respect to particular transformations. The covariance of the equations of motion under a coordinate transformation is the defining property of a Lie symmetry.

It is found to be insightful to write the Euler-Lagrange equations explicitly as

$$[L]_k = \frac{\partial L}{\partial q^k} - \frac{\partial^2 L}{\partial \dot{q}^k \partial q^j}\dot{q}^j - \frac{\partial^2 L}{\partial \dot{q}^k \partial \dot{q}^j}\ddot{q}^j := V_k - W_{kj}\ddot{q}^j = 0. \tag{2.6}$$

From this we observe that if the *Hessian*

$$W_{kj} := \frac{\partial^2 L}{\partial \dot{q}^k \partial \dot{q}^j} \tag{2.7}$$

has an inverse $\bar{W}^{jl}$, the N Euler-Lagrange equations can be expressed as

$$\ddot{q}^j = \bar{W}^{ji} V_i = F^j(q, \dot{q}, t).$$

This "normal form" allows us to apply theorems about the existence and uniqueness of solutions of ordinary differential equations. Lagrangians for which $\det W \neq 0$ are called regular, and singular otherwise. Also, a system described by a regular (singular) Lagrangian is called regular (singular). This is justified since regularity of a Lagrangian is a coordinate-independent statement for invertible coordinate transformations $q \to \hat{q}$: The determinant of $\hat{W}$ is the product $\det W (\det J)^2$ with the Jacobian J of the coordinate transformation. Also the addition of a total derivative to a Lagrangian does not change its character of being regular or singular.

The Hessian also plays a role in the inverse problem of the calculus of variations which asks for conditions under which a set of second order equations $\ddot{q}^j = F^j(q, \dot{q}, t)$ can be derived from a Lagrange function. These conditions–in the literature known as the Helmholtz conditions in view of a publication of Hermann von Helmholtz from 1895–require the existence of a nonsingular matrix (w_{ij}) that obeys a set of differential equations. As you can guess, if a Lagrangian exists, then taking for (w_{ij}) the Hessian, the Helmholtz conditions are fulfilled. On the other hand, if for given F^j a matrix (w_{ij}) obeying the Helmholtz conditions can be found, the Lagrange function can be constructed explicitly and uniquely–up to an overall multiplicative constant and up to a total derivative. The solution of the inverse problem is far from trivial: Only in 1941, J. Douglas succeeded in solving completely the two-dimensional case.

Since this book is mainly about variational symmetries, we always assume the existence of a Lagrangian–and "luckily"–or by "basic principles"–for all fundamental interactions Lagrangians are known. But this remark already leads outside classical physics. There are even surprising arguments that quantization requires Lagrangians [279].

2.1.3 Hamiltonian Formulation

Mechanics in the Hamiltonian form is couched in terms of generalized positions and momenta instead of generalized positions and velocities, as is the case for the Lagrangian formulation. The generalized momenta are defined by

$$p_k := \frac{\partial L}{\partial \dot{q}^k}. \tag{2.8}$$

This expresses the momenta p in terms of $(q, \dot{q}, t)$. These functions can be inverted for the velocities, i.e. $\dot{q}^i = \dot{q}^i(q, p, t)$ iff the matrix

$$\frac{\partial p_k}{\partial \dot{q}^i} = \frac{\partial^2 L}{\partial \dot{q}^k \partial \dot{q}^i}$$

has non-zero determinant. Again we see the appearance of the Hessian W_{ij}. Only for regular systems the subsequent derivation of their Hamiltonian formulation is sound. But as we will see later, fundamental physics is inherently singular due to its symmetries; there the transition to a Hamiltonian is a little tricky–to say the least; see Appendix C.

In the regular case the Hamilton function is defined as

$$H(q, p, t) := p_k \dot{q}^k(p, q) - L(q, \dot{q}(p, q, t), t). \tag{2.9}$$

The transformation from $L(q, \dot{q}, t)$ to $H(q, p, t)$ is a *Legendre transformation*. For the Lagrange function of the form (2.3) the numerical value of the Hamilton function H is calculated from

$$H = p_k \dot{q}^k - L = \frac{\partial L}{\partial \dot{q}^k} \dot{q}^k - (a_{ik} \dot{q}^i \dot{q}^k - V) = 2a_{ik} \dot{q}^i \dot{q}^k - a_{ik} \dot{q}^i \dot{q}^k + V = T + V = E$$

as the total energy of the system.

The equations of motion (2.4) can–in the non-singular case–uniquely be expressed by the Hamilton function. To get these Hamiltonian equations we derive for (2.9)

$$dH = \frac{\partial H}{\partial q^k} dq^k + \frac{\partial H}{\partial p_k} dp_k + \frac{\partial H}{\partial t} dt = (p_k d\dot{q}^k + \dot{q}^k dp_k) - \frac{\partial L}{\partial q^k} dq^k - \frac{\partial L}{\partial \dot{q}^k} d\dot{q}^k - \frac{\partial L}{\partial t} dt.$$

Since with the definition of the generalized momentum (2.8) the Euler-Lagrange equations are equivalent to

$$\dot{p}_k = \frac{\partial L}{\partial q^k}$$

the two terms in the previous expression proportional to $d\dot{q}^k$ cancel, so that

$$\delta H = (\dot{q}^k \delta p_k - \dot{p}_k \delta q^k) - \frac{\partial L}{\partial t}\delta t.$$

Thus the Hamilton equations of motion become

$$\dot{q}^k = \frac{\partial H}{\partial p_k} \qquad \dot{p}_k = -\frac{\partial H}{\partial q^k}. \tag{2.10}$$

The $2N$-dimensional space spanned by the q^k and p_k is called the "phase space". In taking the time as an additional variable one arrives at the $(2N+1)$-dimensional "state space". The motion can be pictured as that of a $2N$-dimensional "phase fluid": Each stream-line of the moving fluid represents the motion of the system starting from specific initial conditions. The fluid as a whole represents the complete solution.

A central concept of the phase space formulation of classical mechanics is the Poisson bracket, defined for two phase functions A, B by

$$\{A, B\} := \frac{\partial A}{\partial q^k}\frac{\partial B}{\partial p_k} - \frac{\partial B}{\partial q^k}\frac{\partial A}{\partial p_k}. \tag{2.11}$$

The Poisson brackets form an algebra in the mathematical sense (see Appendix A.1.1) in identifying the algebra operation $\Box$ with the bracket operation {, }. This Poisson bracket algebra obeys additionally the properties of a Lie algebra, and especially the Jacobi identity

$$\{A, \{B, C\} + \{B, \{A, C\} + \{C, \{A, B\} = 0. \tag{2.12}$$

The Poisson brackets of the coordinates and their canonically conjugate momenta become the "fundamental brackets"

$$\{q^k, q^l\} = 0, \quad \{p_k, p_l\} = 0, \quad \{q^k, p_l\} = \delta^k_l. \tag{2.13}$$

The Hamiltonian equations (2.10) can be written in terms of Poisson brackets as

$$\dot{q}^k = \{q^k, H\} = \frac{\partial H}{\partial p^k} \qquad \dot{p}_k = \{p_k, H\} = -\frac{\partial H}{\partial q^k} \tag{2.14}$$

by which the time evolution of a state space function $F(q, p, t)$ can be expressed as

$$\dot{F} := \frac{dF(q, p, t)}{dt} = \{F, H\} + \frac{\partial F}{\partial t}. \tag{2.15}$$

The Poisson bracket (2.11) is not only a nice means to write down the Hamilton equations in a compact form, but it constitutes–so to say–the backbone of the phase-space formulation of classical mechanics. This will be expounded in the sequel.

Canonical Transformations

We saw that point transformations $q \to \hat{q}(q,t)$ leave the Euler-Lagrange equations invariant. They also leave the Hamiltonian equations (structurally) invariant. There is, however, a larger class of invariance transformations, especially canonical transformations[4]. These are defined as those invertible transformations

$$\hat{q}^j = \hat{q}^j(q,p), \qquad \hat{p}_j = \hat{p}_j(q,p)$$

which leave the fundamental brackets (2.13) invariant. Before deriving properties of canonical transformations, let us introduce another more condensed notation. Since with the canonical transformations the concept of distinct coordinates and momenta fades into oblivion, it makes sense to collect all $2N$ phase space variables into one set $(x^\alpha) = (q^1, ..., q^N, p_1, ..., p_N)$. In this notation the fundamental brackets (2.13) can be written as

$$\{x^\alpha, x^\beta\} = \Gamma^{\alpha\beta} \qquad \text{with} \qquad \Gamma := \begin{pmatrix} 0_N & 1_N \\ -1_N & 0_N \end{pmatrix}. \tag{2.16}$$

In terms of the matrix Γ, the Poisson bracket for two phase space functions A and B becomes

$$\{A, B\} = \Gamma^{\alpha\beta} \frac{\partial A}{\partial x^\alpha} \frac{\partial B}{\partial x^\beta}.$$

The condition for $\hat{x}(x)$ being a canonical transformation simply is $\{\hat{x}^\alpha, \hat{x}^\beta\} = \Gamma^{\alpha\beta}$. Denoting by $X^{\alpha\beta} := \frac{\partial \hat{x}^\alpha}{\partial x^\beta}$, a canonical transformation obeys

$$X\,\Gamma\,X^T = \Gamma. \tag{2.17}$$

This allows to derive that canonical transformations form a group: Take as group elements two matrices X_1 and X_2. Then

$$(X_1X_2)\Gamma(X_1X_2)^T = X_1X_2\Gamma X_2^T X_1^T = X_1\Gamma X_1^T = \Gamma,$$

and the existence of an identity element and an inverse is obvious. The matrices X_i are a representation of the symplectic group $\mathbf{Sp(2N)}$.

Let us now derive under which conditions a transformation is a canonical transformation. Since in most of this text we are interested in continuous transformations

[4] Some textbooks define canonical transformations by the property of leaving the Hamiltonian equations invariant, but this is not true in general, see ([14]). The canonical transformations can also be defined as contact transformations with respect to the Lagrangian.

we restrict our considerations to infinitesimal transformations $\hat{x}^\alpha = x^\alpha + \delta_c x^\alpha$. Then–to first order–the defining relation (2.17) amounts to

$$\frac{\partial(\delta_c x^\alpha)}{\partial x^{\alpha'}}\Gamma^{\alpha'\beta} + \frac{\partial(\delta_c x^\beta)}{\partial x^{\alpha'}}\Gamma^{\alpha\alpha'} \stackrel{!}{=} 0,$$

which after multiplication and summation by $\Gamma^{\beta\gamma}\Gamma^{\alpha\gamma'}$ (and some renaming of indices) becomes

$$\frac{\partial}{\partial x^\alpha}\left(\Gamma^{\beta\gamma}\,\delta_c x^\gamma\right) - \frac{\partial}{\partial x^\beta}\left(\Gamma^{\alpha\gamma}\,\delta_c x^\gamma\right) \stackrel{!}{=} 0.$$

Hence $\Gamma^{\beta\gamma}\,\delta_c x^\gamma$ is the gradient of an (infinitesimal function) g, and

$$\hat{x}^\alpha = x^\alpha + \Gamma^{\alpha\beta}\frac{\partial g}{\partial x^\beta} = x^\alpha + \{x^\alpha, g\} =: x^\alpha + \delta_g x^\alpha. \tag{2.18}$$

This is the most general infinitesimal canonical transformation. The phase space function g is called the generator for infinitesimal canonical transformations. The commutator of two infinitesimal transformations generated by g and h is

$$[\delta_g, \delta_h]x^\alpha = (\delta_g\delta_h - \delta_h\delta_g)x^\alpha = \{\delta_h x^\alpha, g\} - \{\delta_g x^\alpha, h\} = \{\{x^\alpha, h\}, g\} - \{\{x^\alpha, g\}, h\},$$

which due to the Jacobi identity (2.12) can be written as

$$[\delta_g, \delta_h]x^\alpha = \{x^\alpha, \{h, g\}\} = \delta_{\{h,g\}}x^\alpha. \tag{2.19}$$

Finally, for an arbitrary phase space function F

$$\delta_g F(x) = F(x + \delta_g x) - F(x) = \frac{\partial F}{\partial x^\alpha}\,\delta_g x^\alpha = \{F, g\}. \tag{2.20}$$

2.1.4 Principle of Stationary Action

The action S is defined as the functional

$$S\{q\} = S[q^k(t_1, t_2)] := \int_{t_1}^{t_2} L(q, \dot{q}, t)dt,$$

that is a mapping of functions $q^k(t)$ to real numbers. Based on previous work of P.L. de Maupertuis, L. Euler and L. Lagrange in the 18th century, W. Hamilton in 1832 formulated the principle of stationary action: "The classical path q^k_{class} between t_1 and t_2 is the one for which S is stationary". This indeed seems to be a primary principle of physics. It can be traced back to quantum mechanics: According to the formulation with Feynman path integrals (see Appendix D.1), a particle virtually takes all possible paths between the endpoints. The ones near the classical paths dominate the exponential in the transition amplitude

$$\langle q_b t_b | q_a t_a \rangle = \sum_{path} \exp \frac{i}{\hbar} S_{path}.$$

Nevertheless, it remains still a mystery why Nature knows about an action, to begin with.

Euler-Lagrange Equations

Consider different paths $q + \delta q$ and require $\delta S \stackrel{!}{=} 0$. The variation δS is

$$\delta S = \int_{t_1}^{t_2} \{L(q + \delta q, \dot{q} + \delta \dot{q}, t) - L(q, \dot{q}, t)\} dt.$$

For infinitesimal δq^k

$$L(q + \delta q, \dot{q} + \delta \dot{q}, t) = L(q, \dot{q}, t) + \left(\frac{\partial L}{\partial q^k} \delta q^k + \frac{\partial L}{\partial \dot{q}^k} \delta \dot{q}^k \right) + \cdots$$

The requirement $\delta S \stackrel{!}{=} 0$ is thus equivalent to the requirement

$$\int_{t_1}^{t_2} dt \left(\frac{\partial L}{\partial q^k} \delta q^k + \frac{\partial L}{\partial \dot{q}^k} \delta \dot{q}^k \right) \stackrel{!}{=} 0$$

$$\overbrace{\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^k} \delta q^k \right) - \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^k} \right) \delta q^k}$$

$$\begin{aligned} \delta S &= \int_{t_1}^{t_2} dt \left[\left(\frac{\partial L}{\partial q^k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^k} \right) \delta q^k + \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^k} \delta q^k \right) \right] \\ &= \int_{t_1}^{t_2} [L]_k \delta q^k + \left(\frac{\partial L}{\partial \dot{q}^k} \delta q^k \right) \Bigg|_{t_1}^{t_2} \stackrel{!}{=} 0. \end{aligned} \tag{2.21}$$

This must be valid for arbitrary variations δq^k. Assuming now that the variation at the initial and final points of time vanish $(\delta q^k(t_1) = 0 = \delta q^k(t_2))$ the second term in this identity vanishes, and the remaining equations $[L]_k = 0$ are nothing but the Euler-Lagrange equations (2.4). For later purposes we might characterize the dynamical equations also by requiring

$$\delta L = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^k} \delta q^k \right).$$

This derivation of the equations of motion should not be completely new to you, since it is taught in the basic course of theoretical classical mechanics. But mostly it is mentioned only in passing that this holds true only by assuming appropriate boundary conditions. As you will see later, boundary terms contain additional information of the physical system. They are integral part of Noether currents and play an essential role in theories which are invariant with respect to general coordinate transformations, Einstein's general relativity being a prime example.

Higher Derivatives

We were assuming that the Lagrangian of the system depends on the coordinates and on at most their first time derivatives. It is obvious how the previous results can be extended to situations in which the Lagrangian depends on higher derivatives. For example in case of $L(q, \dot{q}, \ddot{q})$ one derives straightforwardly

$$\delta L = [L]_k \delta q^k + \frac{d}{dt} B \tag{2.22}$$

where now

$$[L]_k \;:=\; \frac{\partial L}{\partial q^k} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^k} \;+\; \frac{d^2}{dt^2}\frac{\partial L}{\partial \ddot{q}^k} \tag{2.23}$$

and the boundary term is

$$B = \left(\frac{\partial L}{\partial \dot{q}^k} - \frac{d}{dt}\frac{\partial L}{\partial \ddot{q}^k} \right) \delta q^k + \frac{\partial L}{\partial \ddot{q}^k} \delta \dot{q}^k .$$

From (2.22) the equations of motion $[L]_k = 0$ only follow if both δq^k and $\delta \dot{q}^k$ vanish at t_1 and t_2. Even if we assume that these stringent conditions can be fulfilled, we are still faced with dynamical equations of fourth order due to the last term in (2.23). Nevertheless, there are exceptions. Take for instance the Lagrange function $L(q, \dot{q}, \ddot{q})$ as derived from a function $\bar{L}(q, \dot{q})$ in the form

$$L(q, \dot{q}, \ddot{q}) = \bar{L}(q, \dot{q}) - \frac{d}{dt}\left(q \frac{\partial \bar{L}}{\partial \dot{q}} \right) . \tag{2.24}$$

Instead of plugging this into the previous expressions we repeat the variation (indices suppressed):

$$\begin{aligned}
\delta S &= \int_{t_1}^{t_2} dt \left[\frac{\partial \bar{L}}{\partial q} \delta q + \frac{\partial \bar{L}}{\partial \dot{q}} \delta \dot{q} \right] - \delta \left(q \frac{\partial \bar{L}}{\partial \dot{q}} \right) \Bigg|_{t_1}^{t_2} \\
&= \int_{t_1}^{t_2} dt \left(\frac{\partial \bar{L}}{\partial q} - \frac{d}{dt}\frac{\partial \bar{L}}{\partial \dot{q}} \right) \delta q + \left(\frac{\partial \bar{L}}{\partial \dot{q}} \delta q \right) \Bigg|_{t_1}^{t_2} - \delta \left(q \frac{\partial \bar{L}}{\partial \dot{q}} \right) \Bigg|_{t_1}^{t_2} \\
&= \int_{t_1}^{t_2} dt \, [\bar{L}]_k \delta q^k - q^k \delta p_k \Big|_{t_1}^{t_2} \overset{!}{=} 0 .
\end{aligned}$$

Thus indeed the Euler derivatives $[\bar{L}]_k$ are of second order. The equations of motion follow if the variation of the momenta p_k vanish at the boundary (rather then the δq^k). The specific Lagrangian (2.24) is not just a curiosity; instead, it displays the typical structure of the original Hilbert action of general relativity and its relation to the Einstein action; see (7.70).

Moreover we learn that the addition of a total derivative to the action–although not changing the equations of motion–may require different boundary data in order that the variational problem is well-posed. The generic relation between boundary data and second order terms can be found in the following way [375]: Assume that a second order Lagrangian can be written as

$$L_C(q, \dot{q}, \ddot{q}) = L_q(q, \dot{q}) - \frac{df(q, \dot{q})}{dt}.$$

Here L_q is a first order Lagrangian for which q has to be kept fixed at the boundaries, and L_C a Lagrangian for which a given function $C(q, \dot{q})$ is kept fixed. L_C and L_q bring about the same (second-order) equations of motion if

$$f(q, C) = \int p(q, C)\, dq + F(C),$$

where $F(C)$ is an arbitrary function. Here, it is assumed that one can solve $C(q, \dot{q})$ as $\dot{q} = \dot{q}(q, C)$, and express the momenta as $p = p(q, C)$. The previous case (2.24) immediately follows with $C = p$.

Hamilton Equations

The Hamilton equations can also be derived by variational methods if the action is expressed in terms of phase space coordinates. Rewrite

$$S = \int L(q, \dot{q}, t)dt = \int (p_k \dot{q}^k - H)dt = \int (p_k dq^k - H dt).$$

Now

$$\begin{aligned} \delta S &= \int \left\{ dq^k \delta p_k + p_k \delta dq^k - \frac{\partial H}{\partial q^k} \delta q^k dt - \frac{\partial H}{\partial p_k} \delta p_k dt \right\} \\ &= \int \delta p_k \left(dq^k - \frac{\partial H}{\partial p_k} dt \right) - \int \delta q^k \left(dp_k + \frac{\partial H}{\partial q^k} dt \right) + \int d(p_k \delta q^k). \end{aligned}$$

Assuming again that the δq^k vanish at the boundary, the otherwise arbitrariness of the variations of the phase space variables leads us to conclude that the two integrands vanish separately, giving rise to the equations

$$dq^k = \frac{\partial H}{\partial p_k} dt \qquad\qquad dp_k = -\frac{\partial H}{\partial q^k} dt$$

which are indeed nothing but the Hamilton equations of motion (2.10). Notice that the boundary term we are dropping here is the same in both the Lagrangian and the Hamiltonian case, namely $d(p_k \delta q^k)$.

Boundary Terms and Canonical Transformations

Since the action is invariant with respect to canonical transformations $(q, p) \to (\hat{q}, \hat{p})$, it must be the case that

$$p_k dq^k - H dt = \hat{p}_k d\hat{q}^k - \hat{H} dt + dF. \tag{2.25}$$

Writing this as $dF = p_k dq^k - \hat{p}_k d\hat{q}^k + (\hat{H} - H)dt$ we find

$$\frac{\partial F}{\partial q^k} = p^k \qquad \frac{\partial F}{\partial \hat{q}^k} = -\hat{p}^k \qquad \frac{\partial F}{\partial t} = \hat{H} - H$$

where F is considered as depending on the old and the new coordinates, i.e. $F(q, \hat{q}, t)$. It turns out that F generates a canonical transformation. Three other types of generators for canonical transformations can be obtained by taking other ways of distributing partial derivatives in (2.25). By this, in principle one can obtain generating functions depending on any of the combinations $(q, \hat{q}), (q, \hat{p}), (\hat{q}, p), (\hat{q}, \hat{p})$. For instance define

$$G := d(F + \hat{p}_k \hat{q}^k) = p_k dq^k + \hat{q}^k d\hat{p}_k + (\hat{H} - H)dt$$

where now $G(q, \hat{p}, t)$ fulfills

$$\frac{\partial G}{\partial q^k} = p_k \qquad \frac{\partial G}{\partial \hat{p}_k} = \hat{q}^k \qquad \frac{\partial G}{\partial t} = \hat{H} - H. \tag{2.26}$$

The canonical transformations generated by G include the point transformations $\hat{q}^k = \hat{q}^k(q)$ by the choice $G = f^k(q)\hat{p}_k$ and the identity transformations by the further choice $f^k(q) = q^k$. The transformation infinitesimally deviating from the identity transformation can be written as

$$G = q^k \hat{p}_k - g(q, \hat{p}) = q^k \hat{p}_k - g(q, p)$$

where the latter relation is valid, because g is infinitesimal. Then

$$\frac{\partial G}{\partial q^k} = \hat{p}_k - \frac{\partial g}{\partial q^k} = p_k \qquad \frac{\partial G}{\partial \hat{p}_k} = q^k - \frac{\partial g}{\partial \hat{p}_k} = \hat{q}^k,$$

which reproduces (2.18).

Hamilton-Jacobi Equations

If the generating function $G(q, \hat{p}, t)$ in (2.26) is chosen in such a way that the transformed Hamiltonian $\hat{H}$ vanishes, it is called Hamilton's principal function $\bar{S} = G(q, \hat{p}, t)$. Then the first and the third relation in (2.26) yield

$$H(q, p, t) + \frac{\partial \bar{S}}{\partial t} = 0 = H\left(q^k, \frac{\partial \bar{S}}{\partial q^k}, t\right) + \frac{\partial \bar{S}}{\partial t} \tag{2.27}$$

the latter expression defining the Hamilton-Jacobi equation. Since by assumption the transformed Hamiltonian is identically zero, the corresponding Hamilton equations are simply $\dot{\hat{q}}^k = 0 = \dot{\hat{p}}_k$. Therefore the new variables are constants of motion: $\hat{p}_k = \alpha_k$, $\hat{q}^k = \beta^k$. Therefore the Hamilton-Jacobi equation (2.27) is a differential equation for $\bar{S}(q, \alpha, t)$. Assuming that we can find a solution of this equation, we may solve the relation $\hat{q}^k = \partial \bar{S}(q, \alpha, t)/\partial \alpha_k$ as $q^i(\alpha, \beta, t)$ and

$$p_i = \left.\frac{\partial \bar{S}(q, \alpha, t)}{\partial q^i}\right|_{q^i = q^i(\alpha, \beta, t)}.$$

Therefore the solutions for the coordinates and the momenta are explicitly expressed by conserved quantities.

Since $\bar{S}$ is a function only of q^i and t we find due to the first relation in (2.26) and to (2.27)

$$d\bar{S} = \frac{\partial \bar{S}}{\partial q^i} dq^i + \frac{\partial \bar{S}}{\partial t} dt = \left(p_i \frac{dq^i}{dt} - H\right) dt = L dt$$

revealing that $\bar{S}$ is the indefinite time integral of the Lagrangian. Therefore the action can be expressed as $S = \bar{S}(t_2)$ - $\bar{S}(t_1)$. In the case that the Hamiltonian H is independent of time, the principal function becomes $\bar{S}(q, \hat{p}, t) = W(q, \hat{p})$ - Et with the energy E and Hamilton's characteristic function W, which obeys

$$H\left(q^k, \frac{\partial W}{\partial q^k}\right) = E.$$

*2.1.5 *Classical Mechanics in Geometrical Terms*

All of the preceding findings for classical mechanics were written in terms of generalized coordinates, velocities, momenta, ... These are local expressions which hide the essential geometric structures of analytical mechanics. Indeed the Lagrangian and the Hamiltonian description can properly be formulated on a tangent and a cotangent bundle. These geometric concepts are in more detail explained in Appendix E.

Lagrangian Dynamics

The configuration space is assumed to be a manifold $\mathbb{Q}$ (with coordinates q^i). In order to describe the dynamics (locally given by $q^i(t)$) we need to consider the configuration-velocity space which is the tangent bundle $T\mathbb{Q}$ with local coordinates $(q^i, \dot{q}^i)$. The time development of any function $f(q, \dot{q}) \in \mathcal{F}(T\mathbb{Q}) : T\mathbb{Q} \to \mathbb{R}$ (let us for simplicity assume that the configuration-velocity space functions do not explicitly depend on time) is defined in terms of the vector field

$$\Delta = \dot{q}^i \frac{\partial}{\partial q^i} + a^i(q, \dot{q})^i \frac{\partial}{\partial \dot{q}^i} \tag{2.28}$$

where the a^i, having the meaning of accelerations, will be determined below as

$$\dot{f} = \dot{q}^i \frac{\partial f}{\partial q^i} + a^i(q, \dot{q}) \frac{\partial f}{\partial \dot{q}^i} = \pounds_\Delta f.$$

The Lagrangian is a function on the tangent bundle[5] $L : T\mathbb{Q} \to \mathbb{R}$. Introduce the (Cartan) one-form

$$\theta_L := \frac{\partial L}{\partial \dot{q}^i} dq^i. \tag{2.29}$$

This gives rise to a natural two-form

$$\omega_L := -d\theta_L = \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial q^k} \right) dq^i \wedge dq^k + \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^k} \right) dq^i \wedge d\dot{q}^k. \tag{2.30}$$

Now

$$\begin{aligned} \pounds_\Delta \theta_L &:= \pounds_\Delta \left(\frac{\partial L}{\partial \dot{q}^i} \right) dq^i + \frac{\partial L}{\partial \dot{q}^i} d\pounds_\Delta q^i \\ &= \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) dq^i + \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i \doteq \frac{\partial L}{\partial q^i} dq^i + \frac{\partial L}{\partial \dot{q}^i} d\dot{q}^i = dL, \end{aligned}$$

where, indicated by the $\doteq$ notation, the coordinate version of the Euler-Lagrange equations was used. Therefore the coordinate-free form of the Lagrange equation is

$$\pounds_\Delta \theta_L = dL. \tag{2.31}$$

---

[5] Throughout this subsection it is assumed for simplicity that there is no explicit dependence on time in any of the entities involved. Dropping this assumption amounts to investigating the Lagrangian on the contact manifold $T\mathbb{Q} \times \mathbb{R}$.

This becomes explicitly $-i_\Delta d\theta_L = i_\Delta \omega_L = d(i_\Delta \theta_L - L)$ and can therefore be written in terms of the Lagrange energy E as

$$i_\Delta \omega_L = dE \qquad \text{with} \qquad E := i_\Delta \theta_L - L = \frac{\partial L}{\partial \dot{q}^k}\dot{q}^k - L. \tag{2.32}$$

Let us find out under which conditions the equation $i_X\omega_L = dE$ has a unique solution for the vector field X. On the one hand,

$$dE = \frac{\partial E}{\partial q^i}dq^i + \frac{\partial E}{\partial \dot{q}^i}d\dot{q}^i = -V_i dq^i + W_{ij}\dot{q}^j d\dot{q}^i \tag{2.33}$$

with the notation as in (2.6).
On the other hand, for a generic vector field $X = A^i \frac{\partial}{\partial q^i} + B^i \frac{\partial}{\partial \dot{q}^i}$ we find

$$i_X\omega_L = V_{ij}(A^i dq^j - A^j dq^i) + W_{ij}(A^i dq^j - B^i dq^j)$$

with the abbreviation

$$V_{ij} = \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} \qquad \text{such that} \qquad V_i = \frac{\partial L}{\partial q^i} - V_{ji}\dot{q}^j.$$

The comparison with (2.33) results in

$$\frac{\partial L}{\partial q^i} - V_{ji}\dot{q}^j = (V_{ij} - V_{ji})A^j + W_{ij}B^j \tag{2.34a}$$

$$W_{ij}\dot{q}^j = W_{ij}A^j. \tag{2.34b}$$

The integral curves of X are given by

$$\frac{dq^i}{d\lambda} = A^i, \qquad \frac{d\dot{q}^i}{d\lambda} = B^i.$$

Thus one could be tempted to identify from (2.34b) $\dot{q}^j$ with A^j. But this is only allowed if the Hessian W_{ij} can be inverted. Only in this case (2.34a) do have a unique solution with

$$A^i = \dot{q}^i, \qquad B^i = \bar{W}^{ij}V_j$$

which is indeed the Lagrangian vector field (2.28) with $a^i = B^i = \ddot{q}^i(q, \dot{q})$.

Hamiltonian Dynamics

Hamiltonian mechanics takes place on the cotangent space $T^*\mathbb{Q}$, coordinized by (q^i, p_i). The Hamiltonian is a function $H : T^*\mathbb{Q} \to \mathbb{R}$. Introduce a vector field

$$\nabla = \dot{q}^i \frac{\partial}{\partial q^i} + \dot{p}_i \frac{\partial}{\partial p_i}$$

and define the two-form

$$\omega = dq^i \wedge dp_i$$

which derives from the (Liouville) one-form $\theta_H = p_i dq^i$ as $\omega = -d\theta_H$. In denoting coordinates in $T^*\mathbb{Q}$ as (x^α) we find that the components of $\omega = 1/2\, \omega_{\alpha\beta}\, dx^\alpha \wedge dx^\beta$ are identical to the symplectic matrix Γ as in (2.16).

Now calculate

$$i_\nabla \omega = i_\nabla dq^i \wedge dp_i - dq^i \wedge i_\nabla dp_i = \dot{q}^i dp_i - \dot{p}_i dq^i \doteq \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial q^i} dq^i = dH.$$

This reveals that the coordinate independent Hamilton equations of motion are

$$i_\nabla \omega = dH. \tag{2.35}$$

Indeed they are formulated completely in terms of the vector field ∇ and the two-form ω (both being defined within the cotangent space geometry), and the cotangent space function H.

The two-form ω is a representation of the Poisson bracket structure of Hamiltonian dynamics in the following sense: Associate to every function $f \in \mathcal{F}(T\mathbb{Q})$ a vector field X_f by

$$i_{X_f} \omega = df, \qquad \text{in coordinates} \qquad X_f = \frac{\partial f}{\partial p_i} \frac{\partial}{\partial q^i} - \frac{\partial f}{\partial q^i} \frac{\partial}{\partial p_i}. \tag{2.36}$$

A vector field X_f associated to a cotangent space function f is called a Hamiltonian vector field[6] or a Hamiltonian with respect to f. The equations (2.35) show that the vector field ∇ is the Hamiltonian vector field with respect to the Hamilton function.

Calculate $\pounds_{X_g} f$ and find that in coordinates this reproduces the Poisson bracket $\{f, g\}$, thus

$$\{f, g\} = \pounds_{X_g} f = i_{X_g} df = i_{X_g} i_{X_f} \omega, \tag{2.37}$$

and for two Hamiltonian vector fields:

$$[X_f, X_g] = X_{\{f,g\}}. \tag{2.38}$$

[6] Not all vector fields are Hamiltonian, which tells that there are more conceivable motions on $T^*\mathbb{Q}$ beyond those described by Hamiltonian dynamical systems.

The dynamical evolution of a space function can be written as $\pounds_\nabla f = \{f, H\}$.

Canonical transformations are a subset of diffeomorphisms $F : T^*\mathbb{Q} \to T^*\mathbb{Q}$, namely those for which the symplectic form is preserved: $F^*\omega = \omega$. Indeed the pull-back of $i_\nabla \omega$ then yields

$$F^* i_\nabla \omega = i_{F^*\nabla} F^*\omega = i_{F^*\nabla}\omega = dF^* H$$

showing that the Hamilton equations (2.35) are preserved. Further taking the exterior derivative of the relation $i_{X_f}\omega = df$ that defines the Hamiltonian vector field with respect to the function f,

$$0 = d(i_{X_f}\omega) = \pounds_{X_f}\omega.$$

This demonstrates that the symplectic form is invariant under any Hamiltonian flow. In using the relation (E.3) between a diffeomorhism and a vector field and the relation (2.36) between a vector field and a function one can associate to any function f a one-parameter group of transformations Φ^f. In the language of analytical mechanics the function f is called the infinitesimal generator of the canonical transformation.

Legendre Transformation

In the previous two subsections, the Lagrangian and the Hamiltonian dynamics were described in purely geometric terms, but separately of each other. What is known as the Legendre transformation from the one description to the other, needs to be grasped in geometric terms as a specific mapping

$$\begin{aligned} \mathcal{F}L : T\mathbb{Q} &\to T^*\mathbb{Q} \\ (q, \dot{q}) &\mapsto \mathcal{F}L(q, \dot{q}) = \left(q, \hat{p} = \frac{\partial L}{\partial \dot{q}}\right). \end{aligned} \tag{2.39}$$

This is a mapping of the vector with components $\dot{q}^k$ in $T_q\mathbb{Q}$ onto its covector $\hat{p}_k = \partial L/\partial \dot{q}^k$ in $T^*_q\mathbb{Q}$. In other words, it maps the one form $\theta_L = (\partial L/\partial \dot{q}^k)$ $dq^k \in TQ$ onto the one-form $\hat{p}_i dq^i \in T^*\mathbb{Q}$. On the other hand, independently of any Lagrangian, in the cotangent bundle there exists a Liouville form $\theta_H = p_i dq^i$. Obviously both θ_L and the Legendre mapping depend on the Lagrangian. It is not at all obvious that these dependencies act in such a way that θ_L is always sent to the same canonical one-form θ_H. And indeed, for this to happen one can show that a necessary condition is–lo and behold–the nonsingularity of the Hessian W_{ij}. This is also seen in that the pull-back $\omega_L = \mathcal{F}L^*\omega$ given by (2.30) has components which in matrix form can be written as

$$\begin{pmatrix} A & W \\ -W & 0 \end{pmatrix}$$

(with $A = V\text{-}V^T$ and V_{ij} as in (2.34a)) since all components proportional to $d\dot{q}^i \wedge d\dot{q}^k$ do vanish. This matrix is nonsingular only for det $W \neq 0$. Only in this case is the two-form ω_L symplectic (non-degenerate and closed). The non-singularity of W is of course also the condition necessary to solve $\hat{p}_k = (\partial L/\partial \dot{q}^k)(q, \dot{q})$ uniquely as $\dot{q}^i = \dot{q}^i(q, \hat{p})$ and to identify $\hat{p}_k \equiv p_k$. For regular systems, the Hamiltonian is the projection of the Lagrangian energy $H = \mathcal{F}L(E)$. Furthermore, the connection between the Lagrangian and the Hamiltonian dynamics becomes

$$\dot{q}^k = \mathcal{F}L^* \left(\frac{\partial H}{\partial p_k}\right) \qquad \frac{\partial L}{\partial q^k} = -\mathcal{F}L^* \left(\frac{\partial H}{\partial q^k}\right). \tag{2.40}$$

To emphasize again: only with the existence of a regular Lagrangian one can define this canonical isomorphism between $T\mathbb{Q}$ and $T^*\mathbb{Q}$. This is reminiscent to what is known from differential geometry: If a manifold is equipped with a metric, this metric mediates maps between the tangent and the cotangent bundle, see Appendix E.5.4. In regular classical mechanics, this metric is visible in the kinetic energy term $T = g_{ij}\dot{q}^i\dot{q}^j$.[7]

2.2 Symmetries and Conservation Laws

2.2.1 Conservation Laws

Physic was successful–or even possible–because one was able to find "laws of nature". In a very broad sense, a law of physics encodes many observations (experiments and their results) on a physical system in a compact mathematical relation [184]. The more measurements on one and the same system are encoded the better the law is established experimentally (and may be sanctioned as a theory). The more systems are encoded by one and the same law, the more profound the unification. The least one requires is that "under the same circumstances" it does not matter whether the experiment takes place:

- TODAY or TOMORROW
- in BERLIN or in NEW YORK
- in the NORTH-SOUTH or the EAST-WEST direction
- on the COAST of the Baltic Sea or on a SAILBOAT moving uniformly with respect to the coast.

The term "under the same circumstances" needs a comment. If we investigate a physical system at another time, or in another position, orientation, relative movement, we must make sure that everything surrounding and possibly influencing the system must also be transformed correspondingly. In Chap. 11 of [181] R.P. Feynman describes a grandfather's clock as an example. If you do not place it standing upright, its pendulum will hit the case, and the clock will not work at all. So this system seems not to work in the same manner, independently of its orientation. But, if you include the earth in your consideration, you are again in the position to formulate

[7] The analogy goes even further, in that the Liouville one-form acts as a soldering form.

direction-independent statements about the system. This is related to the distinction of objects in terms of background and dynamical structures already mentioned in the Introduction.

Homogeneity of Time and Energy Conservation

If the outcome of an experiment on a physical system does not depend on when it is performed ("TODAY or TOMORROW"), the Lagrange function can not depend on time explicitly. Therefore

$$\frac{dL}{dt} = \underbrace{\frac{\partial L}{\partial q^i}}_{\left([L]_i + \frac{d}{dt}\frac{dL}{d\dot{q}^i}\right),} \dot{q}^i + \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i$$

by the use of the Euler derivatives $[L]_i$. Thus

$$\frac{dL}{dt} = [L]_i \dot{q}^i + \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\right)$$

or

$$[L]_i \dot{q}^i = \frac{d}{dt}\left(L - \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\right) = \frac{d}{dt}(-E) \doteq 0,$$

in other words, for trajectories or "on-shell" (that is, for solutions of the equations of motion) the energy E is conserved.

Homogeneity of Space and Momentum Conservation

Start from

$$L = \sum_\alpha \frac{m_\alpha \dot{\vec{x}}_\alpha}{2} \; - \; V(\vec{x})$$

and consider a translation of all position vectors by a constant vector $\vec{d}$:

$$\vec{x}_\alpha \to \vec{x}_\alpha + \vec{d},$$

i.e. $\delta\vec{x}_\alpha = \vec{d}, \quad \delta\dot{\vec{x}}_\alpha = 0$. Thus the variation of the Lagrange function becomes

$$\delta L = \sum_\alpha \frac{\partial L}{\partial \vec{x}_\alpha} \cdot \delta\vec{x}_\alpha = \vec{d} \cdot \sum_\alpha \frac{\partial L}{\partial \vec{x}_\alpha}.$$

If the measurement on the system at two different positions (“in BERLIN or in NEW YORK”) leads to the same result, the Lagrange function can only depend on relative positions $\vec{r}_{\alpha\beta} = \vec{x}_\alpha - \vec{x}_\beta$. Therefore

$$\sum_\alpha \frac{\partial L}{\partial \vec{x}_\alpha} = 0.$$

On the other hand $\partial L/\partial \vec{x}_\alpha$ can be expressed by the Euler-derivatives and the momenta in the form

$$\frac{\partial L}{\partial \vec{x}_\alpha} = [\vec{L}]_\alpha + \frac{d}{dt}\frac{\partial L}{\partial \dot{\vec{x}}_\alpha} = [\vec{L}]_\alpha + \dot{\vec{p}}_\alpha.$$

From the fact that

$$\sum_\alpha ([\vec{L}]_\alpha + \dot{\vec{p}}_\alpha) = \dot{\vec{P}} + \sum_\alpha [\vec{L}]_\alpha = 0$$

one derives, that on trajectories $([\vec{L}]_\alpha \doteq 0)$ the total momentum $\vec{P}$ is conserved.

A special case is present if the Lagrange function does not depend on a specific coordinate (called a cyclic coordinate) $\bar{q}^\kappa$. In this case the momentum $\bar{p}_\kappa$ conjugate to $\bar{q}^\kappa$ is itself a conserved quantity.

Isotropy of Space and Angular Momentum Conservation

The variation of the Lagrangean $L(\vec{x}, \dot{\vec{x}})$, namely

$$\delta L = \frac{\partial L}{\partial \vec{x}_\alpha} \cdot \delta \vec{x}_\alpha + \frac{\partial L}{\partial \dot{\vec{x}}_\alpha} \cdot \delta \dot{\vec{x}}_\alpha$$

can be expressed–like in the previous section–through the Euler-derivatives and the momenta as

$$\delta L = ([\vec{L}]_\alpha + \dot{\vec{p}}_\alpha) \cdot \delta \vec{x}_\alpha + \vec{p}_\alpha \cdot \delta \dot{\vec{x}}_\alpha.$$

Consider an infinitesimal rotation: $\delta\vec{\varphi}$ points into the direction of the rotation axis, $\delta\varphi = |\delta\vec{\varphi}|$ is the magnitude of the rotation, see Fig. 2.1. For each vector $\vec{a}$ we have because of $\delta\vec{a} \perp \vec{a}$ and $\delta\vec{a} \perp \delta\vec{\varphi}$

$$\delta\vec{a} = \gamma\, \delta\vec{\varphi} \times \vec{a} \qquad \text{where} \quad \gamma = \text{const}.$$

Therefore,

$$\delta L = \gamma \sum_\alpha ([\vec{L}]_\alpha + \dot{\vec{p}}_\alpha) \cdot (\delta\vec{\varphi} \times \vec{x}_\alpha) + \vec{p}_\alpha \cdot (\delta\vec{\varphi} \times \dot{\vec{x}}_\alpha).$$

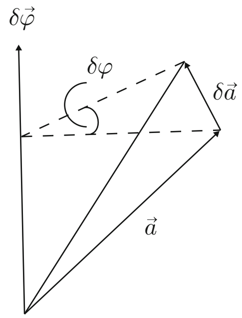


Fig. 2.1 Vector conventions for infinitesimal rotations

We can extract from this expression $\delta\vec{\varphi}$ if we make use of the relation $\vec{a}\cdot(\vec{b}\times\vec{c}) = (\vec{a}\times\vec{b})\cdot\vec{c}$:

$$\begin{aligned}\delta L &= \gamma\ \delta\vec{\varphi}\cdot\sum_\alpha \left\{\vec{x}_\alpha \times ([\vec{L}]_\alpha + \dot{\vec{p}}_\alpha) + \dot{\vec{x}}_\alpha \times \vec{p}_\alpha\right\}\\ &= \gamma\ \delta\vec{\varphi}\cdot\sum_\alpha \left\{x_a \times [\vec{L}]_\alpha + \frac{d}{dt}(\vec{x}_\alpha \times \vec{p}_\alpha)\right\}.\end{aligned}$$

If the variation δL vanishes for arbitrary $\delta\vec{\varphi}$ we conclude that the total angular momentum $\vec{J} = \sum_\alpha(\vec{x}_\alpha \times \vec{p}_\alpha)$ is a constant for solutions of the dynamical equations. The argument can be extended to finite (non-infinitesimal) rotations, exemplifying that if the experiment reveals the same observations "in the NORTH-SOUTH and the EAST-WEST direction" necessarily the total angular momentum is conserved.

Galilei Relativity and Uniform Center-of-Mass Velocity

Galilei relativity is the invariance of classical mechanics with respect to Galilei translations. These are transformations from one system of coordinates to another system which moves with a constant velocity $\vec{v}$ with respect to the former. An example is a "SAILBOAT moving uniformly with respect to a COAST". Quantitatively

$$\vec{x}'_\alpha = \vec{x}_\alpha + \vec{v}t.$$

Introduce the quantity $\vec{Q} = \sum_\alpha(m_\alpha\vec{x}_\alpha - \vec{p}_\alpha t)$ and consider the expression

$$\frac{d}{dt}(\vec{v}\cdot\vec{Q}) = \vec{v}\cdot\frac{d}{dt}\left(\sum_\alpha m_\alpha\vec{x}_\alpha - \vec{p}_\alpha t\right) = \vec{v}\cdot\sum_\alpha(m_\alpha\dot{\vec{x}}_\alpha - \dot{\vec{p}}_\alpha t - \vec{p}_\alpha).$$

Now by definition, $(m_a\dot{\vec{x}}_\alpha - \vec{p}_\alpha) \equiv 0$, and $\dot{\vec{p}}_\alpha$ can be expressed by the Newtonian derivatives (2.2) so that

$$\frac{d}{dt}(\vec{v}\cdot\vec{Q}) = \vec{v}\cdot\sum_\alpha(\vec{\nabla}_\alpha V - \vec{N}_\alpha).$$

If V depends on the coordinate distances $r_{ij} = |\vec{x}_i - \vec{x}_j|$ only, $\sum_\alpha \vec{\nabla}_\alpha V$ vanishes. Therefore, again on trajectories ($\vec{N}_\alpha \doteq 0$) the complete left-hand side vanishes. Since $\vec{v}$ is arbitrary we deduce $\frac{d}{dt}\vec{Q} = 0$. Now

$$\vec{Q} = M\vec{R} - \vec{P}t \tag{2.41}$$

with the center of mass $\vec{R}$ ($M = \sum_\alpha m_\alpha$) and the total momentum $\vec{P}$. Since the total momentum is conserved, we find that the center of mass moves with uniform velocity.

"Deriving" the Lagrangian from Properties of Space-Time and from Galilei Relativity

The next consideration is not directly related to conservation laws. But in order to derive the standard conserved quantities in previous subsections we used characteristics of space and time and assumed that the Lagrangian has the form $L = T\text{-}V$. Indeed, one may turn the arguments around to derive the structure of the Lagrangian[8]. To begin with, consider a free mass point. Because of homogeneity of space and time, the Lagrangian cannot depend on the position $\vec{x}$ and on the time variable t. Therefore it can only depend on the velocity $\dot{\vec{x}}$. Furthermore, because of the isotropy of space, the Lagrangian cannot depend on the direction of velocity. Therefore $L = L(\dot{x}^2)$. The change of the Lagrangian under an infinitesimal transformation to another inertial system with $\dot{\vec{x}} \to \dot{\vec{x}}' = \dot{\vec{x}} + \vec{\epsilon}$ is

$$L' = L(\dot{x}'^2) = L(\dot{x}^2 + 2\dot{\vec{x}} \cdot \vec{\epsilon} + \epsilon^2) = L(\dot{x}^2) + \frac{\partial L}{\partial \dot{x}^2} 2\dot{\vec{x}} \cdot \vec{\epsilon} + \mathcal{O}(\epsilon^2).$$

Requiring that in both inertial systems the dynamics stays the same, the Lagrangians L and L' are allowed to only differ by a total derivative. Therefore $\frac{\partial L}{\partial \dot{x}^2}$ must be a constant, or $L = a\dot{x}^2$; the constant a is identified as $a = m/2$, with m being the mass. (But be aware that this identification can only be justified by including a second point particle.) In any case we derived the standard kinetic energy part T. This can be generalized to a system of free mass points, and we notice that the kinetic part of the Lagrangian transforms under Galilei translations (2.1) as

$$\begin{aligned} L(\dot{\vec{x}}) \;\to\; L' &= \sum_\alpha \frac{m_\alpha}{2} \frac{d\vec{x}'^2_\alpha}{dt'} = \sum_\alpha \frac{m_\alpha}{2}\left[\dot{\vec{x}}^2_\alpha + 2\dot{\vec{x}}_\alpha \vec{v} + v^2\right] \\ &= L + \frac{d}{dt}\sum_\alpha \frac{m_\alpha}{2}\left[2\vec{x}_\alpha \vec{v} + v^2 t\right]. \end{aligned}$$

The last term is a total time derivative and we say that the Lagrangian for a system of free mass points is quasi-invariant by going from inertial system to another one.

[8] In the sequel I follow [332], but you can find this line of reasoning in other textbooks on classical mechanics as well.

For a system of mass points without external forces the potential $V(\vec{x})$ only depends on the relative distances $r_{\alpha\beta} = |\vec{x}_\alpha - \vec{x}_\beta|$, so that this term is itself invariant under Galilei translations.

Instead of assuming from the very start the homogeneity of space and time one can arrive at the same result [459] by considering the Euler-Lagrange equations for a free particle as derived from a Lagrangian $L(\dot{x}, x, t)$:

$$\frac{\partial^2 L}{\partial \dot{x}^\alpha \partial \dot{x}^\beta}\,\ddot{x}^\beta + \frac{\partial^2 L}{\partial \dot{x}^\alpha \partial x^\beta}\,\dot{x}^\beta + \frac{\partial^2 L}{\partial \dot{x}^\alpha \partial t} - \frac{\partial L}{\partial x^\alpha} = 0.$$

Let us demand that this dynamical equation is invariant under Galilei transformations. Now, only the first term depends on $\ddot{x}$, and $\ddot{x}$ does not change under Galilei transformations. Therefore the coefficient must be a constant

$$\frac{\partial^2 L}{\partial \dot{x}^\alpha \partial \dot{x}^\beta} = k_{\alpha\beta}.$$

This can be integrated to

$$L = \frac{1}{2} k_{\alpha\beta}\,\dot{x}^\alpha \dot{x}^\beta + \dot{x}^\beta F_\beta(x,t) + G(x,t).$$

The Euler-Lagrange equations then become

$$k_{\alpha\beta}\,\ddot{x}^\beta + \dot{x}^\beta \left[\frac{\partial F_\alpha}{\partial x^\beta} - \frac{\partial F_\beta}{\partial x^\alpha}\right] + \frac{\partial F_\alpha}{\partial t} - \frac{\partial G}{\partial x^\alpha} = 0.$$

By the previous argumentation, the first term is invariant with respect to Galilei transformations. The second term is the only one depending on velocities. This is required to vanish, and therefore $[F_{\alpha,\beta} - F_{\beta,\alpha}] = 0$ or

$$F_\alpha = \Phi_{,\alpha} \qquad \text{and} \qquad F_{\alpha,t} - G_{,\alpha} = K_\alpha$$

with a constant K_α. The latter condition is solved by $G = \Phi_{,t} - K_\alpha x^\alpha + C(t)$. Inserted into the previous Lagrangian this becomes

$$\begin{aligned} L &= \frac{1}{2} k_{\alpha\beta}\,\dot{x}^\alpha \dot{x}^\beta + \dot{x}^\beta \Phi_{,\beta} + \Phi_{,t} - K_\alpha x^\alpha + C(t) \\ &= \frac{1}{2} k_{\alpha\beta}\,\dot{x}^\alpha \dot{x}^\beta - K_\alpha x^\alpha + \frac{d}{dt}\left(\Phi + \int C\,dt\right). \end{aligned}$$

The last term, being a total derivative, can be dropped. The equations of motion become $k_{\alpha\beta}\,\ddot{x}^\beta + K_\alpha = 0$. Next we require that these equations are invariant with respect to rotations $x^\alpha \to R^\alpha_{\ \beta} x^\beta$. The transformed Lagrangian and equations of motion become

$$L' = \frac{1}{2} k_{\alpha\beta} R^\alpha_{\ \gamma} R^\beta_{\ \delta}\,\dot{x}^\gamma \dot{x}^\delta - K_\alpha R^\alpha_{\ \beta} x^\beta \qquad k_{\alpha\beta} R^\alpha_{\ \delta} R^\beta_{\ \gamma}\,\ddot{x}^\gamma + K_\alpha R^\alpha_{\ \delta} = 0.$$

Invariance under rotations is ensured if $K_\alpha R^\alpha_{\ \delta} = K_\delta$ and $k_{\alpha\beta} R^\alpha_{\ \delta} R^\beta_{\ \gamma} = k_{\delta\gamma}$. And this is the case if $K_\alpha = 0$ and $k_{\alpha\beta} = k\delta_{\alpha\beta}$, that is if $L = \frac{k}{2}\dot{x}^2$.

2.2.2 Noether Theorem–A First Glimpse

In the previous section we became acquainted with relations between certain invariance properties of the Lagrange function and conservation laws. In each case, we argued in a different manner, but common to all the cases is the fact that the conservation law holds only "on-shell", that is only for solutions of the equations of motion. Emmy Noether[9] (1882-1935) was able, amongst others, to relate conservation laws directly to the underlying symmetry group in the case of continuous symmetries.

The Noether theorems (actually two theorems are usually distinguished) relate to continuous symmetries of an action. Let us assume that we are dealing with point transformations[10]

$$q^k \mapsto \hat{q}^k(q,t) \qquad\qquad t \mapsto \hat{t}(t,q).$$

The action functional

$$S[q] = \int_{t_1}^{t_2} dt L(q], \qquad L[q] := L(q^k, \dot{q}^k, t)$$

becomes in the new variables

$$S[q] = \int_{\hat{t}(t_1)}^{\hat{t}(t_2)} d\hat{t} \ \left(\frac{dt}{d\hat{t}}\right) L[q(\hat{q},\hat{t})] := \hat{S}[\hat{q}].$$

A variational symmetry fulfills $S[\hat{q}] \stackrel{!}{=} S[q]$, i.e.

$$\underbrace{\int_{\hat{t}(t_1)}^{\hat{t}(t_2)} d\hat{t} L[\hat{q}]}_{\displaystyle\int_{t_1}^{t_2} dt \frac{d\hat{t}}{dt} L[\hat{q}],} \stackrel{!}{=} \int_{t_1}^{t_2} dt L[q]$$

so that

$$\left(\frac{d\hat{t}}{dt}\right) L[\hat{q}] \stackrel{!}{=} L[q] + \frac{d}{dt}\Sigma_S(q,t). \tag{2.42}$$

---

[9] In the literature you can spot people who show their acquaintance with German by knowing that in many words the "oe" means "ö". But this is not always true, especially for proper names.

[10] The theorem also holds for generalized transformations in which the "new" variables also depend on the "old" velocities $\dot{q}^i$. Later it will be shown that if velocity-dependent transformations are allowed one can derive the conservation of the Runge-Lentz vector of the Kepler problem from a Noether theorem.

The Lagrange function is called invariant iff $\Sigma_S = 0$, and quasi-invariant otherwise. By notation, the boundary term depends on the symmetry transformation.

The Noether theorems only hold for continuous symmetry transformations, namely those which are continuously attainable from the identity transformations. This is for instance not true for time reversal and space inversion. In case of continuous symmetries, we can restrict ourself to transformations near the identity:

$$\hat{q}^k = q^k + \delta_S q^k(q,t) = q^k + \epsilon\eta^k(q,t) \qquad \hat{t} = t + \delta_S(t,q) = t + \epsilon\xi(t,q)$$

with the understanding that every one-parameter continuous set of symmetry transformations is characterized by the parameter ϵ and functions $\eta^k(q,t)$ and $\xi(t,q)$. For these infinitesimal transformations the requirement (2.42) reads

$$\left(1 + \frac{d}{dt}\delta_S t\right) L[\hat{q}] \stackrel{!}{=} L[q] + \frac{d}{dt}\sigma_S(q,t),$$

where the notation σ_S indicates that we are arguing infinitesimally. With the definition

$$\delta_S L := L[\hat{q}] - L[q]$$

the previous expression can be rewritten as

$$\delta_S L + L[q]\frac{d}{dt}\delta_S t \stackrel{!}{=} \frac{d}{dt}\sigma_S(q,t).$$

After applying the chain rule it becomes

$$\delta_S L - \frac{dL}{dt}\delta_S t \stackrel{!}{=} \frac{d}{dt}(\sigma_S - L\delta_S t). \tag{2.43}$$

The terms on the left-hand side are explicitly

$$\begin{aligned} \delta_S L &= \frac{\partial L}{\partial q}\delta_S q + \frac{\partial L}{\partial \dot{q}}\delta_S \dot{q} + \frac{\partial L}{\partial t}\delta_S t. \\ \frac{dL}{dt}\delta_S t &= \left(\frac{\partial L}{\partial q}\dot{q} + \frac{\partial L}{\partial \dot{q}}\ddot{q} + \frac{\partial L}{\partial t}\right)\delta_S t. \end{aligned}$$

(Here the coordinate indices are dropped for convenience–or–laziness; they will be re-introduced at the end result.) The entire left-hand side of (2.43) becomes

$$\delta_S L - \frac{dL}{dt}\delta_S t = \frac{\partial L}{\partial q}\bar{\delta}_S q + \frac{\partial L}{\partial \dot{q}}\bar{\delta}_S \dot{q}, \tag{2.44}$$

with

$$\bar{\delta}_S q := \delta_S q - \dot{q}\delta_S t = \epsilon(\eta - \dot{q}\xi) \qquad \bar{\delta}_S \dot{q} := \delta_S \dot{q} - \ddot{q}\delta_S t. \tag{2.45}$$

This so-called form variation arises here as a convenient abbreviation. As a matter of fact it has a geometric meaning, denoting the difference between the coordinates at two moments of time separated by δt:

$$\bar{\delta} q := \hat{q}(t) - q(t).$$

In contrast, the total variation δq is the difference between the new and the original coordinate compared at the same point of time: $\delta q = \hat{q}(\hat{t})$ - $q(t)$. The $\bar{\delta}$-variation[11] commutes–in contrast to the δ-variation–with the derivatives:

$$\bar{\delta} \frac{d}{dt} q = \frac{d}{dt} \bar{\delta} q.$$

Let's see how this comes about: From $\delta q = \hat{q}(\hat{t})$ - $q(t)$

$$\begin{aligned}(\delta q)^{\cdot} = \frac{d}{dt}\delta q &= \frac{d\hat{q}}{d\hat{t}}\frac{d\hat{t}}{dt} - \frac{dq}{dt} = \frac{d\hat{q}}{d\hat{t}} + \frac{d\hat{q}}{d\hat{t}}\frac{d}{dt}\delta t - \frac{dq}{dt} \\ &= \delta \frac{d}{dt} q + \frac{d\hat{q}}{d\hat{t}}\frac{d}{dt}\delta t = \delta \dot{q} + \frac{d\hat{q}}{d\hat{t}}(\delta t)^{\cdot}.\end{aligned}$$

On the other hand, from (2.45)

$$\frac{d}{dt}\bar{\delta} q = (\bar{\delta} q)^{\cdot} = (\delta q)^{\cdot} - (\dot{q}\delta t)^{\cdot} \qquad\qquad \bar{\delta}\frac{d}{dt} q = (\bar{\delta}\dot{q}) = \delta\dot{q} - \ddot{q}\delta t$$

so that

$$\frac{d}{dt}\bar{\delta} q - \bar{\delta}\frac{d}{dt} q = (\delta q)^{\cdot} - (\dot{q}\delta t)^{\cdot} - \delta\dot{q} + \ddot{q}\delta t = \frac{d\hat{q}}{d\hat{t}}(\delta t)^{\cdot} - \dot{q}(\delta t)^{\cdot} = \delta\dot{q}(\delta t)^{\cdot} \sim 0$$

and this vanishes because it is a term of second order in the infinitesimal transformations considered here.

In terms of the form variation $\bar{\delta}$ the condition (2.43) becomes

$$\Delta_S L := \bar{\delta}_S L - \frac{d}{dt}\left(\sigma_S - L\delta_S t\right) \stackrel{!}{=} 0. \tag{2.46}$$

This is one way to state the invariance of a Lagrangian with respect to a symmetry transformation δ_S. Another form, useful as a further step towards Noether's theorems is to rewrite (2.44) in terms of the Euler-derivative of the Lagrangian:

$$\delta L - \frac{dL}{dt}\delta t = \underbrace{\left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right)}_{[L]_q.}\bar{\delta} q + \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\bar{\delta} q\right)$$

---

[11] The $\bar{\delta}$ notation was already introduced by E. Noether; today it is most often called the "active" variation, mathematicians denote it as Lie variation.

Then (2.43) can be written in the form

$$[L]_q \bar{\delta}_S q + \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\bar{\delta}_S q + L\delta_S t - \sigma_S\right) \equiv 0,$$

or, if the indexed position variables q^k are re-introduced,

$$[L]_k \bar{\delta}_S q^k + \frac{d}{dt} J_S \equiv 0 \tag{2.47}$$

with

$$J_S(t, q, \dot{q}) := \frac{\partial L}{\partial \dot{q}^k}\bar{\delta}_S q^k + L\delta_S t - \sigma_S. \tag{2.48}$$

In the expression for J_S we identify three terms, which can easily be understood by comparing this with the variation procedure leading to the Euler-Lagrange equations (2.21). In that derivation the time t was not varied, and thus the last term did not show up. Therefore also $\bar{\delta} = \delta$ and (2.48) boils down to (2.21), specialized to $\delta = \delta_S$. The current (2.48) can alternatively be expressed as

$$J_S(t, q, \hat{p}) := \hat{p}_k \delta_S q^k + (L - \hat{p}_k \dot{q}^k)\delta_S t - \sigma_S = \hat{p}_k \delta_S q^k - H_C \delta_S t - \sigma_S \tag{2.49}$$

with the generalized momenta $\hat{p}_k := \frac{\partial L}{\partial \dot{q}^k}$. Here H_C is the Hamiltonian in the case of regular Lagrangians and the "canonical" Hamiltonian in the case of singular systems. As shown in Appendix C, the canonical Hamiltonian is a function of the momenta and the coordinates, despite the fact that the Legendre transformation from the Lagrangian to the Hamiltonian is not invertible.

If in the identity (2.47) we insert the Euler-Lagrange equations in the form (2.6)

$$\Big(V_i(q, \dot{q}) - \ddot{q}^k W_{ki}(q, \dot{q})\Big)\,\bar{\delta}_S q^i + \frac{dJ_S(q, \dot{q}, t)}{dt} = 0$$

and observe the dependence on the second derivatives $\ddot{q}$, this splits into the identities

$$V_i \bar{\delta}_S q^i + \dot{q}^i \frac{\partial J_S}{\partial q^i} + \frac{\partial J_S}{\partial t} \equiv 0 \qquad\qquad \frac{\partial J_S}{\partial \dot{q}^i} - W_{ik}\bar{\delta}_S q^k \equiv 0. \tag{2.50}$$

For regular systems the symmetry transformations are formally related to the currents by

$$\bar{\delta}_S q^i = \bar{W}^{ik}\frac{\partial J_S}{\partial \dot{q}^k}. \tag{2.51}$$

Inserting these into the first part of (2.50) returns the on-shell conservation of J_S because of $V_i \bar{W}^{ik} = \ddot{q}^k$. Things become different for singular systems; see Appendix C.5. Equation (2.51) has advantages in formal proofs. Inserting into it the explicit form (2.48) of the current, it becomes $\frac{\partial \sigma_S}{\partial \dot{q}^i} = 0$ in the present case of velocity-independent transformations.

It is advantageous to rephrase the Noether condition in terms of the functions η and ξ as

$$\delta_\epsilon t = \epsilon\,\xi(t,q) \qquad \delta_\epsilon q^k = \epsilon\,\eta^k(t,q) \qquad \bar{\delta}_\epsilon q^k = \epsilon\,(\eta^k - \dot{q}^k \xi) =: \epsilon\,\chi^k. \quad (2.52)$$

The infinitesimal transformations are generated by the differential operator

$$X = \xi(t,q)\frac{\partial}{\partial t} + \eta^k(t,q)\frac{\partial}{\partial q^k}$$

in that $Xt = \xi$ and $Xq^k = \eta^k$. (On the other hand, the one-parameter subgroup is recovered by integrating the vector field X; this is made explicit in Sect. 2.2.4.) For a function $F(t,q)$ we obtain its variation as $\delta_\epsilon F = \epsilon XF$. This can be extended to functions $G(t,q,\dot{q},\ddot{q},\cdots)$ as $\delta_\epsilon G = \epsilon\overline{X}G$ with the vector field

$$\overline{X} = X + \eta^k_{(1)}(t,q,\dot{q})\,\frac{\partial}{\partial \dot{q}^k} + \eta^k_{(2)}(t,q,\dot{q},\ddot{q})\,\frac{\partial}{\partial \ddot{q}^k} + \cdots \quad (2.53)$$

provided the coefficient functions $\eta^k_{(a)}$ are chosen consistently: Consider at first

$$\frac{d\hat{q}}{d\hat{t}} = \frac{dq + \epsilon d\eta}{dt + \epsilon d\xi} = \frac{\dot{q} + \epsilon\dot{\eta}}{1 + \epsilon\dot{\xi}} = (\dot{q} + \epsilon\dot{\eta})(1 - \epsilon\dot{\xi} + \mathcal{O}(\epsilon^2)) = \dot{q} + \epsilon\dot{\eta} - \epsilon\dot{q}\dot{\xi} + \mathcal{O}(\epsilon^2).$$

Thus if (q^k, t) transform according to (2.52) the transformation of $\dot{q}^k$ is fixed to

$$\delta_\epsilon \dot{q}^k = \epsilon\,\eta^k_{(1)}(t,q) \qquad \text{with} \qquad \eta^k_{(1)} = \dot{\eta}^k - \dot{q}^k\dot{\xi}.$$

Notice that for all quantities g the $\dot{g}$ always stands for the total t-derivative. Although a little superfluous at this stage, I introduce the differential operator

$$d_t g := \left(\frac{\partial}{\partial t} + \dot{q}^k\frac{\partial}{\partial q^k} + \ddot{q}^k\frac{\partial}{\partial \dot{q}^k} + \cdots\right) g. \quad (2.54)$$

For the higher terms in (2.53) one finds

$$\eta^k_{(I)} = d_t\eta^k_{(I-1)} - (d_t^I q^k)\,(d_t\xi) \qquad \text{with} \qquad \eta^k_{(0)} = \eta^k.$$

In using this expression one is able to show that (2.53) can be written as

$$\overline{X} = \xi d_t + \chi^k\frac{\partial}{\partial q^k} + (d_t\chi^k)\frac{\partial}{\partial \dot{q}^k} + (d_t d_t\chi^k)\frac{\partial}{\partial \ddot{q}^k} + \cdots \quad (2.55)$$

where $\chi^k = \eta^k - \dot{q}^k\xi$ is the expression that became introduced together with the $\bar{\delta}$-transformation; see (2.52). In the mathematics literature the generator $\overline{X}$ is called the "prolongation" of the symmetry generating operator X. All those objects $I(t,q,\dot{q},\ddot{q},...)$ are called invariant for which $\overline{X}I = 0$. The concept of using infinitesimal generators for the investigation of symmetries was an essential discovery of Sophus Lie in his attempt to give a systematics to solution methods of differential equations.

After this mathematical prologue let us get back to the Noether theorem: In a Taylor expansion of the right-hand-side in (2.42) one obtains the identity

$$\xi \frac{\partial L}{\partial t} + \eta^k \frac{\partial L}{\partial q^k} + \eta^k_{(1)} \frac{\partial L}{\partial \dot{q}^k} + L \frac{d}{dt}\xi = \overline{X} L + L\, d_t \xi = d_t \sigma \qquad (2.56)$$

(with $\sigma_S = \epsilon\sigma$). This can be interpreted as a partial differential equation to be fulfilled by the functions η^k and ξ in a Noether symmetry transformation. You can also read it as a condition which must be fulfilled by the surface part σ for given η and ξ, and gives a clue to derive a possible surface part. Notice that together with $\overline{X}$ also $\hat{X} = \overline{X} + \lambda d_t$ fulfills (2.56) as long $d_t \lambda = 0$. Also notice that (2.56) holds as well if the ξ and η^k depend on velocities $\dot{q}$, that is for non-point transformations.

Noether Charges

In writing $J_\epsilon = \epsilon\, C$ we have

$$C(q, \dot{q}, t) = \frac{\partial L}{\partial \dot{q}^k} \chi^k + L\xi - \sigma,$$

and because of (2.47) this is on-shell conserved, in other words, it is a constant of motion; nowadays also called "Noether charge". This can compactly be derived by means of the symmetry generator X: The condition (2.56) for a Noether symmetry becomes with (2.55)

$$\begin{aligned} 0 &= \left[\xi\, d_t L + \chi^k \frac{\partial L}{\partial q^k} + (d_t \chi^k) \frac{\partial L}{\partial \dot{q}^k} \right] + L d_t \xi - d_t \sigma \\ &= d_t \left(L\xi - \sigma) + \chi^k \frac{\partial L}{\partial q^k} + d_t (\chi^k \frac{\partial L}{\partial \dot{q}^k} \right) - \chi^k\, d_t \frac{\partial L}{\partial \dot{q}^k} \\ &= d_t \left(L\xi - \sigma + \chi^k \frac{\partial L}{\partial \dot{q}^k} \right) + \chi^k\, [L]_k \doteq d_t C. \end{aligned}$$

Notice, that the Noether charge may be void of any physical meaning. There are examples where the charge is simply a numerical constant or where the charge vanishes on-shell.

The previous expression of the conserved charge for a one-parameter group of transformations can be generalized to the case of r symmetry transformations

- Write the $(\delta_S t, \delta_S q)$ generically like

  $$\delta_\epsilon t = \epsilon^a \xi_a(t, q) \qquad\qquad \delta_\epsilon q^k = \epsilon^a \eta^k_a(t, q)$$

  with r constant parameters ϵ^a ($a = 1, ..., r$). Correspondingly there are the generators of infinitesimal symmetry transformations $X_a = \xi_a \frac{\partial}{\partial t} + \eta^k_a \frac{\partial}{\partial q^k}$. The functions η^k_a and ξ_a are not arbitrary since we assumed that the δ_S are symmetry transformations. As such they must constitute a group. In infinitesimal form this is expressed in that the commutator of two infinitesimal transformations is another infinitesimal transformation:

$$[X_a, X_b] = X_a X_b - X_b X_a = \Upsilon_{abc} X_c.$$

(In principle we must allow for further terms being proportional to the equations of motion; these are dropped here. We will reconsider them again in 3.3.4.) If the Υ_{abc} are constants, the symmetry group is an r-dimensional Lie group. The Υ_{abc} are the structure constants of the associated Lie algebra. With $X_a q^k = \eta_a^k$ and $X_a t = \xi_a$ this becomes the system of differential equations

$$(\xi_a \frac{\partial}{\partial t} \eta_b^k + \eta_a^j \frac{\partial}{\partial q^j} \eta_b^k) - (a \leftrightarrow b) = \Upsilon_{abc} \eta_c^k$$

$$(\xi_a \frac{\partial}{\partial t} \xi_b + \eta_a^j \frac{\partial}{\partial q^j} \xi_b) - (a \leftrightarrow b) = \Upsilon_{abc} \xi_c. \tag{2.57}$$

It can be proven (see e.g. [485]) that the prolongations of the infinitesimal generators obey the same algebra, that is $[\overline{X}_a, \overline{X}_b] = \Upsilon_{abc} \overline{X}_c$, and therefore (2.57) are the only conditions that need to be fulfilled.

- Also, the conserved currents J_S from (2.48) can be expanded with the infinitesimal parameters ϵ^a: Write $\sigma_S = \epsilon^a \sigma_a$ and $J = \epsilon^a C_a$. Then

$$C_a(t, q, \dot{q}) := \frac{\partial L}{\partial \dot{q} k} \chi_a^k + L \xi_a - \sigma_a \tag{2.58}$$

where the C_a constitute r conserved Noether charges. Observe, that these need not to be independent; as a matter of fact there might be less than r independent conserved charges. One can verify that C_a is an invariant of the Noether symmetry itself: $\overline{X}_a C_a = 0$.

- The symmetry variation of the coordinates can be expressed by the Noether charges

$$\bar{\delta}_\epsilon q^k = \{q^k, \epsilon^a C_a\} \tag{2.59}$$

where $\{,\}$ is the Poisson bracket. This can directly be verified by writing the conserved charge as $C_a(t, q, p) = \hat{p}_j \eta_a^j - H_C \xi_a - \sigma_a$:

$$\{q^k, \epsilon^a C_a\} = \epsilon^a \{q^k, \hat{p}_j\} \eta_a^j - \epsilon^a \{q^k, H_C\} \xi_a,$$

this indeed leading to $\bar{\delta}_\epsilon q^k = \epsilon^a (\eta_a^k - \dot{q}^k \xi_a)$ at least in the regular case for which the $\hat{p}_j$ can be identified with the canonical phase space momenta p_j, and for which the Hamilton equations of motion hold with $H(t, q, p) = H_C$.

- The commutator of two $\bar{\delta}$-transformations becomes by (2.59) and the Jacobi identity for the Poisson brackets

$$(\bar{\delta}_2\bar{\delta}_1 - \bar{\delta}_1\bar{\delta}_2)q^k = \epsilon_1^a \epsilon_2^b \Big[\{\{q^k, C_a\}, C_b\} - \{\{q^k, C_b\}, C_a\}\Big] = -\epsilon_1^a \epsilon_2^b \{\{C_a, C_b\}, q^k\}.$$

  Since we require that this is again a symmetry transformation $\bar{\delta}_3 q^k = \{q^k, \epsilon_3^c C_c\}$ we obtain

$$\{C_a, C_b\} = \Upsilon_{abc} C_c + Z_{ab}.$$

  The constants Z_{ab} are related to the central charges (a term explained in the group theory appendix.)

Example: Galilei Symmetry Group and its Noether Charges

The conservation laws derived in the previous section are attributable to Noether symmetries of Newtonian mechanics with respect to the Galilei group. A Lagrange function of the form

$$L = \frac{1}{2}\sum_{\alpha=1}^{n} m_\alpha \dot{\vec{x}}_\alpha^2 - \sum_{\alpha<\beta} V(|\vec{x}_\alpha - \vec{x}_\beta|) \tag{2.60}$$

is (quasi-)invariant with respect to the independent transformations

$\delta_\tau t = \tau$	$\delta_\tau \vec{x}_\alpha = 0$	time translations
$\delta_a t = 0$	$\delta_a \vec{x}_\alpha = \vec{a}$	space translations
$\delta_R t = 0$	$\delta_R x_\alpha^i = R^i{}_j x_\alpha^j \quad (R^i{}_j = -R^j{}_i)$	space rotation
$\delta_v t = 0$	$\delta_v x_\alpha^i = v^i t$	Galilei boost

where $(\tau, \vec{a}, R, \vec{v})$ are ten infinitesimal constants corresponding to the ϵ_a above. The generators corresponding to these infinitesimal transformations are

$$H = \partial_t \tag{2.61}$$
$$T_i = \partial_i \tag{2.62}$$
$$M_{ij} = x^i \partial_j - x^j \partial_i \tag{2.63}$$
$$G_{0i} = t\partial_i. \tag{2.64}$$

These form the Galilei algebra (2.75a, 2.75b, 2.75c). You may convince yourself that the Lagrangian (2.60) is invariant with respect to time translations, space translations and rotations. This means that the surface terms $\sigma_\tau, \sigma_a, \sigma_R$ in (2.58) are zero. The Lagrangian is only quasi-invariant with respect to Galileian boosts, namely

$$\delta_v L = \sum m_\alpha\, \delta_v \dot{\vec{x}}_\alpha \cdot \dot{\vec{x}}_\alpha = \sum m_\alpha\, \vec{v} \cdot \dot{\vec{x}}_\alpha$$

such that $\sigma_v = m_\alpha \vec{v} \cdot \vec{x}_\alpha$. The (quasi)-invariances of the Lagrangian (2.60) lead to the following conserved objects:

$$\tau(\sum \vec{p}_\alpha \cdot \dot{\vec{x}}_\alpha - L) = \tau E \tag{2.65a}$$

$$\vec{a} \cdot \sum \vec{p}_\alpha = \vec{a} \cdot \vec{P} \tag{2.65b}$$

$$R \sum \vec{x}_\alpha \times \vec{p}_\alpha = R^i{}_k J^k \tag{2.65c}$$

$$\vec{v} \cdot (\sum m_\alpha \vec{x}_\alpha - \vec{p}_\alpha t) = \vec{v} \cdot \vec{Q}. \tag{2.65d}$$

Here we identify the conserved energy $C_\tau = E$, the conserved (total) linear momentum $\vec{C}_a = \vec{P}$ and angular momentum $\vec{C}_R = \vec{J}$ in the first three relations. The conserved $\vec{C}_v = \vec{Q}$ in the last expression is related to the center of mass $\vec{R}$ and the total momentum $\vec{P}$ as in (2.41). One should be aware that the existence of the ten "standard" conservation laws (2.65a, 2.65b, 2.65c, 2.65d) is tightly bound to the form of the Lagrangian (2.60). Specifically it is essential that the forces on any of the particles depend on the mutual distances $r_{\alpha\beta} = |\vec{x}^i_\alpha - \vec{x}^i_\beta|$.

Covariance with respect to Galilei transformations is not only a characteristic of non-relativistic classical mechanics but also of non-relativistic classical field theory (e.g. fluid dynamics) and of non-relativistic quantum mechanics [446]. More about the latter is dealt with in Subsect. 4.3.4.

You may verify that the equations of motion following from the Lagrange function (2.60) are form-invariant (covariant) with respect to Galilei transformations. But, in general there are more symmetries in the equations of motion than in the action; more about this later when we will consider specifically the Kepler problem.

2.2.3 Symmetry and Canonical Transformations

The relation (2.51), valid for regular systems, is an identity in the tangent-bundle. It can be written as a cotangent-bundle expression by defining $G(q, p(q, \dot{q}), t) := J_S(q, \dot{q}, t)$, namely

$$\bar{\delta}_S q^k = \bar{W}^{ki} \frac{\partial G(q, p(q, \dot{q}))}{\partial \dot{q}^i} = \bar{W}^{ki} \frac{\partial p_j}{\partial \dot{q}^i} \frac{\partial G}{\partial p_j} = \frac{\partial G}{\partial p_k} = \{q^k, G\}.$$

The variation of the momenta are calculated as

$$\bar{\delta}_S \hat{p}_k = \bar{\delta}_S \left(\frac{\partial L}{\partial \dot{q}^k} \right) = V_{kj} \delta q^j + W_{kj} \delta \dot{q}^j = -\frac{\partial G}{\partial q^k} - [L]_j W_{ki} \frac{\partial^2 G}{\partial p_i \partial p_j} \doteq \{p_k, G\}.$$

Therefore (at least for regular systems) the Noether symmetry transformations can on-shell be written as infinitesimal canonical transformations. Off-shell it reproduces the canonical transformation of the momenta if

$$W_{ki} \frac{\partial^2 G}{\partial p_i \partial p_j} = \frac{\partial(\bar{\delta}_S q^j)}{\partial \dot{q}^k} = 0.$$

The interrelation of symmetry transformations and canonical transformations is visible also from the following line reasoning: In subsection 2.1.3, we derived the variation of an arbitrary phase-space function under an infinitesimal canonical transformation $\hat{x}^\alpha = x^\alpha + \delta_g x^\alpha$ in the form of (2.20):

$$\delta_g A(x) = \{A, g\}. \tag{2.66}$$

By this, A is invariant under g-transformations if $\{A, g\} = 0$. In the special case where A is the Hamiltonian we derive

$$\delta_g H = 0 \quad \leftrightarrow \quad \{H, g\} = 0 \quad \leftrightarrow \quad \frac{d}{dt} g \doteq 0,$$

where the last double arrow holds if g does not depend on time explicitly. Thus: Every infinitesimal canonical generator that has a vanishing Poisson bracket with the Hamiltonian leaves it invariant and is a conserved quantity. Expressed in another way, the generating function of an infinitesimal canonical transformation is a conserved quantity for those systems, whose Hamilton function is invariant with respect to this transformation.

This observation allows us to derive the ten conserved quantities in classical mechanics from homogeneity and isotropy of the space-time continuum:

- The momentum components p_x, p_y, p_z are canonically conjugate to the x, y, z-components of the space coordinates and therefore the generating functions of infinitesimal spatial translations in the x, y, z-directions. The total momentum is thus conserved for those systems whose Hamilton function is invariant with respect to infinitesimal spatial translations.
- The angular momentum components J_x, J_y, J_z are canonically conjugate to the rotation angles $\alpha_x, \alpha_y, \alpha_z$ and therefore the generating functions of infinitesimal rotations around the x, y, z-axis. The total angular momentum is thus conserved for those systems whose Hamilton function is invariant with respect to infinitesimal rotations.
- The object $\vec{G} = \sum_\alpha (m_\alpha \vec{x}_\alpha - \vec{p}_\alpha t)$ is the generator of infinitesimal Galilei-boosts

  $$\hat{\vec{x}}_\alpha = \vec{x}_\alpha + \vec{v}t \qquad \hat{\vec{p}}_\alpha = \vec{p}_\alpha + m_\alpha \vec{v},$$

  with an infinitesimal velocity $\vec{v}$. If the Hamilton function is invariant with respect to infinitesimal Galilei-boosts, $\vec{G}$ is conserved in time, implying together with the conservation of the total momentum that the center of mass moves uniformly.
- The Hamilton function H is canonically conjugate to the time variable t and therefore the generating function of infinitesimal time translations $\hat{t} = t + \tau$. The total energy is conserved for those systems whose Hamilton function is invariant with respect to time translations. And this is the case if the Hamilton function does not depend on time explicitly.

The conserved quantities obey an algebra: If together with g also g' is a Noether charge, the Jacobi identity

$$\big\{\{H,g\},g'\big\}+\big\{\{g',H\},g\big\}+\big\{\{g,g'\},H\big\}=0$$

reveals that $\{g,g'\}$ is a Noether charge too. If there are finitely many independent conserved quantities, each commutator can be written as a conserved quantity again. In case of classical mechanics, this leads again to the algebra of the Galilei group.

2.2.4 Conservation Laws and Symmetries

Conservation Laws and Symmetries in Which Sense?

This Sect. 2.2. carries the title "Symmetries and Conservation Laws", and indeed as derived previously, the Noether symmetries lead to conserved quantities, the Noether charges. To observe that conserved quantities follow from symmetries of an action is quite intriguing. But it also raises many questions such as: Does a symmetry necessarily imply the existence of a conserved quantity? Is there a way to find (all) symmetries of the action? Can one find all conservation laws by Noether's theorem? Is every conservation law a consequence of a symmetry? As we will see, posed this way, the questions cannot really be answered–or put another way–we need to be more precise in questioning[12]. The Noether theorem itself only tells that for an r-parameter Lie group, r linear combinations of Euler-derivatives are expressible as total time-derivatives (or–in field theories–as divergences).

- The mere existence of a symmetry is not at all sufficient to establish a conservation law. Take for instance the equation of motion for a particle under the influence of a frictional force
$$m\ddot{\vec{x}}=-k\dot{\vec{x}}.$$
This equation is invariant under space and time translations, but neither the momentum $m\dot{\vec{x}}$ nor the energy $\frac{m}{2}\dot{x}^2$ are conserved. This is not a counter-example to Noether's theorem, since we are arguing with the equations of motion and thus with Lie symmetries. In any case, the equations of motion can be derived from the Lagrangian
$$L=\frac{1}{2}m\dot{\vec{x}}^2\exp\left[\left(\frac{k}{m}\right)t\right].$$
This Lagrangian is covariant with respect to space translations, and–in agreement with Noether's first theorem–indeed the generalized momentum $\vec{p}=\partial L/\partial\dot{\vec{x}}$ is

[12] E. Wigner warned against a "facile identification" of symmetry and conservation principles [557] and P. Havas wrote a "Folklore, Fiction, and Fact" article on this topic [255].

on-shell conserved. And although the Lagrangian is not form invariant with respect to time translations, there is a conserved quantity

$$E = \frac{\vec{p}^2}{2m} = \frac{1}{2} m \dot{\vec{x}}^2 \exp\left[2\left(\frac{k}{m}\right)t\right]$$

which, however, is not the numerical value of the Hamiltonian, although for $k = 0$ it becomes identical to the "usual" energy. For a more elaborate analysis of this example see [255].

- When asking for all symmetries of either the action or the equations of motion we need to specify whether we have in mind symmetries with respect to point transformations $\delta q^k(q, t)$, $\delta t(t, q)$ only, or whether more general transformations are to be considered.
- Even if all variational symmetries of a Lagrangian were known, does the Noether theorem guarantee that all conservation laws are found? Not at all: For a system with N degrees of freedom there are N equations of motion, that is N second order differential equations for functions $q^i(t)$. In their solutions there are $2N$ free constants, which can be determined by the initial values of the coordinates and the velocities. If the system exhibits symmetry under Galilei transformations, at most 10 of these constants may be expressed through the standard Noether charges (2.65a, 2.65b, 2.65c, 2.65d). The authors of [332] emphasize that the distinctive role of the Galilei symmetry group induced constants of motion may be seen in that these do have the important property of being additive.
- It also may happen that the conserved charges are algebraically related, although being due to different symmetry transformations (as will be shown later on examples). In still another case a conservation law might be identically fulfilled; sometimes called a "strong" conservation law in order to distinguish it from a "weak" conservation law being valid on the solutions of the dynamical equations only.
- There are examples where two different Lagrangians–not related by a total derivative–lead to the same equations-of-motion, and where two completely different Noether symmetry transformations lead to the same conserved charge. Take the standard Lagrangian for the two-dimensional oscillator

$$L = \frac{1}{2}\left[(\dot{q}_1^2 + \dot{q}_2^2) - \omega^2(q_1^2 + q_2^2)\right]$$

which is invariant under rotations, giving rise to the conservation of angular momentum $q_1\dot{q}_2 - q_2\dot{q}_1$. The same dynamics is derived from the Lagrangian

$$L' = \dot{q}_1\dot{q}_2 - \omega^2 q_1 q_2.$$

This Lagrangian is not invariant under rotations, but instead under $(q_1, q_2) \rightarrow (e^{\alpha}q_1, e^{-\alpha}q_2)$, from which again the conservation of angular momentum results.

- The "inverse Noether theorem" deals with the question under which circumstances the existence of a conserved quantity relates to a Noether symmetry[13]. The problem of the inverse Noether theorem relates to generalization of Noether's theorem to velocity-dependent transformations [463]. Again, the Hessian (2.7) plays a crucial role here. It was proven [67], that if $C(q, \dot{q}, t)$ is a constant of motion and if the Hessian W^{ij} is invertible, the infinitesimal transformation associated with C, namely
$$\bar{\delta} q^i = \epsilon \bar{W}^{ij} \frac{\partial C}{\partial \dot{q}^j} \qquad\qquad \delta t = L^{-1} \left[\epsilon C - \frac{\partial L}{\partial \dot{q}^j} \bar{\delta} q^j \right]$$
is a symmetry transformation for the Lagrangian L. Assume, for example that the energy $E = (\partial L / \partial \dot{q}^j) \dot{q}^j - L$ is conserved. Then the previous defining equations result in $\bar{\delta} q^i = \epsilon \dot{q}^i$ and $\delta t = -\epsilon$ (and thus $\delta q^i = 0$).

Lie Symmetries and Noether Symmetries

The previous observations about the (non)-relation between conservation laws and symmetry transformations seems to leave us behind in a hard-to-reach landscape. However, as it turns out, the investigation of symmetries of differential equations is more feasible than the investigation of the Lagrangian from which they are possibly derived as Euler-Lagrange equations; see the comprehensive [399]. Remember that Lie symmetries are symmetries of the equations of motion in the sense that if $q^k(t)$ is a solution, the transformed

$$\hat{q}^k(\hat{t}) = q^k(t) + \epsilon \eta^k(t, q) + \mathcal{O}(\epsilon^2) \qquad\qquad \hat{t} = t + \epsilon \xi(t, q) + \mathcal{O}(\epsilon^2)$$

is a solution, too ("mapping solutions to solutions"). Although the following expressions can be defined for a system of differential equations $\Delta^\beta(x_1, ..x_p; u^\alpha(x), u^\alpha_{,i}(x), ..., u^\alpha_{,i_1 \dots i_q}(x)) = 0$ with p independent variables and derivatives of the dependent variables $u^\alpha(x)$ of arbitrary order, I will take the example of classical mechanics with one independent variable t and functions q^k appearing up to the second derivative, that is
$$\Delta^k(t; q, \dot{q}, \ddot{q}) \equiv \Delta^k(\hat{t}; \hat{q}, \hat{\dot{q}}, \hat{\ddot{q}}) = 0.$$

Then we can directly use the operator $\overline{X}$ as given by (2.53) or alternatively by (2.55) to define a Lie symmetry group by transformations which obey

$$\overline{X} \Delta^k{}_{|\Delta^j = 0} = 0. \tag{2.67}$$

All independent vector fields that fulfill this relation are point symmetries of the equations of motion. Furthermore, one is dealing with a Noether-Bessel-Hagen point

[13] Do not get confused here: Noether herself proved that her theorems do have a converse in the sense that the existence of r Euler derivatives which are divergences implies the invariance of an action.

symmetry associated to a Lagrangian L, if a "boundary term" $\sigma(t, q)$ exists such that

$$\overline{X}L + L d_t \xi = d_t \sigma. \tag{2.68}$$

Thus, if one knows all Lie symmetries (with respect to a class of transformations) admitted by the Euler-Lagrange equations then one can find all variational symmetries by checking which of these Lie symmetries are symmetries of the action.

One-Dimensional Free Particle and Harmonic Oscillator

Rather unexpected results do follow from the previous considerations for the most simple system in classical mechanics, namely the one-dimensional free particle and the harmonic oscillator

$$L = \frac{1}{2}\dot{q}^2 - \frac{\omega^2}{2}q^2 \qquad\qquad \Delta = \ddot{q} + \omega^2 q = 0$$

(after an appropriate rescaling such that $m = 1$). In this case, after a straightforward calculation the condition (2.67) becomes explicitly

$$\eta_{tt} + (2\eta_{tq} - \xi_{tt})\dot{q} + (\eta_{qq} - 2\xi_{qt})\dot{q}^2 - \xi_{qq}\dot{q}^3 + \omega^2\eta - \omega^2 q(\eta_q - 2\xi_t - 3\xi_q\dot{q}) = 0.$$

Here the indices on η and ξ denote derivatives with respect to t and q. The previous condition contains terms proportional to $\dot{q}^n$ $(n = 0, ..., 3)$. These must vanish separately, and thus we get a system of differential equations for η and ξ. These are called the 'defining equations' for the symmetry generators.

For the free particle with $\omega = 0$ the most general solution of the defining equations result in

$$\xi(t, q) = \alpha_0 + \alpha_1 t + b_1 t^2 + \beta_0 q + \beta_1 t q, \qquad \eta(t, q) = a_0 + a_1 q + \beta_1 q^2 + b_0 t + b_1 q t$$

with eight constants $\alpha_i, \beta_i, a_i, b_i$. Thus, for the one-dimensional free particle equation there are eight independent point transformations defined by the vector fields

$$\begin{aligned}
X_1 &= \frac{\partial}{\partial t} & X_2 &= t\frac{\partial}{\partial t} & X_3 &= q\frac{\partial}{\partial t} \\
X_4 &= t^2\frac{\partial}{\partial t} + tq\frac{\partial}{\partial q} & X_5 &= qt\frac{\partial}{\partial t} + q^2\frac{\partial}{\partial q} & & \\
X_6 &= \frac{\partial}{\partial q} & X_7 &= q\frac{\partial}{\partial q} & X_8 &= t\frac{\partial}{\partial q}.
\end{aligned}$$

Although of course this example is not at all relevant for fundamental physics it is quite instructive to immediately interpret the effect of the vector fields in terms of finite transformations. These can of course be found by using the exponential map:

$$\hat{q} = e^{aX} \cdot q \qquad\qquad \hat{t} = e^{aX} \cdot t.$$

Sometimes there is an easier procedure: Remember that the components of the vector fields are defined as

$$\frac{d\hat{q}}{d\epsilon}\Big|_{\epsilon=0} = q + \epsilon\eta + \epsilon^2 \cdots = \eta \qquad\qquad \frac{d\hat{t}}{d\epsilon}\Big|_{\epsilon=0} = t + \epsilon\xi + \epsilon^2 \cdots = \xi.$$

Therefore–knowing ξ and η from the vector fields X, we need to solve the first order equations

$$\frac{d\hat{q}}{da} = \eta(\hat{t}(a), \hat{q}(a)) \qquad\qquad \frac{d\hat{t}}{da} = \xi(\hat{t}(a), \hat{q}(a))$$

with initial values $\hat{q} = q$ and $\hat{t} = t$ for $a = 0$. For the generators X_1 and X_6 the finite transformations are easily found as

$$T_1 : \hat{q} = q; \quad \hat{t} = t + a \qquad\qquad T_6 : \hat{q} = q + a; \quad \hat{t} = t,$$

that is constant translations of t and q, respectively[14]. For X_2 and X_7 the transformations are rescalings:

$$T_2 : \hat{q} = q; \quad \hat{t} = e^a t \qquad\qquad T_7 : \hat{q} = e^a q; \quad \hat{t} = t.$$

The finite transformations corresponding to X_3 and X_8 are

$$T_3 : \hat{q} = q; \quad \hat{t} = t + aq \qquad\qquad T_8 : \hat{q} = q + at; \quad \hat{t} = t$$

which we might call Galilei boosts (although only the latter is a genuine one). More tricky are the integrations of the differential equations implicated by X_4 and X_5. Let us instead exponentiate the infinitesimal transformations

$$\begin{aligned}
\hat{q} = e^{aX_4} \cdot q &= \left[1 + atq\partial_q + \frac{1}{2}(atq\partial_q)(atq\partial_q) + ...\right] q \\
&= q + atq + ... + (at)^k q + ... = \frac{q}{1 - at} \\
\hat{t} = e^{aX_4} \cdot t &= \left[1 + at^2\partial_t + \frac{1}{2}(at^2\partial_t)(at^2\partial_t) + ...\right] t \\
&= t + at^2 + ... + a^k t^{k+1} + ... = \frac{t}{1 - at}.
\end{aligned}$$

---

[14] Of course every generator has its own constant a. Thus, for example the constant in T_1 has the dimension of time ($a = t_0$) and for T_6 it has dimension length ($a = q_0$).

As a result we observe that X_4 and X_5 generate the specific projective transformations

$$T_4 : \hat{q} = \frac{q}{1-at}; \quad \hat{t} = \frac{t}{1-at} \qquad T_5 : \hat{q} = \frac{q}{1-aq}; \quad \hat{t} = \frac{t}{1-aq}.$$

It can be shown that the generators $X_1, ...X_8$ obey an algebra which is isomorphic to $sl(3, \mathbb{R})$ [357]. It can also be shown that the finite transformations are the projective transformations

$$\hat{t} = \frac{At + Bq + C}{Lt + Mq + N} \qquad \hat{q} = \frac{Dt + Eq + F}{Lt + Mq + N}.$$

These are all those transformations that map a straight line in a plane into a straight line. Further it can be proven that a Lie group of order eight is the largest symmetry group for point transformations for second order equations of motion [357].

Which of the generators above are Noether symmetries? In calculating according to (2.68) the expressions $X_\alpha L + LD\xi_\alpha$ one finds that X_1, X_4, X_6, and X_8 are Noether symmetries with $\sigma_\gamma = \{0, \frac{1}{2}q^2, 0, q\}$. Further the linear combination $X_+ = 2X_2 + X_7$ is found to be a generator of a Noether symmetry. Thus there are five independent Noether symmetry transformations for the free particle in one-dimension. This should come as a surprise since–knowing about the Galilei group–we could expect only three symmetries related to time translation, 1D space translations and 1D Galilei boosts. In [357] it is shown, that five is the maximal number of independent generators for the variational symmetries of a one-dimensional system. Which are the conserved charges according to (2.58)? One finds for instance $C_6 = \dot{q}$, $C_8 = (\dot{q}t - q)$, which are indeed obviously on-shell conserved. The other charges depend algebraically on these: $C_1 = -{}^1\!/\!_2(C_6)^2$, $C_4 = -{}^1\!/\!_2(C_8)^2$, $C_+ = -C_6C_8$. We should be surprised to get more than two independent constants of motion since the solution $q(t) = v(t - t_0)$ has two constants; we find $C_6 = v$, $C_8 = vt_0$. Let me remark here that already Noether observed that the action of the generator X_8 leading to a quasi-invariance can be realized by the generator

$$\tilde{X}_8 = -2\frac{q}{\dot{q}^2}\frac{\partial}{\partial t} + \left(t - 2\frac{q}{\dot{q}}\right)\frac{\partial}{\partial q}.$$

This is no longer a point transformation, since now $\tilde{\xi} = \tilde{\xi}(t, q, \dot{q})$, $\tilde{\eta} = \tilde{\eta}(t, q, \dot{q})$ depend on velocities. In [285] you find a constructive proof that any quasi-invariant point transformation is equivalent to a velocity-dependent transformation that leaves the Lagrangian strictly invariant.

If you track the previous considerations for the harmonic oscillator, you will find that there are seven independent symmetry generators for the equations of motion, five Noether symmetries, and–not a surprise–two independent constants of motion, for instance

$$E = \frac{1}{2}(\dot{q}^2 + \omega^2 q^2) \doteq \frac{1}{2} k Q^2 \qquad G = q\omega \cos\omega t - \dot{q}\sin\omega t \doteq \sin\omega t_0\, \omega Q$$

given the solution $q(t) = Q\sin\omega(t+t_0)$ with two constants Q and t_0. Without going into the full calculations observe that energy conservation can be traced to two different transformations: The Lagrangian is invariant with respect to $\delta_\tau\{q,t\} = \{0,\tau\}$ and thus the energy E is conserved with $\dot{E} = -\dot{q}\,[L]_q$. However, the Lagrangian is also quasi-invariant with respect to $\delta_\epsilon\{q,t\} = \{\epsilon\sin\omega t, 0\}$ with $\sigma = \omega q\cos\omega t$. Therefore the Noether theorem delivers the constant of motion G with $\dot{G} = \sin\omega t\ [L]_q \doteq 0$. Thus, although the two constants are simply related as $E \propto G^2$, the respective symmetry transformations are completely unrelated.

Free Point Particle in Three Dimensions

We know that the action of the free point particle is invariant under transformations of the 10-dimensional Galilei group. But is this the maximal symmetry group? To find this maximal group we first determine the Lie point symmetries, and then check which of these are Noether point symmetries. (In [300] these are straightforwardly derived from the quest that $\int dt(\frac{dq^i}{dt})^2 = \int d\hat{t}(\frac{d\hat{q}^i}{d\hat{t}})^2$.)

The defining equations for the vector field components of $X = \xi\partial_t + \eta^i\partial_i$ are found to be

$$\xi_{kj} = 0 \qquad \eta^i_{jk} = \delta^i_j\,\xi_{kt} + \delta^i_k\,\xi_{jt} \qquad 2\eta^i_{tk} = \delta^i_k \xi_{tt} \qquad \eta^i_{tt} = 0.$$

The most general solution contains 24 parameter. The 24 independent vector fields are

$$\partial_t \qquad t\partial_t \qquad t^2\partial_t + tq^i\partial_i \qquad q^i\partial_t \qquad tq^j\partial_t + q^jq^i\partial_i \qquad \partial_i \qquad t\partial_i \qquad q^K\partial_I.$$

Again it turns out that only a subset of these generators are generators of variational symmetries, namely the 12 generators

$$\begin{aligned} &H = \partial_t \qquad && T_i = \partial_i \qquad && L_k = \epsilon_{kij} q^i \partial_j \\ &G_i = t\partial_i \qquad && S = 2t\partial_t + q^i \qquad && \partial_i C = t^2\partial_t + tq^i\partial_i. \end{aligned}$$

One verifies that $\{H, T_i, L_i\}$ fulfill the Euclid algebra

$$\begin{aligned} &[H, T_i] = 0 \qquad && [H, J_i] = 0 \qquad && [H, T_i] = 0 \\ &\left[T_i, T_j\right] = 0 \qquad && \left[T_i, L_j\right] = \epsilon_{ijk} T_k \qquad && \left[L_i, L_j\right] = \epsilon_{ijk} L_k. \end{aligned}$$

Further $\{H, T_i, L_i, G_i\}$ fulfill the Galilei algebra with

$$[H, G_i] = T_i \qquad \left[T_i, G_j\right] = 0 \qquad \left[G_i, L_j\right] = \epsilon_{ijk} G_k \qquad \left[G_i, G_j\right] = 0$$

and the two additional generators S and C obey

$$\begin{array}{lllll} & [H,S]=2H & [T_i,S]=T_i & \left[L_i,S\right]=0 & \left[G_i,S\right]=-G_i \\ [H,C]=S & \left[T_i,C\right]=G_i & \left[L_i,C\right]=0 & \left[G_i,C\right]=0 & [S,C]=2C. \end{array}$$

This algebra is also known as Schrödinger algebra $\mathfrak{sch}(1,3)$ (without central extension), since prior to its discovery for the classical free particle it was found for the free Schrödinger equation. The Schrödinger algebra plays a similar role in non-relativistic physics as the conformal algebra plays in relativistic physics; see also Sect. 4.3.4.

The Lagrangian $L = \frac{m}{2}(\dot{q}^i)^2$ is invariant with respect to the transformations generated by H, T_i, L_k, S. It is quasi-invariant for G_i (with $\Sigma_{G_i} = mq^i$) and for C (with $\Sigma_C = \frac{m}{2}q^2$). Which are the charges? The

$$C_\alpha = m\dot{q}_i\eta^i_\alpha - \frac{1}{2}m\xi_\alpha q^2 - \Sigma_\alpha$$

become explicitly

$$\begin{aligned} &C_H = -\frac{m}{2}\dot{q}^2 =: -E \qquad C_{T_i} = m\dot{q}_i =: P_i \qquad C_{L_i} = m\epsilon_{kli}q^l\dot{q}^i =: J_i \\ &C_{G_i} = m(t\dot{q}_i - q_i) =: Q_i \\ &C_S = m\dot{q}_i(q_i - t\dot{q}_i) = D \qquad C_C = \frac{m}{2}(t\dot{q}_i - q_i)(t\dot{q}^i - q^i) = R. \end{aligned}$$

These 12 charges are not independent but can all be expressed algebraically by P_i and Q_i:

$$E = \frac{1}{2m}P^2 \qquad J_i = \frac{1}{m}\epsilon_{ijk}P_jQ_k \qquad D = -\frac{1}{m}P_iQ^i \qquad R = \frac{1}{2m}Q^2.$$

The very fact, that there are only six independent constants of motion agrees with the fact that the solution of the system of differential equations is $q^i(t) = v^it + q^i_0$ which has six integration constants v^i and q^i_0. The solution can as well be expressed by the six conserved charges as $q^i(t) = \frac{1}{m}(P^it - Q^i)$.

The finite transformations for the 12-parameter symmetry group $\mathbf{G_{12}}$ are

$$\hat{q}^i = \frac{R^{ij}q^j + a^i + v^it}{\gamma t + \delta} \qquad \hat{t} = \frac{\alpha t + \beta}{\gamma t + \delta} \qquad \text{with } \alpha\delta - \beta\gamma = 1, \qquad R^TR = 1.$$

There are two special cases, namely (i) $\beta = 0 = \gamma$, $\alpha = 1 = \delta$ with group transformations

$$g: \qquad \hat{\vec{q}} = R\vec{q} + \vec{a} + \vec{v}t \qquad \hat{t} = t$$

constituting the 9-parameter static Galilei group $\mathbf{G_9}$, and (ii) $\vec{a} = 0 = \vec{v}$, $R = 1$ with

$$\sigma : \qquad \hat{\vec{q}} = \frac{\vec{q}}{\gamma t + \delta} \qquad\qquad \hat{t} = \frac{\alpha t + \beta}{\gamma t + \delta} \qquad\qquad \text{with} \qquad \alpha\delta - \beta\gamma = 1$$

which is the 3-parameter group $\mathbf{SL(2, R)}$ containing time translations ($\gamma = 0$, $\alpha = 1 = \delta$), scale transformations ($\beta = 0 = \gamma$) and so-called expansions ($\beta = 0$, $\alpha = 1 = \delta$). From the group composition one finds ([300]) that $\mathbf{G_9}$ is an invariant subgroup of the full group, and that $\mathbf{G_{12}} = \mathbf{G_9} \ltimes \mathbf{SL(2, R)}$.

The Kepler Problem

Given that the explanation of Kepler's three laws by Newton was the very success story of classical mechanics, we dare to ask how this is related to both Lie and Noether symmetries. The two-body Kepler problem is defined by the Lagrangian

$$L = \frac{1}{2} M \dot{\vec{x}}^2 + \frac{\alpha}{|\vec{x}|}$$

where $\vec{x} = \vec{x}_1 - \vec{x}_2$ and $M = \frac{m_1 m_2}{m_1 + m_2}$ is the reduced mass (see e.g. [332], sect.13). For the three degrees of freedom q^k, the equations of motion are

$$\Delta^k(q, \ddot{q}) = M\ddot{q}^k + \frac{\alpha q^k}{r^3} = 0 \qquad\qquad \text{with} \qquad\qquad r^2 = \sum_{k=1}^{3} q^k q^k.$$

The Lie point symmetry transformations, found from determining the vector field $\overline{X}$ for which on-shell $\overline{X}\Delta^k = 0$, are

$$\overline{X}_k = \epsilon_{kij}\, q^j \frac{\partial}{\partial q^i} + \epsilon_{kij} \dot{q}^j \frac{\partial}{\partial \dot{q}^i} \tag{2.69a}$$

$$\overline{X}_4 = \frac{\partial}{\partial t} \tag{2.69b}$$

$$\overline{X}_5 = t\frac{\partial}{\partial t} + \frac{2}{3} q^j \frac{\partial}{\partial q^j} - \frac{1}{3}\dot{q}^j \frac{\partial}{\partial \dot{q}^j}. \tag{2.69c}$$

The three generators X_k are recognized as those of the three-dimensional rotation group; indeed r^2 is an invariant under these generators: $X_k r^2 = 0$. The generator X_4 corresponds to time translations. But what is the meaning of X_5? We verify that this generator leaves invariant the quantity $s = t^2/r^3$:

$$X_5\left(\frac{t^2}{r^3}\right) = \frac{1}{r^3} t \frac{\partial t^2}{\partial t} + t\frac{2}{3} q^j \frac{\partial}{\partial q^j} \frac{1}{r^3} = \frac{2t^2}{r^3} - \frac{2t^2}{3} q^j \frac{3}{r^4} \frac{\partial r}{\partial q^j} = 0.$$

This reflects a scale invariance: If $t \to \lambda t$, $q^k \to \lambda^{2/3} q^k$ the quantity s is left invariant. And indeed with $\lambda = (1+\epsilon)$

$$\hat{q}^k = \lambda^{2/3} q^k = (1+\epsilon)^{2/3} q^k \sim q^k + \frac{2}{3}\epsilon q^k \qquad\qquad \hat{t} = (1+\epsilon)t = t + \epsilon t$$

we confirm (2.69c). Although s is an invariant of X_5, it is not a generator of a Noether symmetry. One finds $X_5 L + D\xi_5 = (1/3)L$, and this cannot be written as a term $D\sigma$. The other four generators (2.69a, 2.69b) describe Noether symmetries. Their conserved charges are the angular momentum and the energy:

$$J_k = M\epsilon_{kij} q^i \dot{q}^j \qquad E = \frac{M}{2}\dot{q}^2 - \frac{\alpha}{r}.$$

Now, you may be aware that the Kepler problem has a further conserved quantity, known as the Laplace-Runge-Lenz vector. For a long time it was unknown whether this can be derived from a Noether symmetry[15]. Indeed this is possible if one goes beyond point transformations. I remarked already that the Noether-Bessel-Hagen identity (2.56) holds even for the case that the ξ and η^k depend on velocities $\dot{q}$. In fact, this identity now splits into two identities, namely one which contains second derivatives $\ddot{q}$, and the rest. The second derivatives appear linearly, and thus the first part yields three identities

$$L\frac{\partial \xi}{\partial \dot{q}^j} + \frac{\partial L}{\partial \dot{q}^k}\left(\frac{\partial \eta^k}{\partial \dot{q}^j} - \dot{q}^k \frac{\partial \xi}{\partial \dot{q}^j}\right) = \frac{\partial \sigma}{\partial \dot{q}^j}.$$

In case of the Kepler problem, it turns out that transformations with $\xi = 0$ and

$$\eta^{(i)k}(q, \dot{q}) = 2q^i \dot{q}^k - q^k \dot{q}^i - \delta^{ki}(q^j \dot{q}^j)$$

render the Lagrangian quasi-invariant. The Noether charges belonging to this variational symmetry are found to be

$$A^k = -M^2\left[\dot{q}^2 q^k - (q^j \dot{q}^j)\dot{q}^k\right] + M\alpha \frac{q^k}{r}.$$

These are the components of the Laplace-Runge-Lenz vector $\vec{A} = -\vec{p} \times \vec{J} + M\alpha\,\vec{x}/|\vec{x}|$. Its conservation has as the consequence that the two masses revolve on elliptic orbits (first Kepler law). The conservation of angular momenta can be expressed as the second Kepler law, and the invariance of $s = t^2/r^3$ amounts to Kepler's third law. In conclusion, Kepler's laws do have their origin in variational and in Lie symmetries.

Let me point out a peculiarity here, which makes one to understand, that variational symmetries need not to be Noether symmetries. With regard to a compact notation let me modify the generator X_4 to $H := X_4 - D$ (remember that adding to an infinitesimal symmetry generator a multiple of the operator D is again a symmetry

---

[15] This symmetry is in the literature also designated as a "hidden" or a "dynamical" symmetry.

generator). Then the algebra of the seven Noether generators $\{X_k, H, Y^i = \eta^{(i)k} \frac{\partial}{\partial q^k}\}$ becomes

$$\begin{aligned}
&\left[X_i, X_j\right] = \epsilon_{ijk} X_k && \left[X_i, Y^k\right] = \epsilon^k{}_{ij} Y^j \quad \left[X_j, H\right] = 0 \\
&\left[Y^i, Y^j\right] = -2E\, \epsilon^{ijk} X_k + 2J_k\, \epsilon^{ijk} H && \left[H, Y^j\right] = 0.
\end{aligned}$$

The generators form an algebra, but not a Lie algebra because the structure coefficients in $\left[Y^i, Y^j\right]$ are not numerical constants.

2.2.5 *Noether–Geometrically

In this subsection, the geometrical description of the first Noether theorem will be given under the simplifying assumption that there is no explicit time dependence[16]. Thus we will deal with strict invariance of a Lagrangian and exclude quasi-invariance from the considerations.

The Noether Theorem in Lagrangian Form

Point transformations $q^k \to \hat{q}^k(q)$ are diffeomorphism $\mathbb{Q} \to \mathbb{Q}$ inducing transformations

$$\dot{q}^k \to \frac{d}{dt}\hat{q}^k(q) = \frac{\partial \hat{q}^k}{\partial q^j}\dot{q}^k.$$

Therefore, in general, an infinitesimal point transformation is represented by a vector field $X \in \mathfrak{X}(T\mathbb{Q})$ of the form

$$X = X^i(q)\frac{\partial}{\partial q^i} + \left(\frac{d}{dt}X^i(q)\right)\frac{\partial}{\partial \dot{q}^i}. \tag{2.70}$$

If the Lagrangian is invariant under a transformation mediated by this vector field (that is $\delta q^i = X^i$), we have $\delta L := \pounds_X L = 0$; then, in coordinates

$$\begin{aligned}
0 = \pounds_X L = X^i \frac{\partial L}{\partial q^i} + \left(\frac{d}{dt}X^i\right)\frac{\partial L}{\partial \dot{q}^i} &= X^i[L]_i + X^i \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} \\
+ \left(\frac{d}{dt}X^i\right)\frac{\partial L}{\partial \dot{q}^i} &= X^i[L]_i + \frac{d}{dt}\left(X^i \frac{\partial L}{\partial \dot{q}^i}\right)
\end{aligned} \tag{2.71}$$

[16] As for generalizations beyond point transformation in a coordinate-independent way there is by now a rich literature beginning with A. Trautman [510]; see also [14, 192].

showing that on-shell $([L]_i = 0)$ the quantity $\Sigma := X^i(\partial L/\partial \dot{q}^i)$ is conserved. In a chart-independent way this becomes

$$\pounds_\Delta(i_X \theta_L) = 0,$$

where θ_L is the Cartan one-form defined by (2.29).

The Noether Theorem in Hamiltonian Form

The Hamiltonian Noether theorem is an immediate consequence of (2.37): Let H be invariant under the one-parameter flow Φ^g mediated by a vector field X_g according to (E.3):

$$0 = \pounds_{X_g} H = \{H, g\} = -\pounds_\nabla g.$$

Thus g is a constant of motion. The canonical transformation generated by X_g preserves the equations of motion since together with (2.38) $[X_g, \nabla] = X_{\{g,H\}} = 0$. The reverse is also true: If the phase space function g is a constant of motion its associated vector field X_g generates an infinitesimal symmetry. Further, if two infinitesimal symmetries are given with constants of motion (g, g'), then the commutator is also a symmetry with an associated constant of motion $\{g, g'\}$. This can be shown by (2.38) and the Jacobi identity for Poisson brackets.

Notice that in terms of geometry the Hamiltonian version of the Noether theorem is more straightforward than in the Lagrangian version. (It also leads more directly into the quantum version.) If the Lagrangian is regular, the constants of motion are related by $\mathcal{F}L^*(g) = i_{X_g}\theta_L$.

2.3 Galilei Group

2.3.1 Transformations and Invariants of Classical Mechanics

Each of the following $\mathbf{G_i}$ constitutes a symmetry group:

Time translations:	$\mathbf{G}_{\boldsymbol{\varnothing}} : \hat{t} = t + \tau$	(1 parameter)
Space translations:	$\mathbf{G}_{\mathbf{a}} : \hat{\vec{x}} = \vec{x} + \vec{a}$	(3 parameter)
Rotations:	$\mathbf{G}_R : \hat{\vec{x}} = \mathbf{R} \cdot \vec{x}$	(3 parameter since $\mathbf{R} \in \mathbf{SO(3)}$)
Galilei boosts:	$\mathbf{G}_v : \hat{\vec{x}} = \vec{x} + \vec{v}t$	(3 parameter)

Furthermore

Time reversal:	$\mathbf{G}_T = \{1, T\}$	$T : t \to -t$
	is a symmetry if	$L = a^{ik}\dot{q}_i\dot{q}_k - U(q)$.
Space inversion:	$\mathbf{G}_P = \{1, P\}$	$P : q_i \to -q_i$
	is a symmetry if	$U = U(\lvert q_i - q_k \rvert)$.

The largest symmetry group of classical mechanics is the union of all $\mathbf{G}_i$. This is the Galilei group **Gal**. The elements of this group leave distances $\Delta_x := \sqrt{|\vec{x}_i - \vec{x}_j|^2}$ and time intervals $\Delta_t := |t_i - t_j|$ invariant. Δ_x are the invariants of the three-dimensional Euclidean space E^3, Δ_t is the invariant of the line E^1. The product $A := E^1 \times E^3$ is called Aristotelian space-time. Since the Galilei boosts transform in a space "mixing" a time dependence into translations they are symmetries of a space which locally is an Aristotelian space-time, but globally a fibre bundle (with base space E^1 and fibre E^3), which constitutes Galilean space-time [410]. This exemplifies that the symmetries of classical mechanics are related to an underlying geometry. In the next chapter we will see that the symmetries of relativistic physics are also related to some specific geometry, namely Minkowski spacetime. Since $\mathbf{G}_T$ and $\mathbf{G}_P$ are discrete groups the full Galilei group invariance group is both discrete and continuous. The continuous part $\mathbf{Gal}_c$ contains the rotations, Galilei boosts, space translations and time translations. The elements of this 10-parameter Lie group can be denoted as

$$\mathbf{Gal}_c \ni g = (\tau,\ \vec{a},\ \vec{v},\ \mathbf{R})$$

with the composition law

$$g' \circ g = (\tau' + \tau,\ \vec{a}' + \mathbf{R}'\vec{a} + \vec{v}'\tau,\ \vec{v}' + \mathbf{R}'\vec{v},\ \mathbf{R}'\mathbf{R}). \tag{2.72}$$

The neutral element of $\mathbf{Gal_c}$ is $g_0 := (0, \vec{0}, \vec{0}, \mathbf{1})$. The inverse to g is

$$g^{-1} = (-\tau,\ \mathbf{R}^{-1}(\vec{v}\tau - \vec{a}),\ -\mathbf{R}^{-1}\vec{v},\ \mathbf{R}^{-1}). \tag{2.73}$$

The group $\mathbf{Gal}_c$ is non-Abelian. Observe that the group composition (2.72) and the expression for the inverse of a group element (2.73) look rather weird. As a matter of fact, the Galilei group is an awkward group, especially compared to the Poincaré group of which it is a limiting case (speed of light going to infinity)–or, in mathematical terms–a group contraction, as explained in the next chapter.

2.3.2 Structure of the Galilei Group

Some specific subgroups of $\mathbf{Gal}_c$ are

- As a set $\mathbf{Gal}_c = \{\mathbf{G}_R, \mathbf{G}_v, \mathbf{G}_a, \mathbf{G}_\tau\}$, and each $\mathbf{G}_I$ is a subgroup (with further subgroups). Of these subgroups only $\mathbf{G}_R$ is non-Abelian (generically). Further non-trivial subgroups are the sets

$$\{\mathbf{G_ø},\mathbf{G_a}\},\quad \{\mathbf{G_ø},\mathbf{G_R}\}$$
$$\{\mathbf{G_a},\mathbf{G_R}\}\quad \{\mathbf{G_a},\mathbf{G_v}\}$$
$$\{\mathbf{G_R},\mathbf{G_v}\}$$
$$\{\mathbf{G_ø},\mathbf{G_a},\mathbf{G_R}\},\quad \{\mathbf{G_ø},\mathbf{G_a},\mathbf{G_v}\}.$$

- Rotations and translations are the automorphisms of E^3. They form the Euclidean group $\mathbf{G_E} = \mathbf{G_R} \ltimes \mathbf{G}_a$.
- The proper orthochronous[17] Galilei group $\mathbf{Gal}_+^\uparrow$ is generated by rotations and Galilei boosts
$$\vec{x} \mapsto \vec{x}' = \mathbf{R}(\vec{\alpha}) \cdot \vec{x} \qquad \mathbf{R} \in \mathbf{SO(3)}$$
$$\vec{x} \mapsto \vec{x}' = \vec{x} + \vec{v}t.$$
We may compile a group element $(R, v) \in \mathbf{Gal}_+^\uparrow \subset \mathbf{GL(4, R)}$ from a rotation and a boost as
$$(\vec{v}, \mathbf{R}) = (\vec{v}, \mathbf{1}) \circ (\vec{0}, \mathbf{R}) = \begin{pmatrix} \mathbf{1} & \vec{v}^T \\ \vec{0} & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{R} & \vec{0}^T \\ \vec{0} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \vec{v}^T \\ \vec{0} & 1 \end{pmatrix}$$
which is isomorphic to $\mathbf{IG_R}$: $\mathbf{Gal}_+^\uparrow \cong \mathbf{G_R} \ltimes \mathbf{R}^3 \cong \mathbf{SO(3)} \ltimes \mathbf{R}^3$. The group $\mathbf{Gal}_+^\uparrow$ is not simple, and not even semi-simple since it contains a non-trivial Abelian subgroup, namely the Galilei boosts. As we will see, in contrast, the Lorentz group pendant $\mathbf{Lor}_+^\uparrow$, is simple. This seemingly minor difference manifests in quite different representations of these groups.
- The group elements of the inhomogeneous proper Galilei group including space and time translations $(\tau, \vec{a})$ can be built as
$$(\tau,\, \vec{a},\, \vec{v},\, \mathbf{R}(\vec{\alpha})) = \begin{pmatrix} \mathbf{R}(\vec{\alpha}) & \vec{v}^T & \vec{a}^T \\ \vec{0} & 1 & \tau \\ \vec{0} & 0 & 1 \end{pmatrix}. \tag{2.74}$$

The unit element g_0 of the group is identical to the 5×5 unit matrix, and you may verify that these matrices do have the inverse (2.73) and obey the multiplication rule (2.72).

2.3.3 Lie Algebra of the Galilei Group

The Lie algebra associated to $\mathbf{Gal_c}$ is spanned by ten generators. These generators and the algebra can be determined directly from the group elements (2.74) by taking their partial derivatives with respect to the group parameters at the unit element (according to (A.1)). Instead of calling the generators generically X_a, let us rename

[17] These two attributes have an intuitive meaning in case of the Lorentz group; see Chap. 3.4.

them in a way that exhibits their relation to physical quantities[18]:

$$(\tau,\ \vec{a},\ \vec{v},\ \mathbf{R}(\vec{\alpha})) = \exp i \left[\tau H + a^k T^k + v^k G^k + \alpha^k J^k\right].$$

The Lie algebra of the generators $\{H, T^k, G^k, J^k\}$ is

$$[J^j, J^k] = i\epsilon^{jkl} J^l \quad [J^j, T^k] = i\epsilon^{jkl} T_l \quad [J^j, G^k] = i\epsilon^{ijkl} G_l \quad [J_j, H] = 0 \tag{2.75a}$$

$$[H, T^i] = 0 \qquad [H, G^i] = iT^i \tag{2.75b}$$

$$[T^i, T^j] = 0 \qquad [T^i, G^j] = 0 \qquad [G^i, G^j] = 0. \tag{2.75c}$$

This algebra reflects of course the group/subgroup properties stated in the previous subsection. For example $\mathbf{Gal}_+^\uparrow$ is visible in the subalgebra consisting of the J^i and the G^i. Further the first commutator in (2.75a) displays the algebra of the $\mathbf{SO(3)}$ generators, the others show that the T^j and G^j transform as vectors and that H transforms as a scalar with respect to $\mathbf{SO(3)}$. The Galilei algebra can be realized in terms of differential operators as in (2.61).

2.4 Concluding Remarks and Bibliographical Notes

Already in this chapter we have caught a glimpse of how symmetries in an established theory (even if it is "only" classical mechanics) lead to important connections to conservation laws–those known to every high school student–and to group theory, here the Galilei group. Maybe this is an unusual perspective on classical mechanics, but it prepares the ground for things to become substantial in our current understanding of the "world".

As mentioned in the Preface, the heading of this chapter s hould be read as "Symmetries in Classical Mechanics". There are of course good text books on classical mechanics. But only few deal explicitly with the origin of conservation laws from symmetries, one example being the classic [332]. But Landau and Lifschitz do not mention the Noether theorems. These are derived in the widely-used textbooks [230, 305]. The book by José and Saletan is in its level between the book by Goldstein et al. and the more abstract text by Arnold [14].

The connection between the ten classical conservation laws and the corresponding space-time symmetries was already stated by G. Herglotz in 1911. (Herglotz was in the Göttingen group of mathematicians with D. Hilbert, F. Klein, later also joined by E. Noether.) However a proof was provided only later by E. Noether in 1918–except for the Galilei boosts. These could be derived after the extension of Noether's theorem due to E. Bessel-Hagen allowing quasi-invariance of a Lagrange function. For the history of this topic, see [311]. The classic text on variational principles is

[18] Here I distinguish the Lie algebra generators from the conserved quantities that are entailed by the symmetry; except for rotations and angular momenta which receive the same symbol J

[331], in which C. Lanzcos discusses in detail the different forms of what now is called “the principle of least action” in terms of related principles introduced by Maupertuis, d’Alembert, Euler, Hamilton, Jacobi; see also [573]. Most of the topics of this chapter are treated with even more details in [490], although these authors–strange enough–do not refer to E. Noether at all. For the meaning of the geometric notions with respect to the Galilei group consult [220].

Finally a remark about Newton’s assumption about the notions of absolute time and space. Already at his time this was heavily criticized by G.W. Leibniz, and later also for instance by E. Mach. It seems that in light of the success of Newtonian mechanics this criticism did not find the attention which it deserves. The subject matter, however, became topical in recent decades in the context of reconciling general relativity and quantum physics; see e.g. Sect. 2.4 in [451]. In the quantum gravity community more and more attention is given to a relational understanding of space and time, termed with the catchword “background independence”. What does this mean for classical mechanics? As elaborated by J.B. Barbour and B. Bertotti [26] a Leibniz/Mach conception of relational space-time means to replace the Galilei symmetry group by the Leibniz group transformations

$$\vec{x} \;\to\; \vec{x} + \mathbf{A}(\lambda)\vec{x} + \vec{g}(\lambda) \qquad\qquad \lambda \to f(\lambda),$$

where $\mathbf{A}(\lambda)$ is an orthogonal matrix, and $\vec{g}(\lambda)$ and $f(\lambda)$ are arbitrary functions (with the additional condition $\dot{f} > 0$). Observe that this does not define a (finite parameter) Lie group but that it requires for its specification arbitrary functions of the parameter λ. This is the isotropy group of what is investigated as Leibniz spacetime in [172]. Julian Barbour is even more consequent, in denying that time has any meaning as a basic notion of physics; see his [25].

Chapter 3
Electrodynamics and Special Relativity

War es ein Gott der diese Zeichen schrieb?

This chapter is not at all about electrodynamics in the sense that it gives an introduction and overview on this important discipline of theoretical physics and of engineering. I presume that you are acquainted with electrodynamics and special relativity on the level of the textbook by J.D. Jackson [298]. Instead this chapter is about what at the turn of the 19th to the 20th century confused the fundamental physics community, namely the more or less obvious contradiction between classical mechanics and electrodynamics. The solution of this conceptual clash gave birth to special relativity, which as we will see has astounding consequences for today's field theories. If considered under the aspect of "symmetries" this chapter should more appropriately be named after those men and the woman who prepared the ground of our current understanding of symmetries: Poincaré, Lorentz, Einstein, de Sitter, Minkowski, Noether,…

3.1 Electrodynamics à la Maxwell

3.1.1 Maxwell Equations

Since the days of Michael Faraday (1791–1867) and James Clark Maxwell (1831–1879), electrodynamics is formulated in terms of

$$\begin{array}{ll}\text{electric and magnetic fields} & \vec{E}(\vec{x},t),\, \vec{B}(\vec{x},t) \\ \text{charge and current densities} & \varrho(\vec{x},t),\;\; \vec{j}(\vec{x},t).\end{array}$$

Ludwig Boltzmann, who himself contributed to formulating the Maxwell equations in the familiar form (3.1), used this sentence from J. W. von Goethe's "Faust" in order to praise their beauty and symmetry.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_3,

Their dynamics is encoded in the Maxwell field equations, written in a traditional form as

$$\operatorname{div}\vec{E} = \varrho \tag{3.1a}$$

$$\operatorname{rot}\vec{B} = \frac{1}{c}\vec{j} + \frac{1}{c}\frac{\partial\vec{E}}{\partial t} \tag{3.1b}$$

$$\operatorname{rot}\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t} \tag{3.1c}$$

$$\operatorname{div}\vec{B} = 0. \tag{3.1d}$$

Here Gaussian units[1] were adopted and c denotes the vacuum velocity of light. The previous "traditional form" of Maxwell's equations is for various reasons not the form J.C. Maxwell used himself. Aside from the notational difference that he did not apply three-vectors and the operations of vector analysis, the system of equations (3.1) is the *microscopic* variant of Maxwell electrodynamics. It is written in terms of the total charge and total current including the charges and currents at the atomic level (and this structure of matter was of course not known to Maxwell). In the original (macroscopic) form there were two further fields, namely the *displacement* $\vec{D}$ and the *magnetization* $\vec{H}$. These are related to the $\vec{E}$ and $\vec{B}$ fields by *constituent relations* $\vec{D} = \epsilon\vec{E}$ and $\vec{B} = \mu\vec{H}$. Here the permittivity ϵ and permeability μ represent properties of the macroscopic matter.

From the field Eqs. (3.1a, 3.1b), the continuity condition follows:

$$\operatorname{div}\vec{j} + \frac{\partial\varrho}{\partial t} = 0. \tag{3.2}$$

The force on a charge q (Lorentz force) is

$$\vec{F} = q\left(\vec{E} + \frac{\vec{v}}{c}\times\vec{B}\right). \tag{3.3}$$

This is electrodynamics at its core, nowadays well-known to undergraduate students in physics. The very fact that electrodynamics can be written in terms of (three)-vectors is due to its invariance with respect to rotations in space.

3.1.2 Lorentz Boosts

Because the Maxwell equations contain a distinctive velocity, namely the vacuum velocity of light, they are not invariant under Galilei boosts[2] $\vec{x}' = \vec{x} - \vec{\mathfrak{v}}t$ (here the coordinate system K' is assumed to move with a constant velocity $\mathfrak{v}$

---

[1] See the appendix "Units and Dimensions" in [298].

[2] This is not the full truth: If one allows for a specific non-local dependence of the transformed fields on the original fields, Maxwell's equations can be made invariant under Galilei boosts; see [207].

relative to the system K), but instead under Lorentz boosts: If for instance $\vec{\mathfrak{v}}$ has a non-vanishing component in the x-direction, the Lorentz boost does not affect the y- and z-coordinates, but the transformed x- and t-values are

$$x' = \gamma(x - \mathfrak{v}t) \qquad t' = \gamma\left(t - \frac{\mathfrak{v}}{c^2}x\right) \tag{3.4}$$

with the dimensionless factor

$$\gamma := \left(1 - \frac{\mathfrak{v}^2}{c^2}\right)^{-\frac{1}{2}}. \tag{3.5}$$

The transformations (3.4) can be written in a more insightful form by introducing the "rapidity" η

$$\tanh\eta = \mathfrak{v}/c \qquad \gamma = \cosh\eta \qquad \gamma\mathfrak{v} = c\,\sinh\eta,$$

and a modified time coordinate $T = ct$ (which has the dimension of a length):

$$x' = x\cosh\eta - T\sinh\eta \tag{3.6a}$$
$$T' = -x\sinh\eta + T\cosh\eta. \tag{3.6b}$$

It is a little tedious to verify that Maxwell's Eqs. (3.1) are invariant under the transformations (3.4) and I refrain from performing this calculation. Later (in Sect. 3.2.4), you will see how the invariance under generic Lorentz transformations can be made explicit by writing Maxwell's equations in tensorial form.

For $c \to \infty$ or $\mathfrak{v}/c \ll 1$ holds $\gamma \to 1$, and the Lorentz boost turn into $x' \to x - \mathfrak{v}t$, $t' \to t$; that is, Galilei boosts are recovered. The transformations (3.4) were dealt with already at the end of the 19th century, and they were named after Henrik Antoon Lorentz (1853–1928).

The transformations (3.4) form an (Abelian) group. This is most easily seen in the rapidities: If one performs two boosts with $\mathfrak{v}_1$ and $\mathfrak{v}_2$, one after the other, the result is a boost with rapidity $\eta_3 = \eta_1 + \eta_2$. The "addition" of two velocities calculated from $v = dx/dt$ and $v' = dx'/dt'$ is computed to be

$$v' = \frac{v + \mathfrak{v}}{1 + \frac{v\mathfrak{v}}{c^2}}$$

which differs from the Galileian $v' = v + \mathfrak{v}$ by the denominator containing the finite vacuum velocity of light. This expression also shows that c is an invariant velocity ($c' = c$). Observe that the transformations (3.4) allow boost velocities $|\vec{\mathfrak{v}}| < c$, but that $|\vec{\mathfrak{v}}| = c$ is not a valid parameter. Therefore, in the sense of group theory the Lorentz boost group is non-compact.

3.2 Special Relativity

3.2.1 "Deriving" Special Relativity

Of course nothing can prevent that nature might prefer to follow Galilei symmetry in classical mechanics and Lorentz symmetry in electrodynamics. But this is hard to accept, and indeed Einstein proposed a way out of this dilemma by an ingenious idea with what is now called special relativity. The title of his publication, "*Zur Elektrodynamik bewegter Körper*" [151], points directly to the conceptual dilemma between classical mechanics and electrodynamics.

Special relativity can be based on postulates extending the classical mechanics postulate sometimes called Galilei relativity: The three laws of Newton are the same for all inertial systems.

Postulates by Einstein

Einstein based his derivation of special relativity on two postulates

- The first, the relativity principle, is that "*dem Begriff der absoluten Ruhe nicht nur in der Mechanik, sondern auch in der Elektrodynamik keine Eigenschaften der Erscheinungen entsprechen*"[3]. This is sometimes briefly worded as: The laws of physics are the same in all inertial systems.
- The second is that "*sich das Licht im leeren Raume stets mit einer bestimmten, vom Bewegungstzustande des emittierenden Körpers unabhängigen Geschwindigkeit* V *fortpflanze*"[4]. In short: The vacuum velocity of light is constant.

These two postulates imply the Lorentz boosts (3.4). This can be demonstrated by a simple calculation that can be understood and performed by anyone who has only moderate knowledge of algebra. Einstein himself presented it in his famous popular book [155], which "…presumes a standard of education corresponding to that of a university matriculation examination…"[5] essentially as follows:

Take two coordinate systems K and K', which are arranged in such a way that K' moves relative to K with constant velocity $\mathfrak{v}$ in the x-direction; see Fig. 3.1. Consider a light signal which propagates in the positive x-direction: $x = ct$ or $x - ct = 0$. Since according to the second postulate the signal propagates in the other coordinate system at the same speed, we also have $x' - ct' = 0$. Necessarily $x - ct = \lambda(x' - ct')$ with some constant λ. Similarly for a light signal in the negative x-direction, $x + ct = \mu(x' + ct')$. Solving the previous relations for (x', t') yields

$$x' = Ax - Bct, \qquad ct' = Act - Bx$$

[3] "Not only in mechanics but also in electrodynamics the phenomena have no properties corresponding to the concept of absolute rest."

[4] "Light always propagates in empty space with a definite velocity c, independent of the state of motion of the emitting body."

[5] "Die Lektüre setzt etwa Maturitätsbildung …voraus".

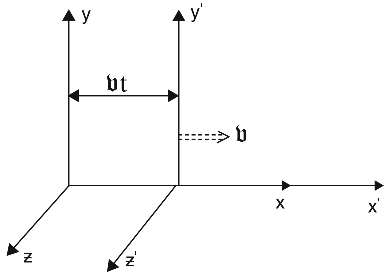


Fig. 3.1 A boost in x-direction

where A and B depend on the constants λ and μ. Since the origin $x' = 0$ moves with velocity $\mathfrak{v} = x/t$, we find from the previous relation $0 = A\mathfrak{v}t - Bct$, by which B is expressed through A (and $\mathfrak{v}$), so that

$$x' = A(x - \mathfrak{v}t) \qquad t' = A\left(t - \frac{\mathfrak{v}}{c^2}x\right). \tag{3.7}$$

Solving these for (x, t) yields

$$x = A^{-1}\gamma^2(x' + \mathfrak{v}t') \qquad t = A^{-1}\gamma^2\left(t' + \frac{\mathfrak{v}}{c^2}x'\right),$$

with γ defined as in (3.5). Finally, since both coordinate systems are on an equal footing (after replacing $\mathfrak{v} \leftrightarrow -\mathfrak{v}$), it must hold that $A \overset{!}{=} A^{-1}\gamma^2$, or $A = \gamma$. This indeed casts (3.7) in the form of a Lorentz boost.

Relativity "Without Light"

It is possible to deduce the form of the Lorentz boosts from other sets of postulates, even without referring to the propagation of light in vacuum. This was stated by W.A. Ignatowsky already as early as 1910, and proven by P. Frank and H. Rothe in 1911. And this is remarkable, given that today we know about interactions other than the electromagnetic one, and that the meaning of vacuum is not without bias. Instead of the previous two postulates start with[6]

- the principle of relativity
- homogeneity of spacetime and isotropy of space
- a condition of "pre-causality": If two events happen at the same place in one reference frame, their time order must be the same in all reference frames.

Thus the focus is shifted away from the special role of light propagation towards properties of space and time. For the relative movement of the two coordinate systems K and K' (see again Fig. 3.1), the relativity principle and the isotropy of space would allow the transformations

$$x = \mathfrak{v}t + \Gamma(\mathfrak{v}^2)x' \qquad \text{or} \qquad x' = \Gamma^{-1}(x - \mathfrak{v}t)$$

[6] In the rest of this section I follow [337, 342, 367]; for similar approaches see [220, 472, 508].

with a scale factor Γ depending on the magnitude of $\mathfrak{v}$. The relativity principle demands that the same expression must hold in the primed coordinates–with $\mathfrak{v}$ replaced by $(-\mathfrak{v})$:

$$x = \Gamma^{-1}(x' + \mathfrak{v}t) = \Gamma^{-1}[\Gamma^{-1}(x - \mathfrak{v}t) + \mathfrak{v}t'].$$

The latter can be solved for t':

$$t' = \Gamma^{-1}\left[t - \frac{1-\Gamma^2}{\mathfrak{v}}x\right].$$

The previous relations make it possible to derive the relation among the velocities in the two coordinate systems:

$$v' = \frac{dx'}{dt'} = \frac{dx - \mathfrak{v}dt}{dt - \frac{1-\Gamma^2}{\mathfrak{v}}dx} = \frac{v - \mathfrak{v}}{1 - \frac{1-\Gamma\,2}{\mathfrak{v}}v},$$

from which again by the exchange of K and K'

$$v = \frac{v' + \mathfrak{v}}{1 + \frac{1-\Gamma^2}{\mathfrak{v}}v'} =: f(v', \mathfrak{v}), \tag{3.8}$$

displaying an unaccustomed addition of velocities. This is expressed in the function $f(v_1, v_2)$ whose properties will be exhibited next. First, if the sign of the velocities v' and $\mathfrak{v}$ is changed, obviously also the sign of v changes. Thus $f(-v_1, -v_2) = -f(v_1, v_2)$. Consider next three bodies A, B, and C in relative motion and denote by v_{AB} the velocity of A with respect to B. Then one can derive from the chain

$$f(v_{\mathrm{CB}}, v_{\mathrm{BA}}) = v_{\mathrm{CA}} = -v_{\mathrm{AC}} = -f(v_{\mathrm{AB}}, v_{\mathrm{BC}}) = -f(-v_{\mathrm{BA}}, -v_{\mathrm{CB}}) = f(v_{\mathrm{BA}}, v_{\mathrm{CB}})$$

that $f(v_1, v_2)$ is a symmetric function of its arguments. Thus (3.8) yields

$$\frac{v' + \mathfrak{v}}{1 + \frac{1-\Gamma^2(\mathfrak{v}^2)}{\mathfrak{v}}v'} = \frac{\mathfrak{v} + v'}{1 + \frac{1-\Gamma^2(v'^2)}{v'}\mathfrak{v}}$$

leading to the algebraic condition

$$\frac{1-\Gamma^2(\mathfrak{v}^2)}{\mathfrak{v}^2} = \kappa = \frac{1-\Gamma^2(v'^2)}{v'^2} \tag{3.9}$$

where κ is a velocity-independent constant. Therefore the velocity addition rule (3.8) takes the form

$$v = \frac{v' + \mathfrak{v}}{1 + \kappa\mathfrak{v}v'} \tag{3.10}$$

from which we see that the constant κ has the dimension [velocity]$^{-2}$. Three cases may be distinguished

- For $\kappa = 0$ we recover the Galilei transformations with $v = v' + \mathfrak{v}$.
- For $\kappa > 0$ write $\kappa = 1/C^2$ with a reference velocity C. Furthermore, with the dimensionless parameter $\beta = \mathfrak{v}/C$, we obtain from (3.9) that $\Gamma^2 = 1 - \beta^2$. We might now just as well introduce the rapidity $\eta = \frac{1}{\sqrt{1-\beta^2}}$ and a rescaled time $T = Ct$, such that finally

$$x' = x \cosh \eta - T \sinh \eta \qquad\qquad T' = -x \sinh \eta + T \cosh \eta.$$

These are identical to the Lorentz-boosts (3.6) if C is identified with vacuum velocity of light c. Within the foregoing derivation C arises as a reference velocity. The value of C remains undetermined. Only experiments (like that of Michelson and Morley) tell us that to a high degree of accuracy $C = c$. And only for $C = c$ are the Maxwell equations invariant with respect to the Lorentz transformations. But observe that in the derivation of the boost equations, various points are left open: For one thing there is no indication that C is the maximum speed of objects. Nor does the result show that signals cannot propagate at a speed higher than C. One must only exclude reference frames with relative speed $\mathfrak{v} > C$, since otherwise Γ would become imaginary. These remarks point in the direction of tachyons (or superluminal objects). Tachyons were disliked and abandoned, essentially because they were thought to lead to causality violations. However, theoretical arguments have become more refined in the meantime and there are even some experimental claims, see [435]. Superluminal objects also seem to be fashionable these days in contexts like cosmology, strings, etc.
- For $\kappa < 0$ write $\kappa = -1/C^2$ with a (different) reference velocity C, and again $\beta = \mathfrak{v}/C$. Then the expression for the resulting velocity

$$v' = v' = \frac{v + \mathfrak{v}}{1 - \frac{v\mathfrak{v}}{C^2}}$$

exhibits some strange features, because the denominator can be positive, negative and even zero. Due to this, various kinds of weird situations may arise: Two positive velocities may sum up to a negative velocity, or to an infinite velocity. The deeper reason of this weirdness can be discovered in the following way: The transformations in the (x, t)-space are

$$x' = \frac{1}{\sqrt{1+\beta^2}}(x - \mathfrak{v}t) \qquad\qquad t' = \frac{1}{\sqrt{1+\beta^2}}\left(t + \frac{\mathfrak{v}}{C^2}x\right).$$

Rescale the time coordinate again as $T = Ct$ and use instead of β the parameter $\cos\phi := \frac{1}{\sqrt{1+\beta^2}}$. Then the previous (x, t)-transformations become the transformations

$$x' = x \cos\phi - T \sin\phi \qquad\qquad T' = x \sin\phi + T \cos\phi. \tag{3.11}$$

But these are just rotations by an angle ϕ in the (x, T)-plane. The composition of boosts amounts to an addition of angles. Thus a consecutive composition of finite boosts could result in zero velocity. Furthermore, time is treated symmetrically with space, and thus no invariant notion of time-orientability is possible. Phrased in another way: Since the invariant is $T^2 + x^2$, it is not possible to say which points precede a given one. For these reasons–by the third postulate above–the case $\kappa < 0$ must be excluded.

The distinction between the cases $\kappa > 0$ and $\kappa < 0$ has a group-theoretical interpretation in that the two-dimensional Lorentz boosts form the group $\mathbf{SO(1,1)}$, which is non-compact, whereas the transformations (3.11) form $\mathbf{SO(2)}$, which is compact.

3.2.2 Minkowski Geometry

Motivation

We saw that the symmetries of classical mechanics are connected with the automorphism groups of an underlying geometry, namely Galilei space-time. Since Lorentz boosts generalize the Galilei boosts, we can expect presumably that we are dealing with a more general geometry.

- The Galilei group contains the transformations

$$G_R: \quad \vec{x}' = \mathbf{R}\vec{x} \qquad G_\tau: \quad t' = t + \tau$$

  as automorphisms of $\mathbb{E}^3$ and $\mathbb{E}^1$, respectively. Lorentz boosts (3.4) mix space coordinates and time coordinates, in contrast to Galilei boosts, in which time translations are not influenced by spatial transformations. In order to get a grip on the new/changed geometry (called $\mathbb{M}^4$), we need to consider invariants.
- Galilei boosts leave distances $\Delta := \sqrt{|\vec{x}_i - \vec{x}_j|^2}$ invariant, infinitesimally expressed as

$$(d\triangle)^2 = dx^2 + dy^2 + dz^2$$

  in Cartesian coordinates, or for instance by

$$(d\triangle)^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2$$

  in spherical coordinates. In the general case the infinitesimal distance can be written with the help of a metric in the form

$$(d\triangle)^2 = g_{ik}\, dx^i dx^k$$

  where for example the (canonical) metric in Cartesian coordinates is simply given by

$$(g_{ik}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

- What could be the canonical metric of $\mathbb{M}^4$? Since for the Lorentz boosts (3.6)

$$x'^2 = x^2 \cosh^2 \eta^2 - 2xT \cosh\eta \sinh\eta + T^2 \sinh^2 \eta$$
$$T'^2 = x^2 \sinh^2 \eta^2 - 2xT \cosh\eta \sinh\eta + T^2 \cosh^2 \eta,$$

we find that

$$T'^2 - x'^2 = T^2 - x^2.$$

Minkowski Space

The canonical metric of the Minkowski space $\mathbb{M}^4$ is defined by

$$(ds)^2 = \eta_{\mu\nu} dx^\mu dx^\nu \qquad (\eta_{\mu\nu}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{3.12}$$

By setting $x^0 = ct$ and $(x^i) = \vec{x}$, $(ds)^2 = c^2 dt^2 - d\vec{x}^2$ indeed becomes an invariant. Choosing the Minkowski metric with $(\eta_{\mu\nu}) = \mathrm{diag}(1, -1, -1, -1)$ is of course a convention. Other authors prefer the $\mathrm{diag}(-1, 1, 1, 1)$ convention. Typically the latter has advantages in the relativity context, whereas the former is more appropriate for dealing with particle physics[7].

Hermann Minkowski (1864–1909) was a professor of mathematics at the universities of Zürich and Göttingen[8], who exhibited the mathematical structures of special relativity. His introductory sentences at a conference in 1908 [372] of a talk with the title "*Raum und Zeit*" could be the libretto of an opera by Richard Wagner: "*M(eine).H(erren)! Die Anschauungen über Raum und Zeit, die ich Ihnen entwickeln möchte, sind auf experimentell-physikalischem Boden erwachsen. Darin liegt ihre Stärke. Ihre Tendenz ist eine radikale: Von Stund an sollen Raum für sich und Zeit für sich zu Schatten herabsinken und nur eine Art Union der beiden soll Selbständigkeit bewahren*"[9]. When first confronted with Minkowski's considerations,

---

[7] Only a few decades ago you could judge from a publication whether the author was educated in general relativity or in particle physics just by his or her use of the metric. Since these disciplines nowadays have converged to a new discipline called "cosmoparticle physics", one should be able to read expressions in both metric conventions and to translate them. I allow myself to be inconsistent in using the "mostly plus" metric in some contexts, but tried to emphasize this deviation properly.

[8] In fact, Einstein attended lectures by Minkowski as a student.

[9] "The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. Henceforth space by itself, and time by itself are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

Einstein uttered "*Seit die Mathematiker über die Relativitätstheorie hergefallen sind, verstehe ich sie selbst nicht mehr*"[10]. Only later, on his tedious route towards general relativity, did he recognize the benefits of Minkowski's formalism. For the developments leading to the Cologne talk and its reception in the scientific community see [110].

Poincaré Transformations

The Minkowski geometry was motivated above by the symmetry invariance with respect to Lorentz boosts. However we would also like to recover the automorphisms of Euclidean space (namely the rotations and the translations.) Thus goal provokes the question of which transformations leave $(ds)^2$ invariant. The answer is: Poincaré transformations, which are transformations of the form

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + a^{\mu} \tag{3.13}$$

with constant a^{μ} and constant matrices $\Lambda^{\mu}{}_{\nu}$, that fulfill

$$\Lambda^{\mu}{}_{\nu}\, \eta_{\mu\rho}\, \Lambda^{\rho}{}_{\sigma} = \eta_{\nu\sigma} \tag{3.14}$$

–or–written in matrix form as $\Lambda^T \eta \Lambda = \eta$, from which it immediately follows that $\det \Lambda = \pm 1$. Because of

$$\Lambda^{\mu}{}_{\nu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}}, \tag{3.15}$$

this in turn guarantees that the transformation $x \to x'$ is invertible. Infinitesimally

$$x'^{\mu} = x^{\mu} + \omega^{\mu}{}_{\nu} x^{\nu} + \epsilon^{\mu} \qquad \text{where} \quad \omega^{\mu\nu} \equiv -\omega^{\nu\mu} \tag{3.16}$$

with the composition rule

$$x''^{\mu} = x^{\mu} + \left(\omega^{\mu}{}_{\nu} + \omega'^{\mu}{}_{\nu}\right) x^{\nu} + \left(\epsilon^{\mu} + \epsilon'^{\mu}\right).$$

It is an easy exercise to verify that transformations of the form (3.13) with the restriction (3.14) leave $(ds)^2$ invariant.

For the proof that all transformations leaving $(ds)^2$ invariant are necessarily of this form, we draw on notions which are due to F. Klein (1849–1925) and W. Killing[11]: For a space with metric $g_{\mu\nu}(x)$, a coordinate transformation $x \to x'$ is called an isometry iff $g'_{\mu\nu}(x') = g_{\mu\nu}(x')$, which can also be written $\bar{\delta} g_{\mu\nu} = 0$. In Appendix E.5.4, it is shown that the isometry condition can be expressed infinitesimally as

$$g_{\mu\nu,\lambda}\xi^{\lambda} + g_{\mu\lambda}\xi^{\lambda}{}_{,\nu} + g_{\lambda\nu}\xi^{\lambda}{}_{,\mu} = 0$$

---

[10] "Since the mathematicians pounced on the relativity theory I no longer understand it myself."; taken from the Einstein biography of C. Seelig from 1960.

[11] Wilhelm Killing (1847-1923) was a professor of mathematics in Münster, a town in Westfalia/Germany. His name is less well known than for instance those of S. Lie and E. Cartan, although results going under their names were found by him independently.

for $x'^{\mu} = x^{\mu} + \xi^{\mu}$ (with $|\xi^{\mu}| \ll 1$). Any solution ξ of this equation is called a Killing vector of the metric space.

Let's get back to the theme of this subsection: Poincaré transformations are the isometries of the Minkowski metric, since the isometry condition becomes in this case

$$\frac{\partial \xi^{\mu}}{\partial x^{\nu}} + \frac{\partial \xi^{\nu}}{\partial x^{\mu}} = 0,$$

and the solution of these differential equations is simply (3.16). Even so, there is a specific situation in which the invariance group is larger than the Poincaré group: $(ds)^2 = 0$ is invariant with respect to conformal transformations; these are treated in Sect. 3.5.1.

The Lorentz boosts which we investigated above (see (3.6a, 3.6b)) are just a specific case of the matrices Λ in (3.13), namely

$$\Lambda_b = \begin{pmatrix} \cosh\eta & 0 & 0 & \sinh\eta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh\eta & 0 & 0 & \cosh\eta \end{pmatrix}.$$

The Poincaré transformations constitute a group with the composition of two group elements $x' = \Lambda x + a$ and $x'' = \Lambda' x' + a'$ as

$$x'' = \Lambda\Lambda' x + (\Lambda' a + a') \qquad \text{or} \qquad \Lambda'' = \Lambda\Lambda' \qquad a'' = \Lambda' a + a'. \tag{3.17}$$

Comparing this with the composition rule (2.72) for the Galilei group, one realizes that the Poincaré group is "nicer"–or to use the term of Minkowski: "more intelligible". The transformations with $a \equiv 0$ constitute a subgroup, called Lorentz or the homogeneous Poincaré group.

Lorentz Tensors

In order to recognize immediately wether a physical theory obeys the postulates of special relativity, it is advantageous to formulate the theory in terms of so-called Lorentz tensors. The basic idea is as follows: If we have an equation in the theory of the generic form $\mathfrak{L} = \mathfrak{R}$ (where possible indices are suppressed), and if after a Lorentz transformation with Λ the equation has the form $\Lambda\mathfrak{L} = \Lambda\mathfrak{R}$, Lorentz invariance is guaranteed. But for this to be the case $\mathfrak{L}$ and $\mathfrak{R}$ must transform in a specific form, namely as Lorentz tensors. This consideration is not peculiar to special relativity. In Newtonian mechanics we use a three-vector notation, and three-vectors are specific tensors in $\mathbb{E}^3$, see [181] Chap. 11. And later in this text, in general relativity, where Minkowski geometry is generalized to Riemann geometry, we will argue using tensors appropriate to this geometry.

Tensors are distinguished by their transformation properties (*sic*)–or in group theoretical terms–how they transform with respect to specific representations of the Lorentz group.

- The simplest tensor is a Lorentz scalar S, which is invariant under Lorentz transformations
$$S \to S' = S.$$
Example (1): $(ds)^2 = \eta_{\mu\nu} dx^\mu dx^\nu$.
Example (2): volume element d^4x, since $d^4x' = \left|\frac{\partial x'}{\partial x}\right| d^4x = |\det \Lambda|\, d^4x = d^4x$.
A scalar transforms according to the trivial representation of the Lorentz group.
- A contravariant vector a^μ transforms according to the defining representation of the Lorentz group, that is as
$$a^\mu \to a'^\mu = \Lambda^\mu{}_\nu a^\nu.$$
Example: coordinates x^μ, which is obvious from (3.13).
- A covariant vector a_μ transforms as
$$a_\mu \to a'_\mu = a_\nu \bar{\Lambda}^\nu{}_\mu,$$
where $\bar{\Lambda}$ is as a matrix identical to Λ^{-1}, i.e. $\bar{\Lambda}^\nu{}_\mu = \frac{\partial x^\nu}{\partial x'^\mu}$.
Example: gradient ∂_μ, since
$$\partial'_\mu = \frac{\partial}{\partial x'^\mu} = \frac{\partial}{\partial x^\nu} \frac{\partial x^\nu}{\partial x'^\mu} = \partial_\nu \bar{\Lambda}^\nu{}_\mu.$$
- For tensors with multiple contravariant and covariant indices one requires an analogous transformation behavior. For instance
$$A'^\mu{}_{\nu\lambda} = \Lambda^\mu{}_\rho \; A^\rho{}_{\sigma\tau} \; \bar{\Lambda}^\varrho{}_\nu \bar{\Lambda}^\tau{}_\lambda.$$
Example: The condition (3.14) shows that the Minkowski metric $\eta_{\mu\nu}$ is a tensor. This tensor is even invariant ($\eta'_{\mu\nu} = \eta_{\mu\nu}$). The only other invariant tensor in $\mathbb{M}^4$ is the totally antisymmetric Levi-Civita tensor $\epsilon^{\mu\nu\rho\sigma}$.

Tensors with the same number of covariant and contravariant indices can be added and subtracted to yield a tensor of the same type, e.g.: $A_\lambda{}^{\mu\nu} + B_\lambda{}^{\mu\nu} = C_\lambda{}^{\mu\nu}$. The product of arbitrary tensors $A^{\cdots}{}_{\cdot\cdot}$ and $B^{\cdot}{}_{\cdot\cdot}$ is again a tensor, which inherits the contravariant and covariant indices from the two factors, e.g. $A^{\mu\nu} B_\lambda = C^{\mu\nu}{}_\lambda$. Immediate and important consequences of these tensor addition and multiplication laws are

- If v^μ is contravariant, $v_\mu = \eta_{\mu\nu} v^\nu$ is a covariant vector. We check this by by considering $v'_\mu = \eta_{\mu\nu} v'^\nu = \eta_{\mu\nu} \Lambda^\nu{}_\rho v^\rho$. From (3.14), we deduce $\eta_{\mu\nu}\Lambda^\nu{}_\rho = \bar{\Lambda}^\nu{}_\mu \eta_{\nu\rho}$, and thus
$$v'_\mu = \bar{\Lambda}^\nu{}_\mu \eta_{\nu\rho} v^\rho = \bar{\Lambda}^\nu{}_\mu \eta_{\nu\rho} v^\rho = \bar{\Lambda}^\nu{}_\mu v_\nu.$$
The metric can therefore be used to pull down upper indices. Accordingly the tensor $\eta^{\mu\nu}$ is used to raise lower indices.

- $(\partial_\mu v^\mu)$ is a Lorentz scalar because this it is a product of a covariant and a contravariant vector.
- The d'Alembert operator $\eta^{\mu\nu}\partial_\mu\partial_\nu := \frac{\partial^2}{c^2\partial t^2} - \nabla^2$ is a Lorentz scalar.

Aside from Lorentz tensors (which, as said before, obtain their properties from the representation theory of the Lorentz group) we will later (in Chap. 5) meet spinors which also transform according to specific representations of the Lorentz group, and which in physics are important to describe particles and fields with half-integer spin.

3.2.3 Relativistic Mechanics

The traditional formulation of classical mechanics is not in the form of Lorentz tensors. In order to manifestly fulfill the postulates of special relativity, we will express the typical entities of classical mechanics (velocity, momentum, force, energy) as tensors[12].

- 4-velocity
  The 3-velocity $v_i = \frac{dx_i}{dt}$ is not a tensor. In order to construct a tensor with the dimension of a velocity, we can use the coordinates x^μ and a time variable that transforms as a scalar. This time variable–called the eigen-time of a mass point–is simply built from the scalar $(ds)^2$ as

  $$d\tau^2 = \frac{1}{c^2}(ds)^2 \,.$$

  By the rules of tensor arithmetics the 4-velocity

  $$u^\mu := \frac{dx^\mu}{d\tau}$$

  is a contravariant vector since dx^μ is a contravariant vector and $d\tau$ is a scalar. Written in $(0, i)$-components the 4-velocity is $(u^\mu) = \gamma(c, v^i)$, with γ given by (3.5). The 4-velocity obeys

  $$u^2 = u_\mu u^\mu = \eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = c^2$$

  because of $c^2 d\tau^2 = \eta_{\mu\nu}dx^\mu dx^\nu$.
- 4-momentum
  From the 4-velocity and the mass m (a scalar) we construct the 4-momentum of a mass point in an obvious manner:

  $$p^\mu := mu^\mu =: \left(\frac{E_R}{c},\, p_R^i\right).$$

[12] In the rest of this chapter, Lorentz tensors will simply be called tensors, since no confusion with Riemann tensors–the basic entities in general relativity–can arise.

From $u^2 = c^2$, the mass-shell condition

$$p^2 - m^2c^2 = 0. \tag{3.18}$$

results. This implies the dispersion relation

$$\frac{E_R^2}{c^2} - p_R^2 = m^2c^2. \tag{3.19}$$

Approximately (for $v/c \ll 1$)

$$p_R^i = \gamma m v^i \simeq m v^i \left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = m v^i + \mathcal{O}\left(\frac{v^2}{c^2}\right),$$

where $mv^i = p_N^i$ is the Newtonian momentum. Furthermore

$$E_R = \gamma mc^2 \simeq mc^2 \left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = mc^2 + \frac{1}{2}mv^2 + \mathcal{O}\left(\frac{v^4}{c^2}\right). \tag{3.20}$$

Here $E_0 = mc^2$ is the rest-energy and $E_N = \frac{1}{2}mv^2$ the kinetic energy in terms of Newtonian mechanics. Observe that it is not true that energy equals mc^2 as often stated to be "Einstein's most famous equations" and printed on T-shirts and elsewhere: $E = mc^2$ is simply wrong!

- 4-force

Finally the "time"-derivative of the 4-momentum is defined as the 4-force:

$$f^\mu := \frac{dp^\mu}{d\tau} = m\frac{d^2x^\mu}{d\tau^2}.$$

The Light Cone

Many properties of relativistic physics are encoded in the light cone. This picture arises for instance from the mass-shell relation (3.18). For ease of notation, in units $c = 1$, this is $p^2 = m^2$ which originates from the time-like property of the four velocity

$$\left(\frac{dx}{d\tau}\right)^2 = 1.$$

This equation describes a (double) cone in four dimensions. Since drawings in 4D are not that easy, we sketch the light cone in 2D, characterized by coordinates (t, x), in Fig. 3.2. Then obviously the upper half represents the future and the lower half the past. The cone is bounded by the two straight lines $x = \pm t$. These are the trajectories of massless particles (e.g. photons). Massive particles with a speed lower than $c = 1$ are bound to the interior of the light cone. The trajectory L1 is called time-like, L2 is light-like, and L3 is space-like.

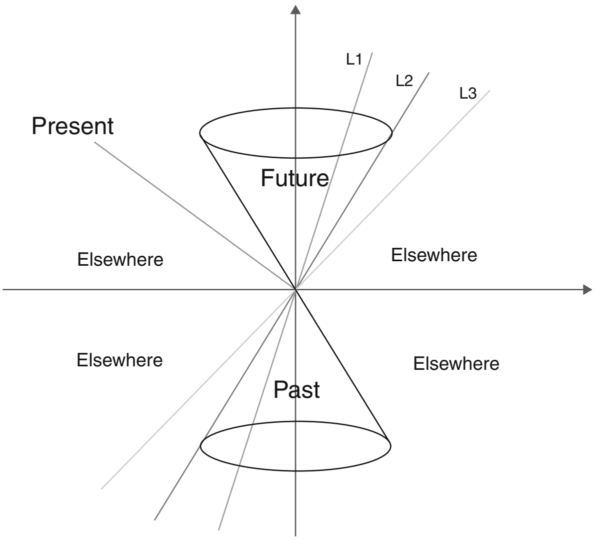


Fig. 3.2 A light cone

Symmetries and Conservation Laws of Relativistic Mechanics

The Minkowski line element $ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$ can be written as

$$ds = c\gamma^{-1}dt \text{ where } \gamma^{-1} = \sqrt{1-\beta^2},\ \ \beta = \left(\frac{\vec{v}}{c}\right)^2,\ \ \vec{v} = \frac{d\vec{x}}{dt} = \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right).$$

Therefore

$$\frac{d\vec{x}}{ds} = \frac{\gamma\vec{v}}{c} \qquad \frac{dt}{ds} = \frac{\gamma}{c}.$$

These equations can be derived from an action which is proportional to the length of the world-line between two points a and b:

$$S = \mu\int_a^b ds = \mu c\int_{t_1}^{t_2}\sqrt{1-\beta^2}dt.$$

The constant μ is found from requiring that for velocities $v^2 \ll c^2$ the Lagrangian is identical to the Lagrangian for a free particle in Newtonian mechanics, that is

$$L = \mu c\sqrt{1-\beta^2} = \mu c\left(1 - \frac{1}{2}\beta^2 + \mathcal{O}(\beta^4)\right) \simeq \mu c - \frac{\mu v^2}{2c}.$$

As a constant, the first term is irrelevant in the Lagrangian, and the second term becomes the Newtonian Lagrangian for $\mu c = -m$. Therefore,

$$L = -mc^2\sqrt{1-\beta^2} \qquad \frac{\partial L}{\partial \dot{x}^i} = \frac{m}{\sqrt{1-\beta^2}}\dot{x}^i \qquad [L]_i = \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}.$$

With the techniques described in Sect. 2.2 one determines the Noether symmetry generators

$$X_0 = \partial_t, \quad X_i = \partial_i, \quad X_{ij} = x^i\partial_j - x^j\partial_i, \quad X_{0i} = t\partial_i + \frac{1}{c^2}x^i\partial_t. \tag{3.21}$$

The calculation is straightforward, but tedious. So at least you may verify that the generators $X_\alpha = \xi_\alpha\partial_t + \eta^i_\alpha\partial_i$ are such that $\overline{X}_\alpha L + L\,\nabla\xi_\alpha$ can be written as total derivatives. The generators fulfill the Poincaré algebra; more details below. The symmetries mediated by (3.21) give rise to conserved charges

$$C_\alpha = \frac{\partial L}{\partial \dot{x}^i}(\eta^i_\alpha - \dot{x}^i\xi_\alpha) + L\xi_\alpha.$$

These are

$$\begin{aligned} C_0 &= -m\gamma c^2 = -E_R, \quad C_i = m\gamma\dot{x}^i = p^i_R, \\ C_{ij} &= \epsilon_{ijk}x^i p^j_R, \quad C_{oj} = m\gamma(t\dot{x}^j - x^j) = Q_R, \end{aligned}$$

which are recognized as the relativistic energy, the components of the relativistic momentum, the relativistic angular momentum $\vec{M} = \vec{x} \times \vec{p}_R$, and as the vector of the relativistic center-of-mass theorem.

3.2.4 *Relativistic Field Theory*

Chapter 5 is completely devoted to relativistic field theories. Here I seize the opportunity to illustrate basic concepts of the Lagrangian formulation of a relativistic field theory using electrodynamics. This will then in later subsections serve to derive the two theorems of E. Noether, the basic properties of the Poincaré group, and its generalizations.

Field Theory

Electrodynamics is the prime example of a field theory. Whereas classical mechanics is described in terms of time-dependent position variables $q_k(t)$ and their derivatives with respect to time, electrodynamics is formulated in terms of fields depending on space and time (e.g. the electric field $\vec{E}(t,\vec{x})$) and their derivatives with respect to time and space.

The notions of point mechanics can be translated to a field theory (in D dimensions) as follows:

$$
\begin{aligned}
\{t\} \quad &\Rightarrow \quad \{x^\mu\} \quad (\mu = 0, \ldots D-1) \\
\{q^k(t)\} \quad &\Rightarrow \quad \{Q^\alpha(x)\} \\
L(q^k, \dot{q}^k, t) \quad &\Rightarrow \quad \mathcal{L}(Q^\alpha, \partial_\mu Q^\alpha, x) \\
S[q] = \int_{t_1}^{t_2} dt\, L \quad &\Rightarrow \quad S[Q] = \int_\Omega d^D x\, \mathcal{L} \\
[L]_k := \frac{\partial L}{\partial q^k} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^k} \quad &\Rightarrow \quad [\mathcal{L}]_\alpha := \frac{\partial \mathcal{L}}{\partial Q^\alpha} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu Q^\alpha)}.
\end{aligned}
$$

The function $\mathcal{L}$ is called Lagrange density, in the sequel also referred to as Lagrangian. The Lagrange function is the spatially integrated Lagrange density: $L[Q] := \int d^3x\, \mathcal{L}[Q]$. The action is a functional of the fields.

For a functional $F[Q]$ its *functional derivative* with respect to Q is defined only if the variation of F can be written in the form

$$\delta F[Q] = \int d^D x\, f_\alpha(x) \delta Q^\alpha(x),$$

that is if no derivatives of the Q^α occur under the integral. Then the functional derivative is

$$\frac{\delta F}{\delta Q^\alpha} := f_\alpha.$$

For the fields themselves, this amounts to the basic definition

$$\frac{\delta Q^\alpha(x)}{\delta Q^\beta(x')} = \delta^\alpha_\beta\, \delta^D(x, x') =: \delta^\alpha_{\beta'}$$

where the latter notation is the one introduced by B. DeWitt.

We already saw in the case of classical mechanics that the equations of motion only follow from the action if the boundary term vanishes. The natural assumption is to keep the endpoints of the variables fixed in the variation of the Lagrangian. We will see, that in field theories one needs to be more careful. As one became aware in the context of the long unsolved question of energy-momentum in general relativity and other diffeomorphism invariant theories, a variational principle is not only defined by its Lagrangian but also by surface terms and boundary conditions. Indeed, in some contexts, the surface terms carry essential information, and one is not allowed to neglect them.

Within the next sections, let us assume that the Lagrangians depend at most on first derivatives of the fields. This is true for the Yang-Mills type Lagrangian of the standard model of elementary particles, but no longer for general relativity. Therefore in the appropriate chapters I will mention those changes that might become necessary in the following arguments.

Let us consider a generic variation of the Lagrangian including a variation of the coordinates:

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial Q}\delta Q + \frac{\partial\mathcal{L}}{\partial(\partial_\mu Q)}\delta(\partial_\mu Q) + \frac{\partial\mathcal{L}}{\partial x^\mu}\delta x^\mu,$$

where possible indices on the fields are suppressed. (At the very end of the calculation it is quite obvious where they are to be inserted in the final expressions.) Since $\delta\partial \neq \partial\delta$ one cannot immediately employ the chain rule on the second term. Similar to the finite dimensional case we consider instead the $\bar{\delta}$ variation

$$\bar{\delta}Q := \delta Q - (\partial_\mu Q)\delta x^\mu$$

for which $\partial_\mu(\bar{\delta}Q) = \bar{\delta}(\partial_\mu Q)$. Then the variation $\delta\mathcal{L}$ becomes

$$\begin{aligned}\delta\mathcal{L} &= \frac{\partial\mathcal{L}}{\partial Q}\bar{\delta}Q + \frac{\partial\mathcal{L}}{\partial Q}Q_{,\mu}\delta x^\mu + \frac{\partial\mathcal{L}}{\partial(\partial_\mu Q)}\bar{\delta}(\partial_\mu Q) + \frac{\partial\mathcal{L}}{\partial(\partial_\nu Q)}(\partial_\nu Q)_{,\mu}\delta x^\mu + \frac{\partial\mathcal{L}}{\partial x^\mu}\delta x^\mu \\ &= \frac{\partial\mathcal{L}}{\partial Q}\bar{\delta}Q - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu Q)}\right)\bar{\delta}Q + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu Q)}\bar{\delta}Q\right) \\ &\qquad + \left[\frac{\partial\mathcal{L}}{\partial Q}Q_{,\mu} + \frac{\partial\mathcal{L}}{\partial(\partial_\nu Q)}(\partial_\nu Q)_{,\mu} + \frac{\partial\mathcal{L}}{\partial x^\mu}\right]\delta x^\mu.\end{aligned}$$

The coefficient in front of the δx^μ term is nothing but $\partial_\mu\mathcal{L}$, whereas the terms proportional to $\bar{\delta}Q$ constitute the Euler derivative of the Lagrangian. Thus

$$\delta\mathcal{L} = [\mathcal{L}]_\alpha\bar{\delta}Q^\alpha + \partial_\mu\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu Q^\alpha)}\bar{\delta}Q^\alpha + \mathcal{L}\delta x^\mu\right] \tag{3.22}$$

where the indices on the fields were reintroduced. In the next section this relation will be taken as a starting point for deducing Noether's theorems in a field theory. Here we obtain the Euler-Lagrange equations $[\mathcal{L}]_\alpha = 0$ assuming that only the fields are varied ($\delta x^\mu = 0$) and that the fields are fixed at the boundary $\partial\Omega$. (This complies with the previous remark about the action functional having a defined functional derivative.) Then indeed

$$\frac{\delta S}{\delta Q^\alpha} = \frac{\partial\mathcal{L}}{\partial Q^\alpha} - \partial_\mu\frac{\partial\mathcal{L}}{\partial_\mu Q^\alpha} = [\mathcal{L}]_\alpha.$$

By introducing the "pseudo-momenta"

$$\Pi^\mu_\alpha := \frac{\partial\mathcal{L}}{\partial_\mu Q^\alpha} \tag{3.23}$$

the expression (3.22) can be written as

$$\delta\mathcal{L} = [\mathcal{L}]_\alpha \bar{\delta} Q^\alpha + \partial_\mu C^\mu \qquad \text{with} \qquad C^\mu = \Pi^\mu_\alpha \bar{\delta} Q^\alpha + \mathcal{L}\delta x^\mu. \tag{3.24}$$

The object C^μ if expressed in terms of the δ variation instead of the $\bar{\delta}$ variation becomes

$$C^\mu = \Pi^\mu_\alpha \delta Q^\alpha - \Theta^\mu{}_\nu \delta x^\nu. \tag{3.25}$$

with the so-called canonical energy-momentum tensor

$$\Theta^\mu_{\ \nu} := \frac{\partial\mathcal{L}}{\partial(\partial_\mu Q^\alpha)}(\partial_\nu Q^\alpha) - \delta^\mu_\nu\ \mathcal{L}. \tag{3.26}$$

This tensor arises, as will be seen later, as a natural object in the context of Noether's theorems in the case of a spacetime translational invariance of a theory. The name 'canonical' stems from the fact that its $\Theta^0{}_0$-component is the Hamiltonian density. As shown below, it also appears in a canonical way in the definition of the angular momentum in the case of a symmetry with respect to Lorentz-transformations. Nevertheless, very often it has been considered to be a formal construct, devoid of any physical meaning.

In order to devise a Hamiltonian formulation of a field theory from its Lagrangian, one first needs a concept of canonical momenta. In case of classical mechanics these are defined as the derivatives of the Lagrange function with respect to the velocities, which themselves are the derivatives of the generalized coordinates with respect to time. In classical mechanics "time" is an undisputedly basic term[13]. However, in relativistic theories this is no longer the case, and one needs to select a "time variable" T. An obvious choice is the coordinate time $T = x^0$. But aside from this "instant" form there are other choices as well, like for example the "front form" (or light-cone time) $T = \frac{1}{\sqrt{2}}(x^0 - x^D)$. In fact there is an infinite choice of possible "time" variables, and later, in Sect. 3.4.5 they will be categorized according to subgroups of the Poincaré group. The definition of the canonical momenta now depends on the choice of T:

$$\Pi^\alpha_{(T)} := \frac{\partial\mathcal{L}}{\partial(\partial_T Q^\alpha)}. \tag{3.27}$$

The canonical momenta are evidently a linear combination of the pseudo-momenta (3.23). Just as in classical mechanics, regular and singular systems are to be distinguished by the determinant of the Hessian. Only if its determinant does not vanish, the implicit-function theorem allows us to express the momenta in terms of the coordinates and the velocities. For field theories, the Hessian becomes infinite dimensional, and to ensure the application of the implicit-function theorem, one needs to define the fields in an appropriate functional space, say Banach spaces. In the regular case,

---

[13] I'm aware that J. Barbour [25] would object to this statement.

the Hamilton density (Hamiltonian, for short) is a function of the fields and their momenta, and is given by

$$H(Q, \Pi_{(T)}) = \sum \int d^{D-1}x\, (\Pi^{\alpha}_{(T)} \partial_T Q^{\alpha} - \mathcal{L}).$$

According to present understanding the four known basic interactions are described by relativistic field theories with local symmetries. This, in turn–and exemplified later–compels their Lagrangians to be singular. Their Hamiltonian formulation needs a specific treatment, and this is the topic of Appendix C.

Poisson brackets of two functionals of the fields and their momenta are defined for "equal times". In case of coordinate time with $P_\alpha := \frac{\partial \mathcal{L}}{\partial(\partial_0 Q^\alpha)}$, for instance:

$$\{A(x), B(y)\}_{x^0=y^0} := \int d^{D-1}z \left(\frac{\delta A(x)}{\delta Q^{\alpha}(z)} \frac{\delta B(y)}{\delta P_{\alpha}(z)} - \frac{\delta B(y)}{\delta Q^{\alpha}(z)} \frac{\delta A(x)}{\delta P_{\alpha}(z)} \right).$$

Because of the functional derivatives appearing in these expressions this obviously only can be applied to functionals for which these derivatives exist. The (non-vanishing) canonical Poisson brackets are

$$\{Q^{\alpha}(x), P_{\beta}(y)\}_{x^0=y^0} = \delta^{\alpha}_{\beta}\, \delta(\vec{x} - \vec{y}). \tag{3.28}$$

Tensorial Form of Electrodynamics

Electrodynamics is not only the prime example of a field theory but also the prime example of a relativistic field theory. In this context "relativistic" means that the theory is Poincaré-invariant. We know that this is the case, because the clash with Galilei invariance was the motivation for Einstein's special relativity postulates. But again it will be manifestly visible if the Maxwell equations are written in the form of (Lorentz) tensors. For this purpose, the six components of the electric and magnetic field strength are arranged in a 4×4 matrix as

$$(F^{\mu\nu}) := \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{pmatrix}. \tag{3.29}$$

Introducing a four-component object $(j^{\mu}) := (c\varrho, \vec{j})$ the continuity equation $\operatorname{div}\vec{\mathrm{j}} + \partial_{\mathrm{t}}\varrho = 0$ can be written as $\partial_{\mu} j^{\mu} = 0$ with a covariant Lorentz vector j^{μ} (i.e. the 4-current). The pair of Maxwell Eqs. (3.1a, 3.1b) can be written in tensor form as

$$\partial_{\mu} F^{\mu\nu} = \frac{1}{c} j^{\nu}, \tag{3.30}$$

and the pair (3.1c, 3.1d) are encoded in

$$F_{\mu\nu,\rho} + F_{\nu\rho,\mu} + F_{\rho\mu,\nu} = 0. \qquad \text{or} \qquad \partial_{[\rho}F_{\mu\nu]} = 0. \tag{3.31}$$

The tensorial expression for the Lorentz force (3.3) is

$$f^{\mu} = \frac{q}{c} F^{\mu}{}_{\nu} u^{\nu}.$$

This neat encoding of "electric" and "magnetic" entities exhibits their unification within a Lorentz-invariant theory. The Eqs. (3.31) are solved by a (Lorentz) vector potential A_{μ}:

$$F_{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu}.$$

In the traditional (pre-Minkowki) formulation of electrodynamics, this corresponds to the fact, that (3.1d), i.e. $\mathrm{div}\vec{\mathrm{B}} = \vec{0}$, is solved by $\vec{B} = \mathrm{rot}\vec{\mathrm{A}}$ with a vector potential $\vec{A}$, which in turn, inserted in (3.1c), has the solution $\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} + \mathrm{grad\,V}$ with a scalar potential V. But neither $\{\vec{A}, V\}$ nor A_{μ} are unique: Since

$$A'_{\mu} = A_{\mu} + \partial_{\mu}\chi \tag{3.32}$$

leaves the field strength tensor unchanged, (3.32) is also a solution of (3.31). The substitution (3.32) is called a *gauge transformation*. In order to arrive at unique solutions for the field equations one needs to break the gauge invariance by imposing gauge conditions. Common gauge conditions are for instance the Lorenz gauge $\partial_{\mu} A^{\mu} = 0$ or the Coulomb gauge $\mathrm{grad}\,\vec{\mathrm{A}} = 0$.

Relativistic Field Theory in Lagrange Formulation

The action obeys the postulates of special relativity if the Lagrange density does not explicitly depend on the coordinates (translational invariance) and if it transforms as a Lorentz scalar (we saw already that the volume element is a scalar). The Lagrangian for the free Maxwell theory for example is

$$\mathcal{L}_{ED} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu}. \tag{3.33}$$

Since in this Lagrangian only derivatives of the vector potential A_{μ} but not the vector potential itself are present, the field equations are

$$[\mathcal{L}_{ED}]^{\mu} := -\partial_{\nu} \frac{\partial \mathcal{L}_{ED}}{\partial(\partial_{\nu} A_{\mu})} = \frac{1}{2}\partial_{\nu}\left[F^{\rho\sigma}\frac{\partial}{\partial(\partial_{\nu} A_{\mu})}(\partial_{\rho} A_{\sigma} - \partial_{\sigma} A_{\rho})\right] = \partial_{\nu} F^{\nu\mu} = 0 \tag{3.34}$$

which indeed reproduces (3.30) in the case of vanishing external currents.

In order to mimic the coupling to charges, let us take as a further example the interaction of a charged boson with the electromagnetic field. As derived later in Chap. 5, a charged boson field is represented by a complex scalar field ϕ. The Lagrange density $\mathcal{L}$ is the sum of three terms

$$\mathcal{L} = \mathcal{L}_{ED} + \mathcal{L}_S + \mathcal{L}_I \tag{3.35}$$

where $\mathcal{L}_{ED}$ is the Lagrangian of the free Maxwell field as before, and

$$\mathcal{L}_S = (\partial_\mu \phi)(\partial^\mu \phi^*) - V(\phi \cdot \phi^*) \tag{3.36}$$

is the Lagrange density of the scalar/source field. Examples are the Klein-Gordon field with $V = m^2(\phi \cdot \phi^*)$ and the Higgs field $V = -\mu(\phi \cdot \phi^*) + \lambda(\phi \cdot \phi^*)^2$ with $(\mu, \lambda) > 0$. The term $\mathcal{L}_I$ describes the interaction of the fields $\{A, \phi\}$. The combination $\mathcal{L}_S + \mathcal{L}_I$ in the Lagrange density (3.35) can formally be obtained by replacing the partial derivatives in $\mathcal{L}_S$ by

$$D_\mu \phi := (\partial_\mu + iqA_\mu)\phi, \qquad D_\mu \phi^* := (\partial_\mu - iqA_\mu)\phi^*, \tag{3.37}$$

where q is the charge of the boson. The meaning of this replacement in terms of symmetry considerations will become clear later. The full Lagrange density (3.35) is thus

$$\mathcal{L} = \mathcal{L}_{ED} + (D_\mu \phi)(D^\mu \phi^*) - V(\phi \cdot \phi^*). \tag{3.38}$$

Obviously each of the three terms in this expression is a Lorentz scalar. The field equations derived from this Lagrange density are

$$[\mathcal{L}]^\mu := \frac{\partial \mathcal{L}}{\partial A_\mu} - \partial_\nu \frac{\partial \mathcal{L}}{\partial(\partial_\nu A_\mu)} = 0 \tag{3.39a}$$

$$[\mathcal{L}]^\phi := \frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0, \tag{3.39b}$$

and analogous ones (actually the complex conjugates) for ϕ^*. With

$$\frac{\partial \mathcal{L}}{\partial A_\mu} = iq(\phi D^\mu \phi^* - \phi^* D^\mu \phi) =: -j_\phi^\mu \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial(\partial_\nu A_\mu)} = -F^{\nu\mu} \tag{3.40}$$

the field Eq. (3.39a) become $[\mathcal{L}]^\mu = -j_\phi^\mu + \partial_\nu F^{\nu\mu} = 0$ and these reproduce (3.30).

3.3 Noether Theorems

The paper "*Invariante Variationsprobleme*" of Emmy Noether (1882–1935) was presented to the July 7th, 1918 meeting of the "*Königliche Gesellschaft der Wissenschaften zu Göttingen*" by F. Klein and published as [393]. It is certain that Noether was not a member of this royal society, and it even can be doubted, that she herself

Fig. 3.3 Emmy Noether

was present at this meeting. How this paper arose in the context of D. Hilbert's attempt at laying the foundations of physics is described in [311]. The Noether theorems–although a main cornerstone in this endeavor–somehow did not really enter the physics community until the 1960's, and then due for instance to the work of A. Trautman [509]. For the reception of Noether's paper see [327].

In the description of the Noether theorems, I follow largely [514]–however with changes in notation. It is fair to say that there is much more in the original Noether article than what these days seems important for physics. Noether investigated the invariance of functionals

$$I = \int \dots \int f\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}, \dots\right) dx \tag{3.41}$$

(where the u_m are functions of the variables x^i) under transformations $x \to x(x)$ and $u \to u'(x, u)$. She investigated the consequences of invariance–but not of quasi-invariance, that is invariance up to a divergence. This was carried out by E. Bessel-Hagen in 1921 [40]. He also generalized Noether's theorem in enlarging the class of symmetries to include transformations depending on the derivatives of the $u(x)$. In the mathematics literature, these go under the name "generalized symmetries" [399] and sometimes under the unfortunate misnomer of "Lie-Bäcklund transformations".

3.3.1 Variational Symmetries in Field Theories

The starting point for deriving the Noether-Bessel-Hagen theorems within a Lagrangian field theory is the condition

$$\int_{\hat{\Omega}} d^4\hat{x}\ \mathcal{L}(\hat{Q}, \partial\hat{Q}, \hat{x}) \stackrel{!}{=} \int_{\Omega} d^4x\ \mathcal{L}(Q, \partial Q, x) \tag{3.42}$$

under a one-parameter group of point transformations

$$\hat{x}^\mu = \hat{x}^\mu(Q, \partial Q, x, \epsilon) = x^\mu + \epsilon\xi^\mu(Q, \partial Q, x) + \mathcal{O}(\epsilon^2)$$
$$\hat{Q}^\alpha = \hat{Q}^\alpha(Q, \partial Q, x, \epsilon) = Q^\alpha + \epsilon\eta^\alpha(Q, \partial Q, x) + \mathcal{O}(\epsilon^2).$$

In principle, one could go through the chain of calculations for deriving an expression similar to (2.48). We have, however, already obtained the result of an arbitrary variation of the Lagrangian as in (3.24):

$$\delta\mathcal{L} = [\mathcal{L}]_\alpha \bar{\delta} Q^\alpha + \partial_\mu C^\mu$$

with C^μ expressed either by the δ or the $\bar{\delta}$ variation:

$$C^\mu = \Pi^{\alpha\mu}\bar{\delta}Q^\alpha + \mathcal{L}\delta x^\mu = \Pi^{\alpha\mu}\delta Q^\alpha - \Theta^\mu{}_\nu \delta x^\nu.$$

For a symmetry transformation, $\delta_S\mathcal{L} = \partial_\mu \Sigma^\mu_S$ must hold, and the equalization

$$\partial_\mu \Sigma^\mu_S = [\mathcal{L}]_\alpha \bar{\delta}_S Q^\alpha + \partial_\mu C^\mu(\delta \to \delta_S)$$

results in

$$0 = \Delta_S\mathcal{L} = [\mathcal{L}]_\alpha \bar{\delta}_S Q^\alpha + \partial_\mu J^\mu_S \tag{3.43}$$

with

$$J^\mu_S = \Pi^{\alpha\mu}\bar{\delta}_S Q^\alpha + \mathcal{L}\delta_S x^\mu - \Sigma^\mu_S = \Pi^{\alpha\mu}\delta_S Q^\alpha - \Theta^\mu{}_\nu \delta_S x^\nu - \Sigma^\mu_S. \tag{3.44}$$

This is clearly the generalization of (2.48) to a field theory. In some contexts we will later use (3.43) as the identity

$$\frac{\delta S}{\delta Q^\alpha}\bar{\delta}_S Q^\alpha \equiv \Delta_S\mathcal{L} - \partial_\mu J^\mu_S. \tag{3.45}$$

Let me give another, more abstract, proof of Noether's theorem(s) using the concept of the infinitesimal symmetry generator

$$\overline{X} = \xi^\mu \frac{\partial}{\partial x^\mu} + \eta^\alpha \frac{\partial}{\partial Q^\alpha} + \eta^\alpha_\mu \frac{\partial}{\partial Q^\alpha_{,\mu}} + \ldots \qquad \text{with} \qquad \eta^\alpha_\mu = d_\mu \eta^\alpha - Q^\alpha_{,\nu}\, d_\mu \xi^\nu$$

or alternatively, with the understanding that $d_\mu = \frac{d}{dx^\mu}$,

$$\overline{X} = \xi^\mu d_\mu + \chi^\alpha \frac{\partial}{\partial Q^\alpha} + (d_\mu \chi^\alpha)\frac{\partial}{\partial Q^\alpha_{,\mu}} + \ldots \qquad \text{with} \qquad \chi^\alpha = \eta^\alpha - Q^\alpha_{,\mu}\xi^\mu.$$

The left-hand side of (3.42) can be written as

$$\int_{\hat{\Omega}} d^4\hat{x}\ \mathcal{L}(\hat{Q}, \partial\hat{Q}, \hat{x}) = \int_{\Omega} d^4x\ J\,\mathcal{L}(\hat{Q}, \partial\hat{Q}, \hat{x})$$

where J is the Jacobian of the transformation $x \to \hat{x}$:

$$J = \det\left|\frac{d\hat{x}^\mu}{dx^\nu}\right| = \det|\nabla_\nu \hat{x}^\mu|.$$

With $J = 1$ for $\epsilon = 0$, we get $J = 1 + \epsilon\nabla_\mu \xi^\mu + \mathcal{O}(\epsilon^2)$. Furthermore,

$$\mathcal{L}(\hat{Q}, \partial\hat{Q}, \hat{x}) = \mathcal{L}(Q, \partial Q, x) + \epsilon\left[\frac{\partial\mathcal{L}}{\partial Q^\alpha}\eta^\alpha + \frac{\partial\mathcal{L}}{\partial Q^\alpha_{,\mu}}\eta^\alpha_\mu + \frac{\partial\mathcal{L}}{\partial x^\mu}\xi^\mu\right] + \mathcal{O}(\epsilon^2).$$

Therefore,

$$0 = \int_\Omega d^4x\ \left[J\mathcal{L}(\hat{Q}, \partial\hat{Q}, \hat{x}) - \mathcal{L}(Q, \partial Q, x)\right] = \epsilon\int_\Omega d^4x\left[X\mathcal{L} + \mathcal{L}d_\mu\xi^\mu\right] + \mathcal{O}(\epsilon^2),$$

and thus up to order ϵ, the condition (3.42) becomes

$$0 = \int_\Omega d^4x\ (\eta^\alpha - Q^\alpha_{,\mu}\xi^\mu)\,[\mathcal{L}]_\alpha + \int_\Omega d^4x\ d_\mu\left(\chi^\alpha\frac{\partial\mathcal{L}}{\partial Q^\alpha_{,\mu}} + \mathcal{L}\xi^\mu\right). \tag{3.46}$$

Allowing the existence of a surface term, the condition for X being a generator for a Noether symmetry becomes

$$X\mathcal{L} + \mathcal{L}d_\mu\xi^\mu = d_\mu\Sigma^\mu,$$

which is the generalization of (2.68) from classical mechanics. From (3.46), the two Noether theorems follow:

- Theorem 1: If functions $\xi^\mu_r(x, Q, \partial Q)$, $\eta^\alpha_r(x, Q, \partial Q)$, $\Sigma^\mu_r(x, Q, \partial Q)$ exist, such that for arbitrary constant parameters $\epsilon^1, \epsilon^2, \cdots, \epsilon^N$

  $$\delta x^\mu = \xi^\mu = \xi^\mu_r\,\epsilon^r, \qquad \delta Q^\alpha = \eta^\alpha = \eta^\alpha_r\,\epsilon^r, \qquad \Sigma^\mu = \sigma^\mu_r\,\epsilon^r,$$

  then N identities exist of the form

  $$(\eta^\alpha_r - Q^\alpha_{,\mu}\xi^\mu_r)\,[\mathcal{L}]_\alpha = \nabla_\mu\left(\chi^\alpha_r\frac{\partial\mathcal{L}}{\partial Q^\alpha_{,\mu}} + \mathcal{L}\xi^\mu_r - \sigma^\mu_r\right).$$

  Thus there are N on-shell conserved currents $j^\mu_r = \chi^\alpha_r\frac{\partial\mathcal{L}}{\partial Q^\alpha_{,\mu}} + \mathcal{L}\xi^\mu_r - \Sigma^\mu_r$. (Strictly speaking, in these currents only those α-terms contribute to the sum for which the Q^α are dynamical fields.)

- Theorem 2: If linear operators $\hat{\xi}^\mu_r$, $\hat{\eta}^\alpha_r$, $\hat{\Sigma}^\mu_r$ exist, such that for arbitrary functions ϵ^r which vanish on the boundary $\partial\Omega$

$$\delta x^\mu = \hat{\xi}^\mu_r \cdot \epsilon^r, \qquad \delta Q^\alpha = \hat{\eta}^\alpha_r \cdot \epsilon^r, \qquad \Sigma^\mu = \hat{\sigma}^\mu_r \cdot \epsilon^r,$$

then N identities exist of the form

$$\hat{\eta}^{*\alpha}_r \cdot [\mathcal{L}]_\alpha - \hat{\xi}^{*\mu}_r \cdot \left(Q^\alpha_{,\mu}\,[\mathcal{L}]_\alpha\right) = 0$$

where the $*$ operators are the adjoints of the un-starred ones. These identities reveal that the field equations are not independent of each other.

Both Noether theorems are detailed in the following two subsections.

3.3.2 Global Symmetries and 1st Noether Theorem

From the original [393]: "*Ist das Integral I invariant gegenüber einer $\mathcal{G}_\rho$, so werden ρ linear-unabhängige Verbindungen der Lagrangeschen Ausdrücke zu Divergenzen–umgekehrt folgt daraus die Invarianz von I gegenüber einer $\mathcal{G}_\rho$. Der Satz gilt auch im Grenzfall von unendlich vielen Parametern.*"[14] Here E. Noether refers to the integral (3.41).

Global Symmetries and Conserved Charges

In 3.3.1., the first Noether theorem was stated for symmetry variations of the form $\delta_\epsilon x^\mu = \xi^\mu_r(x, Q, \partial Q)\,\epsilon^r$, $\delta_\epsilon Q^\alpha = \eta^\alpha_r(x, Q, \partial Q)\,\epsilon^r$ with constant parameters ϵ^r. In the following we restrict ourselves to symmetry transformations

$$\delta_\epsilon x^\mu = \mathcal{D}^\mu_r(x)\,\epsilon^r \qquad\qquad \delta_\epsilon Q^\alpha = \mathcal{A}^\alpha_r(Q)\,\epsilon^r. \tag{3.47}$$

(actually Noether's index ρ is called r here). But even in (3.47), the functions $\mathcal{D}^\mu_r$ and $\mathcal{A}^\alpha_r$ are not arbitrary because of the required Lie-group property. More about this consistency condition in Sect. 3.3.4.

Since the ϵ^r are constants, the identities (3.43) become

$$[\mathcal{L}]_\alpha(\mathcal{A}^\alpha_r - Q^\alpha_{,\mu}\mathcal{D}^\mu_r) + \partial_\mu j^\mu_r = 0 \qquad (r = 1, \ldots, R) \tag{3.48}$$

[14] If the integral I is invariant with respect to a [group] $\mathcal{G}_\rho$, then ρ linearly independent combinations of the Lagrange expressions become divergences - and from this, conversely, invariance of I with respect to a [group] $\mathcal{G}_\rho$, will follow. The theorem holds good even in the limiting case of infinitely many parameters.

with

$$j_r^\mu := \Pi^{\alpha\mu}\mathcal{A}_r^\alpha - \Theta^\mu{}_\nu \mathcal{D}_r^\nu - \sigma_r^\mu . \tag{3.49}$$

Here it is understood that the surface term is also expanded: $\Sigma_S^\mu = \sigma_r^\mu \epsilon^r$. As a consequence of (3.48), on-shell (that is for $[L]_\alpha = 0$) there are R conserved Noether currents j_r^μ:

$$\partial_\mu j_r^\mu = \partial_\mu(\Pi^{\alpha\mu}\mathcal{A}_r^\alpha - \Theta^\mu{}_\nu \mathcal{D}_r^\nu - \sigma_r^\mu) = -[\mathcal{L}]_\alpha(\mathcal{A}_r^\alpha - Q^\alpha_{,\mu}\mathcal{D}_r^\mu) \doteq 0. \tag{3.50}$$

The previous sentence needs some cautious supplementary comments: (i) As mentioned before, only those terms proportional to $\mathcal{A}_r^\alpha$ contribute to the sum for which the fields Q^α are dynamical. (ii) The currents might be either trivial numerical constants or vanish themselves on-shell. (iii) The currents are not unique, since any current

$$\tilde{j}_r^\mu = j_r^\mu + \partial_\nu k_r^{\mu\nu}$$

with antisymmetric tensors $k_r^{\mu\nu}$ also obeys the conservation law $\partial_\mu \tilde{j}_r^\mu = 0$. This indeterminacy can be used to redefine the current appropriately, as will be described in more detail for the energy-momentum tensor at the end of this subsection. Since the currents are conserved only on-shell, one sometimes also finds the term 'weak' conservation laws in order to distinguish these from 'strong' conservation arising from divergence identities. The literature, however, is not unequivocal in this point. Concerning the different meanings of these terms, and the terminology 'proper' and 'improper' conservation laws as introduced by D. Hilbert and by E. Noether herself compared to uses by P. Bergmann and A. Trautman, see [55].

Under some moderate conditions on the Q^α, the (non-trivially) conserved currents (3.49) imply the existence of conserved charges

$$\frac{d}{dt}C_r = 0 \qquad \text{with} \qquad \mathrm{C_r} := \int_\Omega \mathrm{d^3x\, j_r^0}.$$

What is meant by "moderate" conditions can be derived from

$$0 = \int_\Omega d^3x\ \partial_\mu j^\mu = \int_\Omega d^3x\ \partial_0 j^0 + \int_\Omega d^3x\ \partial_i j^i = \frac{d}{dt}\int_\Omega d^3x\ j^0 + \oint_{\partial\Omega} dS\ \hat{n}^i_{ij},$$

where in the last step, the Gauß integral theorem was used to convert a volume integral into a surface integral. This integral vanishes if the j^i fall off sufficiently rapidly: If the j^i decrease with increasing distance r in such a way that the net outflow from the region enclosed by Ω decreases faster than the volume increases, the net outflow across $\partial\Omega$ vanishes for sufficiently large r. Notice that although the on-shell conserved currents are non-unique, the charges which are contingently implied are unique.

Algebra of Noether Charges

The Poisson brackets of the Noether charges C_r exhibit a structure which is directly related to the algebra of the underlying Lie group. This can be derived as follows: Start from the Poisson bracket $\{Q^\alpha, \epsilon^r C_r\}$, or explicitly:

$$\begin{aligned}\left\{Q^\alpha(x), \epsilon^r \int d^3y\, j_r^0(y)\right\} &= \epsilon^r \int d^3y \left\{Q^\alpha(x), \left(\Pi^{\alpha 0}(y)\mathcal{A}_r^\alpha - \Theta^0_{\ \nu}(y)\mathcal{D}_r^\nu\right)\right\} \\ &= \epsilon^r \left(\mathcal{A}_r^\alpha - \mathcal{D}_r^\nu(\partial_\nu Q^\alpha)\right) = \delta_S Q^\alpha - (\partial_\nu Q^\alpha)\delta_S x^\nu.\end{aligned}$$

Since the last term is simply $\bar{\delta} Q^\alpha$, we find

$$\bar{\delta} Q^\alpha = \{Q^\alpha, \epsilon^r C_r\}. \tag{3.51}$$

Next, calculate the commutator of two $\bar{\delta}$ transformations:

$$\left(\bar{\delta}_2\bar{\delta}_1 - \bar{\delta}_1\bar{\delta}_2\right) Q^\alpha = \epsilon_1^r\epsilon_2^s \Big[\{\{Q^\alpha, C_r\}, C_s\} - \{\{Q^\alpha, C_s\}, C_r\}\Big] = -\epsilon_1^r\epsilon_2^s\{\{C_r, C_s\}, Q^\alpha\},$$

where in the last step the Jacobi identity for Poisson brackets was used. We require that

$$\bar{\delta}_3 Q^\alpha = \{Q^\alpha, \epsilon_3^t C_t\} = -\epsilon_1^r\epsilon_2^s\{\{C_r, C_s\}, Q^\alpha\},$$

and this is leads to

$$\{C_r, C_s\} = f_{rst} C_t + \gamma_{rs}. \tag{3.52}$$

Here f_{rst} are the structure constants of the algebra and the γ_{rs} are constants (in terms of Lie algebras to be interpreted as central charges; see Appendix A.3.5).

Examples

Example I: Energy-Momentum Conservation

The consequences of the first Noether theorem with respect to the currently well-established Poincaré symmetry of field theories leads directly to ten conserved charges, related to the generators of the Poincaré group.
(i) For infinitesimal translations $\delta x^\mu = \epsilon^\mu$ with its four parameters ϵ^μ, we read off from (3.47) that $\mathcal{D}_r^\mu = \delta_r^\mu$ and $\delta Q^\alpha = 0$. Therefore, the corresponding $\mathcal{A}$ vanish identically, with the consequence that

$$\bar{\delta}_\epsilon Q_\alpha = -\epsilon^\mu \partial_\mu Q_\alpha. \tag{3.53}$$

The Noether current (3.49) belonging to spacetime translational invariance is just the canonical energy momentum tensor $\Theta^{\mu}{}_{\nu}$ itself. It is weakly conserved:

$$\partial_{\mu}\Theta^{\mu}{}_{\nu} \doteq 0. \tag{3.54}$$

The integrated expressions, or the *four-momenta*

$$T^{\mu} := \int d^3x\,\Theta^{0\mu} \tag{3.55}$$

are conserved over time, provided that $\Theta^{i\mu}$ has an appropriate boundary behavior, as discussed before.

(ii) For infinitesimal Lorentz transformations $\delta x^{\mu} = \omega^{\mu}{}_{\nu}x^{\nu}$ with its six parameters $\omega_{\mu\nu} + \omega_{\nu\mu} = 0$, we read off from (3.47) that

$$\mathcal{D}^{\mu}_{[\rho\sigma]} = -x_{[\rho}\delta^{\mu}_{\sigma]}.$$

In a Lorentz-invariant theory, the fields Q transform as Lorentz tensors (this will be made more specific in Sect. 5.3):

$$\hat{Q}^{\alpha}(\hat{x}) = D^{\alpha}{}_{\beta}Q^{\beta}(x) = \left(\delta^{\alpha}_{\beta} - \frac{i}{2}\omega_{\mu\nu}(\Sigma^{\mu\nu})^{\alpha}{}_{\beta}\right)Q^{\beta} \tag{3.56}$$

where D is a representation matrix of the Lorentz group, and $(\Sigma^{\mu\nu})$ is the *spin matrix*. The comparison with (3.47) yields

$$\mathcal{A}^{\alpha}_{[\rho\sigma]} = -(i\Sigma_{\rho\sigma})^{\alpha}{}_{\beta}Q^{\beta}.$$

The Noether current corresponding to (3.49) is the angular momentum tensor

$$M^{\mu}{}_{\rho\sigma} := (\Theta^{\mu}{}_{\rho}x_{\sigma} - \Theta^{\mu}{}_{\sigma}x_{\rho}) + S^{\mu}{}_{\rho\sigma} \tag{3.57}$$

with the *canonical spin tensor*

$$S_{\mu\rho\sigma} := -\frac{\partial\mathcal{L}}{\partial Q^{\alpha}_{,\mu}}(i\Sigma_{\rho\sigma})^{\alpha}{}_{\beta}Q^{\beta}. \tag{3.58}$$

Now the conservation law $\partial_{\mu}M^{\mu}{}_{\rho\sigma} \doteq 0$ implies together with (3.54) that

$$\partial_{\mu}S^{\mu\rho\sigma} \doteq \Theta^{\sigma\rho} - \Theta^{\rho\sigma}. \tag{3.59}$$

The Noether charges defined by

$$M^{\rho\sigma} := \int d^3x\, M^{0\rho\sigma} = \int d^3x\,(x^{\rho}\,\Theta^{0\sigma} - x^{\sigma}\,\Theta^{0\rho}) + \int d^3x\, S^{0\rho\sigma}$$

are the *angular momenta*. These are conserved, assuming appropriate boundary conditions for $M^i{}_{\rho\sigma}$.

One could be tempted to interpret the first term as an orbital part and the second as a spin part. Such an interpretation, however, is doubtful because the energy-momentum and spin distribution can be "relocalized" by adding curls (in other contexts of this book also called superpotentials),

$$\hat{\Theta}^{\mu\nu} = \Theta^{\mu\nu} + \partial_\lambda Y^{\mu\nu\lambda} \qquad \hat{S}^{\mu\rho\sigma} = S^{\mu\rho\sigma} + Y^{[\mu\rho]\sigma} + \partial_\lambda Z^{\mu\rho\sigma\lambda}$$

with the index antisymmetries $Y^{\mu\nu\lambda} = -Y^{\mu\lambda\nu}$ and $Z^{\mu\rho\sigma\lambda} = -Z^{\mu\rho\lambda\sigma} = -Z^{\rho\mu\sigma\lambda}$. These redefined tensors obey the conservation laws (3.54, 3.59) as well: $\partial_\mu \hat{\Theta}^{\mu\nu} \doteq 0$, $\partial_\mu \hat{S}^{\mu\rho\sigma} \doteq \hat{\Theta}^{\sigma\rho} - \hat{\Theta}^{\rho\sigma}$. By a specific choice of the superpotentials one can achieve that $\hat{S}^{\mu\rho\sigma} = 0$, namely by defining the *Belinfante tensor*

$$B^{\mu\nu} = \Theta^{\mu\nu} + \frac{1}{2}\partial_\lambda\left(S^{\mu\nu\lambda} + S^{\nu\mu\lambda} - S^{\lambda\nu\mu}\right).$$

This redefinition of the energy-momentum tensor, due to F.J. Belinfante [42], ensures that the angular momenta (and the four-momenta) depend only on this new energy-momentum tensor:

$$T^\mu = \int d^3x\, B^{0\mu} \tag{3.60a}$$

$$M^{\mu\nu} = \int d^3x\,(x^\mu B^{0\nu} - x^\nu B^{0\mu}). \tag{3.60b}$$

Observe that now – in contrast to the case of the canonical energy-momentum tensor – the angular-momentum has the structure of an orbital part. Indeed up to today there is a general discussion whether or not there is a physically reasonable way to split the total angular momentum of a photon (or – in a broader context – a gluon) into spin and orbital contributions [335].

The Belinfante tensor has the virtue of being symmetric ($B^{\mu\nu}$-$B^{\nu\mu} \doteq 0$) (because of the vanishing spin tensor and the conservation law (3.59)), in contrast to the canonical energy-momentum tensor. Since a symmetric energy-momentum tensor is the source of gravitation in Einstein's general relativity theory, the Belinfante tensor is by many called the "improved" energy-momentum tensor. In Sect. 7.5 it will be shown that the Belinfante tensor is the flat space limit of the Hilbert matter tensor, a tensor defined within general relativity for minimally coupled relativistic matter fields.

The canonical energy-momentum tensor for a scalar field from the Lagrangian

$$\mathcal{L}_\varphi = \frac{1}{2}\partial_\mu\varphi\,\partial^\mu\varphi - V(\varphi) \tag{3.61}$$

is immediately derived as

$$\Theta_{\varphi}^{\mu\nu} = \partial^{\mu}\varphi\partial^{\nu}\varphi - \frac{1}{2}\eta^{\mu\nu}\partial_{\lambda}\varphi\partial^{\lambda}\varphi + \eta^{\mu\nu}V(\varphi). \tag{3.62}$$

This tensor is already symmetric. There is actually no need to "improve" this tensor. And indeed, since the canonical spin tensor vanishes for a scalar field, the canonical energy-momentum tensor and the Belinfante tensor are identical. One finds that its divergence can be expressed as

$$\partial_{\mu}B_{\varphi}^{\mu\nu} = -(\partial^{\nu}\varphi)\,[L]_{\varphi}\,.$$

which of course is compatible with the on-shell conservation of the current.

For source-free electrodynamics, one derives the canonical energy-momentum tensor

$$\Theta_{\mathrm{ED}}^{\mu\nu} = -F^{\mu\lambda}\partial^{\nu}A_{\lambda} + \frac{1}{4}\eta^{\mu\nu}\,F^{\varrho\sigma}F_{\varrho\sigma}. \tag{3.63}$$

In order to calculate the electromagnetic Belinfante tensor with the canonical spin tensor we need the spin matrix

$$(i\,\Sigma^{\mu\nu})_{\rho}{}^{\sigma} = \delta_{\rho}^{\mu}\eta^{\nu\sigma} - \delta_{\rho}^{\nu}\eta^{\mu\sigma},$$

and thus from (3.58) $S^{\mu\nu\lambda} = F^{\lambda\nu}A^{\mu} - F^{\lambda\mu}A^{\nu}$, finally leading to

$$B_{\mathrm{ED}}^{\mu\nu} = -F^{\mu}{}_{\lambda}F^{\nu\lambda} + \frac{1}{4}\eta^{\mu\nu}\,F^{\varrho\sigma}F_{\varrho\sigma}. \tag{3.64}$$

This tensor is identical with the electromagnetic tensor introduced by Minkowski [372] and describing observables, such as the energy density $B_{ED}^{00} = \frac{1}{2}(\vec{E}^2 + \vec{B}^2)$, the Poynting vector $B_{ED}^{i0} = (\vec{E}\times\vec{B})^i$, and the Maxwell stress tensor B_{ED}^{ij}. Further, the conserved charges become

$$\mathcal{E} = \frac{1}{2}\int_{\Omega} d^3x(E^2 + B^2),\quad \vec{\mathcal{P}} = \int_{\Omega} d^3x\,\vec{E}\times\vec{B},\quad \vec{\mathcal{J}} = \int_{\Omega} d^3x\,\vec{x}\times(\vec{E}\times\vec{B}),$$
$$\vec{\mathcal{Q}} = \int_{\Omega} d^3x\left[\frac{\vec{x}}{2}(E^2 + B^2) - t(\vec{E}\times\vec{B})\right]$$

which are the energy, the momentum, the angular momentum and the energy center of the electromagnetic field. Aside from being symmetric and reproducing the standard definitions of energy density and momentum density of the electromagnetic field, the Belinfante tensor (3.64) has the eminent property of being gauge-invariant: making the change $A_{\mu} \to A_{\mu} + \partial_{\mu}\epsilon$ leaves B_{ED} untouched. Indeed, a straightforward derivation of the Belinfante tensor is possible by assuming Poincaré and gauge invariance of the theory [377]. Similar to the case of the scalar field, we find an on-shell expression for the electromagnetic Belinfante tensor:

$$\partial_\mu B^{\mu\nu}_{\text{ED}} = -F^{\nu\mu}\,[L]_\mu\,.$$

It can be shown (see e.g. [420]) that for a Poincaré-invariant theory with $\xi^\mu := \delta x^\mu = \epsilon^\mu + \omega^\mu{}_\nu x^\nu$, the current (3.44) can always be expressed by the Belinfante tensor as

$$J^\mu_\xi = -B^{\mu\nu}\xi_\nu. \tag{3.65}$$

This current is conserved on-shell, revealing that translational invariance is synonymous with $\partial_\mu B^{\mu\nu} = 0$, and that Lorentz invariance derives from the symmetry of the Belinfante tensor. At the same time the current (3.65) exposes neatly the structure of the corresponding Noether charges (3.60).

Example II: Maxwell Current

The Lagrange density (3.38) for the interaction of a charged scalar boson with an electromagnetic field is invariant under

$$\phi \mapsto \phi e^{iq\epsilon} \quad \phi^* \mapsto \phi^* e^{-iq\epsilon} \tag{3.66}$$

because the fields ϕ and ϕ^* always occur pairwise in products. The coefficients in (3.47) are $\mathcal{D} = 0$ and $\mathcal{A}_\phi = iq\phi, \quad \mathcal{A}_{\phi^*} = -iq\phi^*$. The Noether current associated to this global symmetry is

$$j^\mu = \frac{\partial\mathcal{L}}{\partial_\mu\phi}(iq\phi) + c.c. = iq(\phi D^\mu\phi^* - \phi^* D^\mu\phi).$$

We verify the identity $\partial_\mu j^\mu \equiv iq(\phi^*[\mathcal{L}]^{\phi^*} - \phi[\mathcal{L}]^\phi)$ corresponding to (3.50). Notice that the Noether current is identical to the current j^μ_ϕ (3.40) from the field equations. Therefore, we can write

$$j^\mu = (j^\mu_\phi + \partial_\nu F^{\nu\mu}) - \partial_\nu F^{\nu\mu} = [\mathcal{L}]^\mu - \partial_\nu F^{\nu\mu}.$$

Thus, on-shell, $j^0 \doteq \partial_k F^{0k} = \operatorname{div}\vec{\mathrm{E}}$. From this, the conserved quantity

$$C = \int_\Omega d^3x\, j^0 = \int_\Omega d^3x\,(\operatorname{div}\vec{\mathrm{E}}) = \int_{@\Omega} \mathrm{d}\vec{\mathrm{S}}\cdot\vec{\mathrm{E}} \tag{3.67}$$

is derived. It is the charge inside the volume Ω.

3.3.3 Local Symmetries and 2nd Noether Theorem

From the original [393]: "*Ist das Integral I invariant gegenüber einer $\mathcal{G}_{\infty\rho}$, in der die willkürlichen Funktionen bis zur σ ten Ableitung auftreten, so bestehen ρ identische Relationen zwischen den Lagrangeschen Ausdrücken und ihren Ableitungen bis zur σ ten Ordnung; auch hier gilt die Umkehrung.*"[15]

As already mentioned in 3.3.1, the second Noether theorem deals with situations in which the R parameters ϵ^r of the symmetry group are functions instead of constants: $\delta_\epsilon x^\mu = \hat{\xi}^\mu_r \cdot \epsilon^r$, $\delta_\epsilon Q^\alpha = \hat{\eta}^\alpha_r \cdot \epsilon^r$ with linear operators $\hat{\xi}^\mu_r$ and $\hat{\eta}^\alpha_r$. Here we will restrict ourselves to transformations of the form[16]

$$\delta_\epsilon x^\mu = \mathcal{D}^\mu_r(x)\,\epsilon^r(x) + \ldots, \tag{3.68a}$$

$$\delta_\epsilon Q^\alpha = \mathcal{A}^\alpha_r(Q)\,\epsilon^r(x) + \mathcal{B}^{\alpha\mu}_r(Q)\,\epsilon^r_{,\mu}(x) + \ldots \tag{3.68b}$$

As indicated by the dots, terms with higher derivatives in the parameters could be present (Noether: "…up to the σ-th derivative"). In the following, it is assumed that there are only terms with at most first derivatives in the ϵ^r. This is the case for Yang-Mills theories and for general relativity. How a further term with second derivatives would influence the Noether identity is derived in Sect. 3.3.4. Further generalizations are conceivable in the sense of allowing coefficients depending on derivatives of the fields. And, as stated already for global symmetries, the coefficients in (6.68) necessarily obey differential equations due to the group property of the transformations; this will be detailed also in Sect. 3.3.4.

The starting point of the following considerations is once more the invariance condition (3.43)

$$\Delta_\epsilon \mathcal{L} = [\mathcal{L}]_\alpha \bar{\delta}_\epsilon Q^\alpha + \partial_\mu J^\mu_\epsilon \equiv 0 \tag{3.69}$$

together with the Noether current (3.44). It turns out to be advantageous to express the J^μ_S by current 'components' j^μ_r and $k^{\mu\nu}_r$ as

$$J^\mu_\epsilon = j^\mu_r \epsilon^r + k^{\mu\nu}_r \partial_\nu \epsilon^r = [j^\mu_r - \partial_\nu k^{\mu\nu}_r]\epsilon^r + \partial_\nu(k^{\mu\nu}_r \epsilon^r). \tag{3.70}$$

Inserting the transformations (3.68) into the invariance condition (3.69), the resulting expression has terms with ϵ^r, $\partial_\mu \epsilon^r$, $\partial_\mu\partial_\nu\epsilon^r$:

$$\big([\mathcal{L}]_\alpha \bar{A}^\alpha_r + \partial_\mu j^\mu_r\big)\epsilon^r + \big([\mathcal{L}]_\alpha \mathcal{B}^{\alpha\mu}_r + j^\mu_r + \partial_\nu k^{\nu\mu}_r\big)\epsilon^r_{,\mu} + k^{\mu\nu}_r \epsilon^r_{,\mu\nu} \equiv 0, \tag{3.71}$$

[15] If the integral I is invariant with respect to a $\mathcal{G}_{\infty\rho}$ in which the arbitrary functions occur up to the σ-th derivative, then there exist ρ identity relationships between the Lagrange expressions and their derivatives up to the σ-th order. In this case also, the converse holds.

[16] These transformations could be called 'fibre preserving' because the new coordinates $\hat{x}$ only depend on the old coordinates and not on the fields Q.

where

$$\bar{\mathcal{A}}_r^\alpha := \mathcal{A}_r^\alpha - \mathcal{Q}^\alpha_{,\mu}\mathcal{D}_r^\mu.$$

In the case of a global symmetry, the last two terms vanish because of $\epsilon^r_{,\mu} = 0 = \epsilon^r_{,\mu\nu}$. Then the vanishing of the term in front of ϵ^r leads us back to (3.48). For local symmetry transformations, not only the coefficients preceding the ϵ^r are required to vanish, but also those in front of their first and second derivatives in (3.71) have to vanish separately, since the identity holds for arbitrary functions ϵ^r. This gives rise to three sets of identities:

$$k_r^{\mu\nu} + k_r^{\nu\mu} \equiv 0. \tag{3.72a}$$

$$[\mathcal{L}]_\alpha \mathcal{B}_r^{\alpha\mu} + j_r^\mu - \partial_\nu k_r^{\mu\nu} \equiv 0 \tag{3.72b}$$

$$[\mathcal{L}]_\alpha \bar{\mathcal{A}}_r^\alpha + \partial_\mu j_r^\mu \equiv 0, \tag{3.72c}$$

where the first two sets do not exist in the case of global symmetries. The identities (3.72c) reveal that the current components j_r^μ are themselves conserved on-shell:

$$\partial_\mu j_r^\mu \equiv -[\mathcal{L}]_\alpha \bar{\mathcal{A}}_r^\alpha \doteq 0. \tag{3.73}$$

The identities (3.72b) tell us that

$$j_r^\mu \doteq \partial_\nu k_r^{\mu\nu}.$$

The Noether current (3.70) can then by using (3.72b) be written as

$$J_\epsilon^\mu = -[\mathcal{L}]_\alpha \mathcal{B}_r^{\alpha\mu}\epsilon^r + \partial_\nu(k_r^{\mu\nu}\epsilon^r), \tag{3.74}$$

that is as a term vanishing on-shell and a divergence. The expression $(k_r^{\mu\nu}\epsilon^r)$ which on-shell gives the Noether-current may be called "superpotential"; this terminology is borrowed from attempts to define local conservation laws in general relativity.

The two sets of identities (3.72c) and (3.72b) together imply

$$\mathcal{N}_r = [\mathcal{L}]_\alpha(\mathcal{A}_r^\alpha - \mathcal{Q}^\alpha_{,\mu}\mathcal{D}_r^\mu) - \partial_\mu([\mathcal{L}]_\alpha \mathcal{B}_r^{\alpha\mu}) \equiv 0. \tag{3.75}$$

These are the R "identity relationships between the Lagrange expressions and their derivatives" to which Noether refers. From them, the field equations $[\mathcal{L}]_\alpha = 0$ are not independent of each other, which causes trouble in the initial-value problem. Today, the relations (3.75) are called "Noether identities", or sometimes "generalized Bianchi identities", because in general relativity–as you will see in Sect. 7.5.3–they correspond to the (Riemann geometry-based) contracted Bianchi identities. And when Noether's second theorem is mentioned, one mostly has these identities in mind. The importance of the further identities (3.72a) became apparent only in recent years in the aim to give a meaning to Noether charges in gravitational theories. They were discovered by F. Klein [320] for general relativity (for some historical remarks see [413]), investigated also by J. Goldberg [229], exhibited by R. Utiyama [514], mentioned by A. Trautman [510]–all of them not citing F. Klein. (I found the first reference to Klein in [24].) The identities were called Noether's third theorem in [56], see also [55], cascade equations in [306], and Klein identities in [413]. I have chosen to call them Klein-Noether identities in this book. In these relations, a prominent role is played by the currents

$$\Upsilon_r^\mu := [\mathcal{L}]_\alpha \mathcal{B}_r^{\alpha\mu} + j_r^\mu, \tag{3.76}$$

since the sequence of Klein-Noether identities can equivalently be written as

$$\partial_\mu \Upsilon_r^\mu \equiv 0 \qquad \Upsilon_r^\mu - \partial_\nu k_r^{\mu\nu} \equiv 0 \qquad k_r^{\mu\nu} + k_r^{\nu\mu} \equiv 0. \tag{3.77}$$

In this notation, one readily sees that these relations are not independent of each other: the first (R) relations follow from the second set of $(D \times R)$ relations if we use the inex symmetry expressed in the last set of $(D \times D \times R)$ relations. Notice that the currents Υ_r^μ are identically conserved and that the currents j_r^μ are conserved on-shell only.

The explicit expressions for j_r^μ and $k_r^{\mu\nu}$ depend on the coefficients $\mathcal{A}, \mathcal{B}, \mathcal{D}$ in the symmetry transformations (3.68) and on possible surface terms. In expanding these as

$$\Sigma_S^\mu = \sigma_r^\mu \epsilon^r + \rho_r^{\mu\nu} \partial_\nu \epsilon^r$$

we obtain from (3.44):

$$j_r^\mu = \Pi^{\alpha\mu} \bar{\mathcal{A}}_r^\alpha + \mathcal{L}\mathcal{D}_r^\mu - \sigma_r^\mu \tag{3.78a}$$

$$k_r^{\mu\nu} = \Pi^{\alpha\mu} \mathcal{B}_r^{\alpha\nu} - \rho_r^{\mu\nu}. \tag{3.78b}$$

Are there charges accompanied by (on-shell) conserved currents? We found that the currents j_r^μ are conserved on-shell (3.73). Provided that the spatial components j^k fall off sufficiently rapidly, the quantities

$$C_r = \int_V d^3x\, j_\epsilon^0 = \int_{\partial V} k_r^{i0}\, dS_i \tag{3.79}$$

are conserved. Observe that in this case of a local symmetry, the charges can be expressed as a surface integral by the use of the on-shell identity $j_r^\mu \doteq \partial_\nu k_r^{\mu\nu}$ and by the use of Stoke's theorem. (In the language of general relativity the charges may be called "quasi-local".) As we will see, in the case of Yang-Mills theory and for general relativity these charges are not necessarily gauge invariant or generally covariant, respectively.

Aside from the currents j_r, we also have the on-shell conserved Noether currents at our disposal, and we might be tempted to define "charges"

$$C_\epsilon = \int_V d^3x\, J_r^0 = \int_{\partial V} (k_r^{i0}\epsilon^r) dS_i. \tag{3.80}$$

These depend on the symmetry parameter functions $\epsilon^r(x)$, and this makes it hard to interpret them, generically. Off course, if the ϵ^r are constant parameter, we recover the previous charges as $C_\epsilon = \epsilon^r C_r$. Otherwise, without appropriate boundary conditions (both on the fields and on the ϵ), the C_ϵ are not well-defined objects. On the other hand, the ϵ are not completely arbitrary since they need to obey asymptotic conditions by their provenance from the variation of an action. The field equations are only valid, if the fields are fixed on the boundary: $\delta Q^\alpha = 0_{|\partial V}$. This must also be the case for the symmetry transformations (3.68) and restricts the possible $\epsilon^r(x)$ that are allowed in the infinite set of charges (3.80). A deeper analysis is only possible on a case-to-case basis. The very fact that locally-symmetric theories allow for infinitely many charges was first realized in general relativity in the early fifties. This became accepted only slowly, since it leads to the perplexing observation that there exists a plethora of energy-momentum tensors for the gravitational field (more about this in Sect. 7.5.4).

Here we observe different qualities of global and of local symmetries: Whereas (weakly) conserved Noether currents do arise in both cases, only global symmetries directly lead to interpretable conserved charges. Instead the strong suit of local symmetries is to restrict the form of the Lagrangian. This will be seen at an example below, and I will demonstrate this anew in deriving the Yang-Mills strength (see Sect. 5.3.4) and in showing the equivalence of Belinfante and Hilbert energy-momentum tensors in gravitational theories (see Sect. 7.5.4).

The relation (3.71), together with (3.77) points to an important peculiarity of theories with a local symmetry: In choosing $\mu = 0 = \nu$ and observing that for the pseudo-momenta (3.23) the $\Pi^0_\alpha = P_\alpha$ are the canonically-conjugate momenta to the fields Q^α, we obtain R relations

$$P_\alpha \mathcal{B}_r^{\alpha 0} - \rho_r^{00} = 0, \tag{3.81}$$

which show that the momenta are not independent. The relations (3.81) correspond to "primary" constraints as they were dubbed by J.L. Anderson and P.G. Bergmann [9]. In the case of higher derivative terms in (3.68), it would be the highest non-vanishing coefficients which lead to primary constraints ([449]). The existence of constraints among the canonically-conjugate fields is due to the non-invertibility of the Legendre transformation, as it is typically the case for singular systems. A separate appendix (Appendix C) is devoted to the Hamiltonian formulation of singular systems, among them theories with local symmetries such as Yang-Mills theories and general relativity.

Example

In order to demonstrate the rich information content of the Klein-Noether identities, let us assume that we are dealing with a theory in terms of a vector field A_μ and a complex scalar field ϕ which is invariant under the transformations

$$A_\mu \mapsto A_\mu + \partial_\mu \epsilon(x) \tag{3.82a}$$

$$\phi \mapsto \phi + iq\epsilon(x)\phi \qquad \phi^* \mapsto \phi^* - iq\epsilon(x)\phi^*. \tag{3.82b}$$

The coefficients corresponding to (3.68) are $\mathcal{D} = 0$ and

$$\begin{array}{lll} \mathcal{A}_\mu = 0 & \mathcal{A}^\phi = iq\phi & \mathcal{A}^{\phi^*} = -iq\phi^* \\ \mathcal{B}^\nu_\mu = \delta^\nu_\mu & \mathcal{B}_\phi = 0 & \mathcal{B}_{\phi^*} = 0. \end{array}$$

The current constituents in $J^\mu_\epsilon = j^\mu \epsilon + k^{\mu\nu}\epsilon_{,\nu}$ are according to (3.78)

$$j^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} iq\phi + c.c. \qquad k^{\mu\nu} = \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)}.$$

Since $k^{\mu\nu}$ must be antisymmetric due to (3.72a), we immediately learn that derivatives of the vector fields in the Lagrangian $\mathcal{L}(A, \partial A, \phi, \partial\phi)$ can appear only in the combination $F_{\mu\nu} = (\partial_\mu A_\nu - \partial_\nu A_\mu)$. The Klein-Noether identities (3.72b) become

$$[\mathcal{L}]^\mu + \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} iq\phi + c.c.\right) + \partial_\nu \frac{\partial \mathcal{L}}{\partial(\partial_\nu A_\mu)} = \frac{\partial \mathcal{L}}{\partial A_\mu} + \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} iq\phi + c.c.\right) = 0$$

from which, assuming that there is no explicit dependence of the Lagrangian on the vector field, these necessarily must occur in combination with the derivatives of the scalar fields as $D_\mu = \partial_\mu + iqA_\mu$. Finally, the relation (3.72c) is

$$[\mathcal{L}]^\phi iq\phi + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} iq\phi\right) + h.c. = \frac{\partial \mathcal{L}}{\partial \phi} iq\phi + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} iq(\partial_\mu \phi) + h.c. = 0.$$

If we exclude a pathological situation in which the derivatives of a Lagrangian with respect to the fields ϕ are related to the Lagrangian derivatives with respect to $\partial_\mu \phi$, the latter identity splits:

$$\frac{\partial \mathcal{L}}{\partial \phi}\phi - \frac{\partial \mathcal{L}}{\partial \phi^*}\phi^* = 0 \qquad \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}(\partial_\mu \phi) - \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^*)}(\partial_\mu \phi^*) = 0.$$

This reveals that (1) the dependence of the Lagrangian on the scalar fields can only be in terms of $(\phi\phi^*)$, (2) the dependence of the Lagrangian on the derivatives of the scalar field is given in terms of $(\partial_\mu \phi \partial^\mu \phi^*)$. Therefore, the Klein-Noether identities lead to a Lagrangian $\mathcal{L}(F_{\mu\nu}, D_\mu \phi D^\mu \phi^*, \phi\phi^*)$, which then can further be restricted by requiring that the Lagrangian be real and Lorentz invariant. This enforces the fact that the field strengths can appear only in the form $(F_{\mu\nu}F^{\mu\nu})$. In any case, this set of Lagrangians contains (3.38), that is the "usual" one describing the interaction of a charged boson with the electromagnetic field.

What are the Noether charges in this example? Since $\Pi^{\mu\nu} = -F^{\mu\nu}$ and since there are no surface terms, we find with $\mathcal{B}^{\nu}_{\mu} = \delta^{\nu}_{\mu}$, $\mathcal{B}_{\phi} = 0$ that $k^{\mu\nu} = F^{\mu\nu}$, Therefore, according to (3.79), we recover the charge as given by Gauß' law; see (3.67).

3.3.4 *Further Topics Relating to Variational Symmetries*

More General Symmetry Transformations

The previous cases for the Noether theorems can be generalized in various ways (and indeed were treated quite generally by E. Noether herself).

Transformations with Higher Derivatives in the Infinitesimal Parameters

The identity (3.69) immediately allows us to deduce the consequences of more general symmetry transformations of the form (3.68), for instance transformations with second derivatives in the parameters ϵ^r:

$$\delta Q^{\alpha} = \mathcal{A}^{\alpha}_{r}(Q)\cdot\epsilon^{r}(x) + \mathcal{B}^{\alpha\mu}_{r}(Q)\cdot\epsilon^{r}_{\ ,\mu}(x) + \mathcal{C}^{\alpha\mu\nu}_{r}(Q)\cdot\epsilon^{r}_{\ ,\mu\nu}(x). \tag{3.83}$$

As before in (3.70), the current J^{μ}_{S} is split into two parts,

$$J^{\mu}_{S} = \tilde{j}^{\mu}_{r}\epsilon^{r} + \partial_{\nu}(k^{\mu\nu}_{r}\epsilon^{r}) \qquad\qquad \tilde{j}^{\mu}_{r} = j^{\mu}_{r} - \partial_{\nu}k^{\mu\nu}_{r}.$$

Inserting these expressions into the invariance condition $[\mathcal{L}]_{\alpha}\delta_{S}Q^{\alpha} + \partial_{\mu}J^{\mu}_{S} \equiv 0$ we obtain a cascade of Klein-Noether identities directly comparable to (3.77)

$$\begin{aligned}
[\mathcal{L}]_{\alpha}\mathcal{A}^{\alpha}_{r} + \partial_{\mu}\tilde{j}^{\mu}_{r} + \partial_{\mu}\partial_{\nu}k^{\mu\nu}_{r} &\equiv 0\\
[\mathcal{L}]_{\alpha}\mathcal{B}^{\alpha\mu}_{r} + \tilde{j}^{\mu}_{r} + 2\partial_{\nu}k^{\mu\nu}_{r} &\equiv 0\\
[\mathcal{L}]_{\alpha}\mathcal{C}^{\alpha[\mu\nu]}_{r} + k^{[\mu\nu]}_{r} &\equiv 0.
\end{aligned}$$

These in turn imply the Noether identities

$$[\mathcal{L}]_{\alpha}\mathcal{A}^{\alpha}_{r} - \partial_{\mu}\big([\mathcal{L}]_{\alpha}\mathcal{B}^{\alpha\mu}_{r}\big) + \partial_{\mu}\partial_{\nu}\big([\mathcal{L}]_{\alpha}\mathcal{C}^{\alpha\mu\nu}_{r}\big) \equiv 0 \tag{3.84}$$

as a generalization of (3.75). This exhibits the generic relation between the variations of the fields and the ensuing Noether identities, which here for simplicity is stated for a finite-dimensional system with a a one-parameter group of symmetries:

$$\delta q^{i} = \sum_{s=0}^{N}\Phi^{i}_{s}\frac{d^{s}\epsilon}{dt^{s}} \;\longleftrightarrow\; \sum_{s=0}^{N}(-1)^{s}\frac{d^{s}}{dt^{s}}\big(\Phi^{i}_{s}[L]_{i}\big) = 0 \;\longleftrightarrow\; \delta L = \frac{d}{dt}F. \tag{3.85}$$

Higher-Derivative Theories

In the previous context, it was assumed that the Lagrangians depend on the fields and at most on the first derivatives of the fields. This is true for Yang-Mills field theories and for certain formulations of GR or modified gravitational theories. It is not true for the original Hilbert-Einstein Lagrangian which, as we will see in Chap. 7, depends on the metric field and its first and second derivatives. A straightforward calculation reveals that (3.43) is left unchanged if the Noether current (3.44) is augmented by a part containing the second derivatives of the fields:

$$J_S^\mu = \Pi^{\alpha\mu}\bar{\delta}_S Q^\alpha + \mathcal{L}\delta_S x^\mu - \Sigma_S^\mu + \frac{\partial\mathcal{L}}{\partial Q^\alpha_{,\mu\nu}}\bar{\delta}_S Q^\alpha_{,\nu} - \partial_\nu \frac{\partial\mathcal{L}}{\partial Q^\alpha_{,\mu\nu}}\bar{\delta}_S Q^\alpha. \tag{3.86}$$

Gauge Conditions: Breaking the Symmetry

A local symmetry has as a consequence that the fields Q^α and $Q'^\alpha = Q^\alpha + \delta_S Q^\alpha$ describe the same physical state. In other words, the field equations are underdetermined and one is dealing with redundant degrees of freedom. This necessitates the choice of *gauge conditions*. As is well known from electrodynamics, the potentials A_μ are not observable, but only the electric and magnetic field strengths[17]. The invariance $A_\mu \mapsto A_\mu + \partial_\mu\epsilon$ can be broken by imposing conditions such as

$$\begin{aligned}
&\partial^\mu A_\mu = 0 \quad && \text{Lorenz gauge} \\
&\partial_i A_i = 0 \quad && \text{Coulomb gauge} \\
&n^\mu A_\mu = 0 \quad && \text{with } n^\mu \neq 0 \text{ for at least one index } \mu.
\end{aligned}$$

The latter comprises the temporal ($A_0 = 0$), the axial ($A_3 = 0$) and the light cone gauge ($A_0 - A_3 = 0$). A gauge condition must fulfill at least two criteria: (i) it should be attainable, i.e. it must be possible to arrive from arbitrary fields to fields that obey the gauge condition (ii) it should not overspecify the fields. One may further require that a gauge condition completely break the symmetry. The Lorenz gauge does not accomplish complete symmetry breaking: It leaves open symmetry transformations for which the ϵ are solutions of $\partial_\mu\partial^\mu\epsilon = 0$. For the Coulomb gauge the condition $0 = \partial_i A'_i = \partial_i A_i + \partial_i\partial_i\epsilon$ leads to a Poisson equation for ϵ with the solution

$$\epsilon(\vec{x}, t) = -\int d^3y \, \frac{\partial_i A_i(\vec{y}, t)}{4\pi|\vec{x} - \vec{y}|}.$$

[17] This is not the full truth: The Aharanov-Bohm effect reveals traces of the gauge potentials in the form of holonomies.

Furthermore, in this gauge, the field equations assume the form

$$\partial_t \partial_t A_\nu - c^2 \partial_k \partial_k A_\nu - c \partial_t \partial_\nu A_0 = c^2 j_\nu$$

which for A_0 has the solution

$$A_0 = - \int d^3 y \frac{j_0(\vec{y}, t)}{4\pi |\vec{x} - \vec{y}|}.$$

The fact that the scalar potential is just the instantaneous Coulomb potential is the origin of the name Coulomb gauge. Other, less common and partly incomplete gauges are the multipolar (or Poincaré) gauge $x^i A_i = 0$ and the Fock-Schwinger gauge $x^\mu A_\mu = 0$.

Substituting Fields Within the Action

What happens to the original symmetry and the dynamical equations if some fields in terms of other fields are substituted into of a given action? This question has been answered by J. Pons in [419], and I follow his notation: Assume a Lagrangian $\mathcal{L}(\phi, \psi)$ depending on fields ϕ, ψ and their derivatives. Assume that the fields ψ can be expressed as $\psi = F(\phi, \partial_\mu \phi, \ldots)$, and denote $\mathcal{L}_r(\phi) = \mathcal{L}(\phi, \psi \to F)$. Is it true that the original field equations restricted to $\psi \to F$ are equivalent to the field equations following from the reduced Lagrangian $\mathcal{L}_r$? The answer is

$$([\mathcal{L}]_\phi)_{|F} = 0 \quad \text{and} \quad ([\mathcal{L}]\ )_{|\mathrm{F}} = 0 \quad \Longrightarrow \quad [\mathcal{L}_\mathrm{r}]_{\text{Œ}} = 0,$$

but the inverse is not true in general–the deviation being quantified in [419]. A further result is as follows: If one defines an extended Lagrangian $\mathcal{L}_E = \mathcal{L} + \lambda(\psi - F)$ which contains the restriction $\psi \to F$ with a Lagrange multiplier λ, it turns out that the dynamical equations derived from $\mathcal{L}_E$ are equivalent to those derived from $\mathcal{L}_r$:

$$[\mathcal{L}_E]_{(\phi,\psi,\lambda)} = 0 \quad \Longleftrightarrow \quad [\mathcal{L}_r]_\phi = 0, \quad \psi - F = 0, \quad \lambda + [\mathcal{L}]_\psi = 0.$$

There is a very special situation where $([\mathcal{L}]_\phi)_{|F} = 0$ has the same dynamical content as $[\mathcal{L}_r]_\phi = 0$ and this is the case for so-called *auxiliary fields* ψ_a. For these $([\mathcal{L}]_{\psi_a}) \equiv 0$ - or expressed in another way - an auxiliary field can be isolated by its own equation of motion: $[\mathcal{L}]_{\psi_a} \leftrightarrow \psi_a = F_a$. Auxiliary fields are quite common in supersymmetric theories.

These results hold for generic variations. In the case of infinitesimal symmetry transformations $\delta_S(\phi, \psi)$ which, as indicated, generically depend on the fields, one finds (details again in [419]) that the answer to the preservation of symmetries depends on the "tangency condition"

$$\big(\delta_S(\psi - F)\big)_{|F} = 0$$

with the following results: (i) If $\delta_S\phi$ and $\delta_S\psi$ is a Noether symmetry for $\mathcal{L}(\phi, \psi)$ and if the tangency condition holds, then $(\delta_S\phi)_{|F}$ is a Noether symmetry for $\mathcal{L}_r$. (ii) The tangency condition is always fulfilled for auxiliary fields. (iii) For auxiliary fields both Noether and Lie symmetries are preserved.

Substituting Gauge Conditions Within the Action

What happens to the dynamical equations if gauge conditions are substituted into a given action? (We are of course not asking, "What happens to the symmetry", since the motivation for introducing gauge conditions is to break the symmetry.) For holonomic constraints (those which do not depend on the derivatives of the fields) this was investigated in [417] on the level of the Hamiltonian. A complete gauge fixing gives rise to a theory which has at most second-class constraints (for the terminology see Appendix C). It is equivalent to the original theory only if the first-class constraints are added in by hand.

In recent times the R_ξ gauge has become prominent, essentially because of its more or less natural appearance in functional integrals and its power to investigate questions of unitarity and renormalizability in quantum field theories. It amounts to adding to the Lagrangian of electrodynamics some further terms with a field B and a constant ξ:

$$\mathcal{L}_\xi = -\frac{1}{2}\partial^\mu A^\nu(\partial_\mu A_\nu - \partial_\nu A_\mu) + B(\partial^\mu A_\mu) + \frac{\xi}{2}B^2 + A_\mu j^\mu. \tag{3.87}$$

The Euler derivatives are

$$\begin{aligned}
[\mathcal{L}_\xi]^\nu &= \frac{\partial\mathcal{L}}{\partial A_\nu} - \partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu A_\nu)} = j^\nu + \partial_\mu(\partial^\mu A^\nu - \partial^\nu A\mu) - \partial^\nu B \\
[\mathcal{L}_\xi]_B &= \frac{\partial\mathcal{L}}{\partial B} = \partial^\mu A_\mu + \xi B.
\end{aligned}$$

The Euler-Lagrange equation $[\mathcal{L}_\xi]_B = 0$ may be solved for the auxiliary field B, so that

$$\mathcal{L}_\xi \to -\frac{1}{2}\partial^\mu A^\nu(\partial_\mu A_\nu - \partial_\nu A_\mu) + \frac{1}{2\xi}(\partial^\mu A_\mu)^2 + A_\mu j^\mu.$$

For $\xi \to 0$ one obtains the Landau gauge (formally being the same as the Lorenz gauge); for $\xi = 1$ the name "Feynman-'t Hooft" gauge has become established.

Noether Symmetries and Lie Symmetries

The relation between Lie and Noether symmetries becomes apparent by using the technique with symmetry generators and their prolongations as described for finite-dimensional systems in Sect. 2.2. Infinitesimal point transformations near the identity are generically

$$\delta x^{\mu} = \epsilon\, \xi^{\mu}(x, Q)) \quad \delta Q^{\alpha} = \epsilon\, \eta^{\alpha}(x, Q) \quad \bar{\delta} Q^{\alpha} = \epsilon\, \chi^{\alpha} \quad \text{with} \quad \text{Ø} = -\,\text{Q}_{,\text{-}}\text{¸}^{\text{-}}$$

and they are encoded in a generator X and its prolongation (by some ([285]) called the Lie-Bäcklund operator)

$$\overline{X} = \xi^{\mu} d_{\mu} + \chi^{\alpha}\frac{\partial}{\partial Q^{\alpha}} + (d_{\mu}\chi^{\alpha})\frac{\partial}{\partial Q^{\alpha}_{,\mu}} + (d_{\mu}d_{\nu}\chi^{\alpha})\frac{\partial}{\partial Q^{\alpha}_{,\mu\nu}} + \dots$$

Here $d_{\mu} = d/dx^{\mu}$ acts as the total derivative with respect to the coordinates x^{μ}. Following [285], we consider the Euler operators E_{α} and the Noether operators N^{μ}:

$$E_{\alpha} = \frac{\delta}{\delta Q^{\alpha}} = \frac{\partial}{\partial Q^{\alpha}} + \sum_{k\geqslant 1}(-1)^{k}\, d_{\mu_1}\dots d_{\mu_k}\frac{\partial}{\partial Q^{\alpha}_{\mu_1\dots\mu_k}}$$

$$N^{\mu} = \xi^{\mu} + \chi^{\alpha}\frac{\delta}{\delta Q^{\alpha}_{\mu}} + \sum_{k=1}(d_{\mu_1}\dots d_{\mu_k}\chi^{\alpha})\frac{\partial}{\partial Q^{\alpha}_{\mu_1\dots\mu_k}}.$$

It can be shown that there is the operator identity

$$\overline{X} + d_{\mu}\xi^{\mu} = \chi^{\alpha}E_{\alpha} + d_{\mu}N^{\mu}$$

which can be used for proving some of the subsequent statements. Applying the Euler operator to a Lagrangian leads–no surprise–to the Euler-Lagrange expression: $E_{\alpha}\mathcal{L} = [\mathcal{L}]_{\alpha}$.

A Lie symmetry is at hand if

$$\overline{X}[\mathcal{L}]^{\alpha}\Big|_{[\mathcal{L}]^{\beta}} = 0,$$

and this is a Noether symmetry if a Σ^{μ} exists such that

$$\overline{X}\mathcal{L} + \mathcal{L}d_{\mu}\xi^{\mu} = d_{\mu}\Sigma^{\mu}.$$

For every Noether symmetry there is a current

$$J^{\mu} = N^{\mu}\mathcal{L} - \Sigma^{\mu}$$

which is conserved on-shell:

$$d_\mu J^\mu = 0 \qquad \text{for} \qquad [\mathcal{L}] = 0.$$

Variational Transformations in General

In the previous exposition of symmetries, it was (following Noether) more or less tacitly assumed that the symmetry transformations constitute a Lie group. This is definitely true for the gauge theories underlying the standard model of particle physics, albeit not for gravitational theories and supersymmetric theories, where the "structure constants" are no longer constants and/or the algebra of infinitesimal symmetry transformations closes only on-shell.

In order to prepare the consequences for these more general symmetry transformations, let me reflect some basic findings in the notation of B. DeWitt from his famous *Les Houches* lectures 1963 [120]. Assume that the classical action has a stationary point:

$$\frac{\delta S}{\delta Q^\alpha}\bigg|_{\hat{Q}} =: S_{,\alpha} = 0.$$

at $Q = \hat{Q}$. (This is of course equivalent to the vanishing Euler derivatives of the Lagrangian: $[\mathcal{L}]_\alpha = 0$ in the case that no surface terms arise.) If both $\hat{Q}$ and $\hat{Q}+\delta Q$ are solutions of the equations of motion, differing only infinitesimally, we can write

$$\begin{aligned} 0 &= S_{,\alpha}[\hat{Q}] \\ 0 &= S_{,\alpha}[\hat{Q} + \delta Q] = S_{,\alpha}[\hat{Q}] + S_{,\alpha\beta}[\hat{Q}]\,\delta Q^\beta \end{aligned}$$

leading to the "equation of small disturbances" $S_{,\alpha\beta}[\hat{Q}]\,\delta Q^\beta = 0$.

Assume that the theory is invariant under symmetry transformations of the fields, infinitesimally

$$\delta Q^\alpha = \mathcal{R}^\alpha_r(Q) \cdot \epsilon^r(x) = \int dy\, \mathcal{R}^\alpha_r(x, y)\epsilon^r(y). \tag{3.88}$$

Here $\mathcal{R}$ is understood to be a differential operator. The previous cases are covered by

$$\mathcal{R}^\alpha_r(x, y) = \mathcal{A}^\alpha_r \delta(x - y) + \mathcal{B}^{\alpha\mu}_r \partial_\mu \delta(x - y) + \dots$$

Invariance means that $\delta_\epsilon S = S_{,\alpha}\, \delta_\epsilon Q^\beta = 0$, or with (3.88) that

$$S_{,\alpha}\, \mathcal{R}^\alpha_r = \int dx\, S_{,\alpha}(x)\, \mathcal{R}^\alpha_r(x, y) \equiv 0. \tag{3.89}$$

These are–in a condensed notation–the Noether identities with respect to the symmetry transformations (3.88). They imply

$$\frac{\delta}{\delta Q^\beta}(S_{,\alpha}\mathcal{R}^\alpha_r) = S_{,\alpha\beta}\,\mathcal{R}^\alpha_r = 0$$

(on the stationary points of S) which shows that $S_{,\alpha\beta}$ is not invertible. Consider a variation of the field equations:

$$\delta(S_{,\alpha}) = S_{,\alpha\beta}\,\delta Q^\beta = S_{,\alpha\beta}R^\beta_r\cdot\epsilon^r = \partial_\alpha(S_{,\beta}\mathcal{R}^\beta_r)\cdot\epsilon^r - S_{,\beta}\mathcal{R}^\beta_{r,\alpha}\cdot\epsilon^r .$$

The first term vanishes identically because of (3.89), and thus we see again that a variational symmetry maps solutions to solutions. These results are independent of whether the symmetry transformations constitute a Lie group or not.

With regard to applications in supersymmetric theories, one must allow for bosonic and fermionic transformations characterized by Grassmann even and odd infinitesimal parameters ϵ (see Appendix B.2). Consider the commutator of two transformations. A direct computation leads to

$$[\delta_1,\delta_2]Q^\alpha = \left(\mathcal{R}^\alpha_{r,\beta}\mathcal{R}^\beta_s - (-1)^{|\epsilon^r||\epsilon^s|}\mathcal{R}^\alpha_{s,\beta}\mathcal{R}^\beta_r\right)\epsilon^r_1\epsilon^s_2.$$

Here, again in the DeWitt notation,

$$\mathcal{R}^\alpha_{r,\beta}\mathcal{R}^\beta_s = \int dy\,\frac{\delta\mathcal{R}^\alpha_r(x,y)}{\delta Q^\beta(y)}\mathcal{R}^\beta_s(y,z).$$

On the other hand, the commutator stands for an infinitesimal transformation of the form (3.88) and satisfies the Noether identity:

$$S_{,\alpha}\left(\mathcal{R}^\alpha_{r,\beta}\mathcal{R}^\beta_s - (-1)^{|\epsilon^r||\epsilon^s|}\mathcal{R}^\alpha_{s,\beta}\mathcal{R}^\beta_r\right) = 0.$$

Now it can be shown that this implies a relation

$$\mathcal{R}^\alpha_{r,\beta}\mathcal{R}^\beta_s - (-1)^{|\epsilon^r||\epsilon^s|}\mathcal{R}^\alpha_{s,\beta}\mathcal{R}^\beta_r = -\mathcal{R}^\alpha_t\,f_{trs} - S_{,\beta}\,E^{\beta\alpha}_{rs}. \tag{3.90}$$

The algebra is called "closed" if $E^{\beta\alpha}_{rs} = 0$, and "open" otherwise. If the algebra is closed and the f_{trs} do not depend on the fields, one is dealing with a Lie algebra with structure constants f_{trs}. We will see that general relativity has a closed algebra, but with structure functions f_{trs} depending on the field variables (i.e. the metric fields). We will also see that supersymmetric theories tend to have open algebras. The introduction of auxiliary fields may close the algebra, but in some theories they are not easily found (or not found at all as for most of the $N \geqslant 3$ supersymmetries).

In a further step, one needs to ensure that the Jacobi identities are fulfilled

$$\sum_{\text{cyclic over }1,2,3} [\delta_1, [\delta_2, \delta_3]] Q_\alpha = 0.$$

These are satisfied for a Lie algebra, but otherwise may lead to quite complicated consistency conditions which may necessitate the introduction of further structure tensors aside from f and E; see [232]. Interestingly, there is a dexterous bookkeeping device in terms of quantities related to BRST and anti-BRST symmetries and, more generally, to the fields and anti-fields in the Batalin-Vilkovisky quantization; see [36]. A (little) more about this is given in Appendix D.2.3.

3.4 Poincaré Transformations

3.4.1 Poincaré and Lorentz Groups

Let me first just for completeness of this section repeat some facts about Poincaré transformations as stated in Sect 3.2.2: The elements of the Poincaré group **P** have the generic form $p = (\Lambda, a)$. The Λ describe Lorentz transformations and the a denote spacetime translations. In D dimensions the condition $\Lambda^T \eta \Lambda = \eta$ reduces the number of independent elements in Λ to $\frac{D(D-1)}{2}$. The Poincaré group in D dimensions is therefore a $\frac{D(D+1)}{2}$ -dimensional manifold[18]. For $D = 4$, the Poincaré group is a 10-parameter Lie group.

The group composition is

$$(\Lambda', a') \circ (\Lambda, a) = (\Lambda'\Lambda, \Lambda' a + a'). \tag{3.91}$$

The neutral element obviously is $p_n = (\mathbf{1}, 0)$, from which we immediately derive that the inverse to the element $p = (\Lambda, a)$ is $p^{-1} = \Lambda^{-1}(\mathbf{1}, -a)$. Again I point out that the Poincaré group **P** is much simpler than the Galilei group **Gal**. (Where "simple" is meant colloquially, and not in its strict mathematical sense.) You realize what is meant by comparing with (2.72) and (2.73).

Galilei from Poincaré

It is beyond question that the Lorentz boosts (see e.g. (3.4)) become Galilei transformations if $c \to \infty$ (or the rapidity $\eta \to 1$). Nevertheless, the details of deriving the full Galilei group from the Poincaré group (or the Galilei algebra from the Poincaré

[18] Nearly all expressions in this book can be stated in an arbitrary number of dimensions. This may sound stupendous, since "obviously" our world is 4-dimensional. But, who knows: there are Kaluza-Klein theories, string models, brane worlds, …

algebra) must be treated with greater care. Consider for simplicity this transition in one space and one time dimension. The Poincaré group element, according to

$$g(\eta, a, \tau) = P = \begin{pmatrix} \cosh\eta & \sinh\eta & a \\ \sinh\eta & \cosh\eta & \tau \\ 0 & 0 & 1 \end{pmatrix} \overset{c\to\infty}{\longrightarrow} \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & \tau \\ 0 & 0 & 1 \end{pmatrix},$$

becomes an element in the Galilei group describing only displacements in space and time–boosts are missing. The correct limit can be recovered by first applying a similarity transformation with a c-dependent matrix

$$S = \begin{pmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and by a rescaling $\bar{a} = ca$. Then indeed

$$SPS^{-1} \overset{c\to\infty}{\longrightarrow} \begin{pmatrix} 1 & \mathfrak{v} & \bar{a} \\ 0 & 1 & \tau \\ 0 & 0 & 1 \end{pmatrix}$$

is a generic group element in the inhomogeneous Galilei group. This demonstrates that one needs to make precise sense of the limit operation. I will come back to this a little later.

Lorentz Group and Its Partitions

The Lorentz group **Lor** is a subgroup of the Poincaré group **P** with group elements restricted to $(\Lambda, 0) \in \mathbf{Lor}$. The Lorentz group is isomorphic to $\mathbf{O(3,1)}$. As a topological space, **O(3,1)** splits into four connected components which in the sequel are characterized by the notation $+/-$ (positive/negative determinant) and $\uparrow / \downarrow$ (time orientation preserving/non-preserving). Whereas the first criterium is obvious, the second one needs an explanation: From $\eta_{\mu\nu}\Lambda^{\mu}{}_{\varrho}\Lambda^{\nu}{}_{\sigma} = \eta_{\varrho\sigma}$ we specifically have the restriction $\eta_{\mu\nu}\Lambda^{\mu}{}_{0}\Lambda^{\nu}{}_{0} = \eta_{00}$, or explicitly $\Lambda^{0}{}_{0}\Lambda^{0}{}_{0} - \Lambda^{i}{}_{0}\Lambda^{i}{}_{0} = 1$. This relation suggests that we distinguish two cases, namely $\Lambda^{0}{}_{0} \geqslant 1$ and $\Lambda^{0}{}_{0} \leqslant -1$. The following taxonomy for Lorentz subgroups is introduced:

- orthochronous Lorentz group $\mathbf{Lor}^{\uparrow}$ consisting of transformations that do not change the time-orientation, thus obeying $\Lambda^{0}{}_{0} \geq 1$;
- proper Lorentz group $\mathbf{Lor}_{+}$, characterized by $\det\Lambda = +1$;
- proper orthochronous Lorentz group $\mathbf{Lor}_{+}^{\uparrow}$, which does not contain space inversion and time reversal.

This allows us to write **Lor** as a union of disjunct sets:

$$\mathbf{Lor} = \mathbf{Lor}_+^{\uparrow} \cup \mathbf{Lor}_+^{\downarrow} \cup \mathbf{Lor}_-^{\uparrow} \cup \mathbf{Lor}_-^{\downarrow}. \tag{3.92}$$

Only $\mathbf{Lor}_+^{\uparrow}$ contains the neutral element. $\mathbf{Lor}_+^{\downarrow}$ contains the space inversion P, $\mathbf{Lor}_-^{\uparrow}$ contains the time reversal T, and $\mathbf{Lor}_-^{\downarrow}$ contains PT. The space inversion is the Lorentz transformation with the (non-vanishing) components

$$\Lambda^0{}_0 = 1, \qquad \Lambda^1{}_1 = \Lambda^2{}_2 = \Lambda^3{}_3 = -1,$$

(all other components of Λ vanishing), and the time reversal is explicitly

$$\Lambda^0{}_0 = -1, \qquad \Lambda^1{}_1 = \Lambda^2{}_2 = \Lambda^3{}_3 = 1.$$

In the sense of a coset decomposition one may write

$$\mathbf{Lor}^{\uparrow} = \mathbf{Lor}_+^{\uparrow} \cup P\,\mathbf{Lor}_+^{\uparrow}, \qquad \mathbf{Lor}_+ = \mathbf{Lor}_+^{\uparrow} \cup PT\,\mathbf{Lor}_+^{\uparrow},$$

and thus the full Lorentz group can be subdivided in cosets of of $\mathbf{Lor}_+^{\uparrow}$:

$$\mathbf{Lor} = \mathbf{Lor}_+^{\uparrow} \cup P\,\mathbf{Lor}_+^{\uparrow} \cup T\,\mathbf{Lor}_+^{\uparrow} \cup PT\,\mathbf{Lor}_+^{\uparrow}.$$

The set $\mathbf{Lor}_+ = \mathbf{Lor}_+^{\uparrow} \cup \mathbf{Lor}_+^{\downarrow}$ is as a group isomorphic to $\mathbf{SO(3,1)}$. In the decomposition (3.92) only $\mathbf{Lor}_+^{\uparrow}$ is a subgroup, since it is the only set containing the neutral element. This group is simple, i.e. it does not contain non-trivial cosets. (It is true that the Lorentz boosts are subsets of $\mathbf{Lor}_+^{\uparrow}$, but they do not constitute a subgroup.)

The Poincaré group is the inhomogeneous group **ILor** associated to the Lorentz group.

3.4.2 Poincaré Algebra

The algebra of the Poincaré group **P** in four dimensions is spanned by ten generators: The six $M_{\mu\nu} = -M_{\nu\mu}$ generate the algebra of the Lorentz group, and the four T_μ generate the translations (the inhomogeneous part of the Poincaré group). Let us write the group elements as[19]

$$g(\epsilon, \omega) = \exp i\left(\epsilon^\mu T_\mu - \frac{1}{2}\omega^{\mu\nu} M_{\mu\nu}\right). \tag{3.93}$$

[19] The definition of $M_{\mu\nu}$ in (3.93) is made use of by all those authors who use the metric convention chosen in this book; this definition has as a consequence that the Lorentz generators appear with a minus-sign in the group element.

The Lie algebra $\mathfrak{so}(3,1)$ of the Lorentz group is

$$[M_{\mu\nu}, M_{\varrho\sigma}] = i\left[\eta_{\mu\sigma}M_{\nu\varrho} - \eta_{\nu\sigma}M_{\mu\varrho} + \eta_{\nu\varrho}M_{\mu\sigma} - \eta_{\mu\varrho}M_{\nu\sigma}\right] \tag{3.94}$$

where η is the Minkowsi metric. Appending the translations, the algebra is enlarged to the Poincaré algebra $\mathfrak{iso}(3,1)$ with

$$[M_{\mu\nu}, T_\lambda] = i\left[\eta_{\nu\lambda}T_\mu - \eta_{\mu\lambda}T_\nu\right] \tag{3.95}$$

$$[T_\mu, T_\nu] = 0. \tag{3.96}$$

The algebra can be realized from observing their action on fields Q as follows: Generically $\delta Q = \bar{\delta} Q - (\delta x^\mu)\, \partial_\mu Q$. Now for Poincaré transformations with $\delta x^\mu = \epsilon^\mu + \omega^{\mu\nu} x_\nu$ and δQ given in terms of the spin matrix Σ by (3.56)

$$\begin{aligned}
\bar{\delta}_{(\epsilon\omega)} Q &= -\frac{i}{2}\omega^{\mu\nu}\Sigma_{\mu\nu} Q - (\epsilon^\mu + \omega^{\mu\nu} x_\nu)\partial_\mu Q \\
&= -\epsilon^\mu \partial_\mu Q - \frac{1}{2}\omega^{\mu\nu}(x_\nu\partial_\mu - x_\mu\partial_\nu) Q - \frac{i}{2}\omega^{\mu\nu}\Sigma_{\mu\nu} Q \\
&= i\epsilon^\mu (i\partial_\mu) Q - \frac{i}{2}\omega^{\mu\nu}\left(i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu}\right) Q \\
&= i\left(\epsilon^\mu T_\mu - \frac{1}{2}\omega^{\mu\nu} M_{\mu\nu}\right) Q
\end{aligned} \tag{3.97}$$

from which we identify

$$T_\mu \overset{\circ}{=} i\partial_\mu \tag{3.98a}$$

$$M_{\mu\nu} \overset{\circ}{=} L_{\mu\nu} + \Sigma_{\mu\nu} = i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu}. \tag{3.98b}$$

In order to recover the (3+1) version of the Poincaré algebra, the following linear combinations of the generators (T, L) are introduced

$$\begin{aligned}
J^i &:= \frac{1}{2}\epsilon^{ijk} L^{jk} = i\epsilon^{ijk} x^j \partial^k \qquad & K^j &:= \frac{1}{c} L^{0j} = \frac{i}{c}(x^0\partial^j - x^j\partial^0) = G^j - \frac{1}{c^2} x^j \partial_t \\
T^i &:= i\partial^i & H &:= cT^0 = ic\partial_t.
\end{aligned}$$

The choice of factors c in the definitions von K^j and H is motivated by the considerations that K^j should be the generator of boosts with v^j (and not with v^j/c), and that H should generate time translations with t (and not with $x^0 = ct$)[20].

[20] You certainly know that the unit of length is no longer derived from the circumference of the earth's equator and represented in the physical meter prototype preserved in Paris (as I learned in school), nor is it any longer defined in terms of the wavelength of light emitted by a krypton isotope (as I learned at the university), but–since 1983–in terms of the velocity of light. The International Bureau of Weights and Measures (BIPM) states: "The meter is the length of the path travelled by

(Be aware that here we made a specific choice of “time”, namely coordinate time $t = x^0$.) These generators obey the algebra

$$[J^i, J^j] = i\epsilon^{ijk} J^k \; [J^i, T^j] = i\epsilon^{ijk} T^k \quad [J^i, K^j] = \epsilon^{ijk} K^k \;\; [J^i, H] = 0 \tag{3.99a}$$

$$[H, T^i] = 0 \qquad [H, K^j] = iT^j \tag{3.99b}$$

$$[T^i, T^j] = 0 \qquad [T^i, K^j] = \frac{i}{c^2}\,\delta_{ij} H \;\; [K^i, K^j] = -\frac{i}{c^2}\,\epsilon_{ijk} J_k. \tag{3.99c}$$

Subalgebras of the Poincaré algebra are the sets $\{H, T\}, \{H, J\}, \{H, T, J\}, \{T\}, \{T, J\}, \{J\}, \{J, K\}$. The K_i alone do not constitute an algebra. This is compatible with the fact that the Lorentz boosts do not constitute a subgroup.

3.4.3 Galilei and Bargmann Algebra

The Lie algebra of the Galilei group (2.75) can formally be derived from the Lie algebra (3.99) of the Poincaré group by taking the limit $1/c^2 \to 0$. This leads immediately to structural changes in the group algebra. We saw already that one needs to give a precise sense to the limit procedure. This was investigated and E. Wigner [289] in more generality. The procedure is an example of what is called a contraction, see Appendix A.2.4.

One can even argue that the limit $H/c^2 \to 0$ for $c \to \infty$, although mathematically allowed, is physically disputable. Since the numerical value of H is the energy, and since according to (3.20) the energy contains a term mc^2 we would rather expect $H \to \infty$ for $c \to \infty$. Instead with H we should work with $\tilde{H}$ defined by $H = mc^2 + \tilde{H}$, and $H/c^2 \to m$ for $c \to \infty$. In this limit the Poincaré algebra (3.99) passes over to

$$[J^i, J^j] = i\epsilon^{ijk} J^k \quad [J^i, T^j] = i\epsilon^{ijk} T^k \quad [J^i, K^j] = i\epsilon^{ijk} K^k \quad [J^i, \tilde{H}] = 0 \tag{3.100a}$$

$$[\tilde{H}, T^i] = 0 \qquad [\tilde{H}, K^j] = iT^j \tag{3.100b}$$

$$[T^i, T^j] = 0 \qquad [T^i, K^j] = im\delta^{ij} \qquad [K^i, K^j] = 0. \tag{3.100c}$$

This can be considered an algebra if one is willing to interpret the mass m as an additional generator m which commutes with all other generators. Mathematically this 11-dimensional algebra is the central extension of the inhomogeneous Galilei algebra, and m is a central charge; see Appendix A.3.5.

As elaborated in the next chapter, the group related to this algebra plays a distinctive role in non-relativistic quantum physics, where it is called by some the

(Footnote 20 continued)
light in vacuum during a time interval of 1/299792458 of a second.” This, by the way, answers the question about how precisely c is measured.

Bargmann group. It is a subgroup of the *Schrödinger group*, which is the symmetry group of the Schrödinger equation–the latter can even be derived from it. This is due to the fact that the Schrödinger group allows for proper representations in contrast to ray representations of the Galilei group. Furthermore, the existence of the central charge gives rise to a mass superselection rule.

Before dealing with the Poincaré and the Galilei algebra in the broader context of "kinematical" algebras in the section after the next, I want to return to the question of the choice of "time" in a relativistic theory.

3.4.4 Forms of Relativistic Dynamics

The title of this subsection is the same as that of an article by P.A.M. Dirac from 1949 [126] and a lecture given by B. Bakker [22]. Choosing a time variable τ amounts to foliating spacetime into space-like hypersurfaces Σ. Without any further restrictions, there are infinitely many choices. But one may ask the question as to which of the Poincaré generators leave the foliation invariant. The set of all such generators defines a subgroup $\mathbf{G}_\Sigma$ of the Poincaré group, the *stabilizer* of Σ. The generators of this subgroup are called kinematical. The rest, the dynamical operators map Σ onto another hypersurface Σ' and therefore describe a development in τ. They thus make up for the "Hamiltonians", as Dirac called them. Still, in the case that there are no further restrictions to the regularity of the hypersurfaces, a classification of possible choices of the dynamical variables seems hopeless. This changes, however, if one demands that any two points on the hypersurface can be connected by a transformation from $\mathbf{G}_\Sigma$. In group-theoretical terms (see Appendix E.4.4), this means that the stabilizer subgroup acts transitively on Σ, that is $\forall x, y \in \Sigma : \exists g \in \mathbf{G}_\Sigma \rightarrow x = gy$. Under these circumstances there are only five different and inequivalent hypersurfaces, as shown in [339]:

$$
\begin{aligned}
\mathrm{IF}: &\quad x^0 = 0 \\
\mathrm{FF}: &\quad x^0 + x^3 = 0 \\
\mathrm{PF}: &\quad x^2 = a^2 > 0, x^0 > 0 \\
\mathrm{H2}: &\quad (x^0)^2 - (x^1)^2 - (x^2)^2 = a^2 > 0, x^0 > 0 \\
\mathrm{H3}: &\quad (x^0)^2 - (x^3)^2 = a^2 > 0, x^0 > 0.
\end{aligned}
$$

The first three hypersurfaces were already investigated by Dirac, and he called them "instant form", "front form", and "point form". He emphasized that the structure of a relativistically invariant Hamiltonian quantum theory is considerably different for these different forms. Although not privileged, the IF is usually adopted, but also FF has its merits. The hypersurfaces H2 and H3 have not been used in practice–at least as far as I know. Let us investigate the kinematical and dynamical operators.

Table 3.1 Three different forms of dynamics

form	τ	dynamical generators	kinematical generators	N
IF	t	$H, \vec{K}$	$\vec{T}, \vec{J}$	6
FF	$t + x^3/c$	$T^0 - T^3$,	$T^0 + T^3, T^i, J^3$	
		$K^1 - J^2, K^2 + J^1$	$K^1 + J^2, K^2 - J^1, K^3$	7
PF	$(t^2 - \vec{x}^2/c^2 - a^2/c^2)^{1/2}$	T^μ	$M^{\mu\nu}$	6

In the realization (3.98), the action of a Poincaré transformation on a scalar function $\varphi(x)$ is

$$\overline{\delta}\varphi = i\epsilon_\mu \partial^\mu \varphi - \frac{i}{2}\omega_{\mu\nu}(x^\mu \partial^\nu - x^\nu \partial^\mu)\varphi.$$

To each of the previously distinguished forms/hypersurfaces, a time variable $\tau(x)$ is associated. Now, by definition, those generators $T^{\hat{\mu}}$ and $M^{\hat{\mu}\hat{\nu}}$ are "kinematical" for which supercript-wise

$$\partial^{\hat{\mu}}\tau(x) \equiv 0 \qquad (x^{\hat{\mu}}\partial^{\hat{\nu}} - x^{\hat{\nu}}\partial^{\hat{\mu}})\tau(x) \equiv 0. \tag{3.101}$$

For example, for the instant form $\tau = x^0$:

$$\partial^i \tau \equiv 0 \qquad (x^i \partial^j - x^j \partial^i)\tau(x) \equiv 0.$$

Therefore T^i and L^{ij} (that is $\vec{T}$ and $\vec{J}$) are kinematical generators. On the other hand T^0 and L^{0j} (that is H an $\vec{K}$) are dynamical operators. This is of course no surprise but shows that the very idea of defining the kinematical and dynamical operators by (3.101) makes sense. Applying this to the front form and the point form, you may establish Table 3.1 to find the respective dynamical and kinematical generators. The last entry in the table denotes the number of independent kinematical generators. For the two other forms H2 and H3 you would find that this number is $N = 4$. Thus among the five distinguished hypersurfaces the front form has the maximum number of kinematical invariants, and this has certain advantages for formulating relativistic dynamics, especially to describe bound state problems, as exploited in [339].

Observe that for $c \to \infty$ in all cases $\tau \to t$. As a consequence, the instant form is the only appropriate initial surface for Galilei-invariant systems.

3.4.5 Kinematical Groups and Their Mutual Contractions

Kinematical Algebras

In their analysis [18] H. Bacry and J.-M. Lévy-Leblond identified all possible kinematical groups, that is the invariance groups of spacetimes that possess the properties of homogeneity of space and time, the isotropy of space, and of invariance under boosts. These properties require that the kinematical group must be a ten-dimensional Lie group. Let the associated generators be J_i, P_i, H, K_i for space rotations, space translations, time translations, and boosts, respectively[21]. The requirement of isotropy of space enforces that H has to be regarded as a scalar, and P_i, J_i, K_i must be vectors. This in turn yields for the commutation relations involving the J_i:

$$[J_i, H] = 0 \quad [J_i, J_j] = \epsilon_{ijk} J_k \quad [J_i, P_j] = \epsilon_{ijk} \quad P_k [J_i, K_j] = \epsilon_{ijk} K_k. \tag{3.102}$$

However, the other commutation relations are *a priori* not restricted. Now Bacry and Lévy-Leblond made the further assumption that space inversion P and time reversal T are automorphisms of the kinematical group[22]. This means that the discrete maps

$$\begin{aligned} \mathrm{P}&: H \mapsto H,\ P \mapsto -P,\ K \mapsto -K,\ J \mapsto J \\ \mathrm{T}&: H \mapsto -H,\ P \mapsto P,\ K \mapsto -K,\ J \mapsto J \end{aligned}$$

are to be compatible with the algebra. Given these assumptions, Bacry and Lévy-Leblond derive that the remaining commutation relations are of the form

$$\begin{aligned} &[H, P_i] = \alpha K_i && [H, K_i] = \lambda P_i && \\ &[P_i, P_j] = \beta \epsilon_{ijk} J_k && [K_i, K_j] = \mu \epsilon_{ijk} J_k && [P_i, K_j] = \rho \eta_{ij} H \end{aligned}$$

with real constants $\alpha, \beta, \lambda, \mu, \rho$. These constants are not arbitrary because one needs to ensure that the Jacobi identities are obeyed. A tedious but nevertheless straightforward calculation reveals that these are satisfied only if

$$\beta - \alpha\rho = 0 \qquad \mu + \lambda\rho = 0,$$

and thus the possible kinematical groups are completely described by three real parameters α, λ, ρ. The overall sign of these parameters is irrelevant because it can be absorbed by multiplying all generators by a factor (-1). Therefore, one can assume $\rho \geqslant 0$ and need only distinguish the cases $\rho = 0$ and $\rho = 1$ (this is no loss of generality since an overall scaling is possible). The admissible Lie algebras are thus parametrized by

$$[H, P_i] = \alpha K_i \qquad [H, K_i] = \lambda P_i \tag{3.103a}$$

$$[P_i, P_j] = \alpha\rho \epsilon_{ijk} J_k \quad [K_i, K_j] = -\lambda\rho \epsilon_{ijk} J_k \quad [P_i, K_j] = \rho \eta_{ij} H. \tag{3.103b}$$

---

[21] In this subsection I use the notation of Bacry and Lévy-Leblond. You can recover the notation of this book by making the replacements $J_k \to i J^k$, etc. I also denote the translation operator as P although, strictly speaking this is the conserved momentum associated to translation symmetry generator T.

[22] In a later publication [19] the requirement on parity and time reversal was dropped, given that these are not symmetries of nature–at least for weak interaction processes.

together with (3.102). As a further condition, one must impose that the Lorentz boosts generate non-compact subgroups. We saw in Sect. 3.2.1 that this is a necessary condition for establishing a causal order of events. Let me indicate how the further case differentiation identifies non-mutually isomorphic variants of Lie algebras:

1. Cases $\rho = 1$ ("relative-time" Lie algebras)

   - For $\alpha \neq 0, \lambda \neq 0$ there is always a suitable mapping of the generators in (3.102) to a set of ten generators $M_{AB} = -M_{BA}$ with $(A, B = 0, 1, 2, 3, 4)$:

     $$H = \eta M_{04}, \quad P_i = \pi M_{i4}, \quad K_i = \kappa M_{i0}, \quad J_i = \iota \epsilon_{ijk} M_{jk}$$

     with $(i, j = 1, 2, 3)$, so that the original algebra becomes

     $$[M_{AB}, M_{CD}] = \eta_{AD} M_{BC} - \eta_{AC} M_{BD} + \eta_{BC} M_{AD} - \eta_{BD} M_{AC}$$

     with semi-Euclidean metric coefficients η_{AB}. These commutation relations are typical of $\mathfrak{so}(p, 5-p)$ (see Appendix A.2.4). The different further cases (subject to the signs of α and λ) correspond to the algebras of $\mathbf{SO(5)}$, $\mathbf{SO(4,1)}$, $\mathbf{SO(3,2)}$. Actually there are two realizations of $\mathbf{SO(4,1)}$ depending on whether its subgroup $\mathbf{SO(4)}$ is generated by the J and the P or by the J and the K. With the requirement of non-compactness for the boost subgroup, the second $\mathbf{SO(4,1)}$ and the $\mathbf{SO(5)}$ are ruled out as kinematical groups.
   - For $\alpha = 0$ and $\lambda \neq 0$ the algebra (3.103) is isomorphic to either $\mathbf{SO(3,1)} \ltimes \mathbb{R}$ or $\mathbf{SO(4)} \ltimes \mathbb{R}$, depending on the sign of λ. Again, the latter group must be excluded, having non-compact boosts. And the former is just the Poincaré group.
   - The algebra with $\alpha = \lambda = 0$ is the Lie algebra of the Carroll group $\mathbf{C}$.
   - For the algebras with the remaining choices of α and λ see [18].

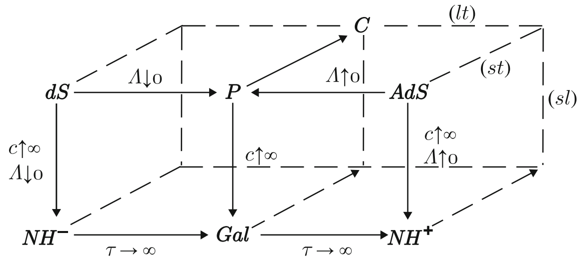
2. Cases $\rho = 0$ ("absolute-time" Lie algebras). For these $[P_i, P_j] = [K_i, K_j] = [P_i, K_j] = 0$.

   - For $\alpha \neq 0, \lambda \neq 0$, the algebra corresponds to the Newton-Hooke groups $\mathbf{NH}^{\pm}$, depending on the sign of α/λ.
   - For $\alpha = 0, \lambda \neq 0$ one gets an algebra isomorphic to the algebra of the Galilei group $\mathbf{Gal}$.
   - For the algebras with the remaining choices of α and λ see [18].

(Some) Mutual Contractions

In all, there are eleven kinematical groups according to the previous analysis. The most general structures identified are the de Sitter algebra $\mathfrak{so}(4, 1)$ and the anti-de Sitter algebra $\mathfrak{so}(3, 2)$. These can be contracted to the other algebras in the Bacry/Lévy-Leblond classification and were also expounded by those authors. According to the Inönu-Wigner procedure (see Appendix 2.4) a contraction is always defined with

Fig. 3.4 Contractions of kinematical groups



respect to a particular subalgebra. Since space isotropy should not be spoiled by contraction, the subalgebras should be rotationally invariant. As there are four of them, namely (J_i, H), (J_i, P_i), (J_i, K_i), and (J_i), there are four variants of physically motivated contractions. Let us consider the obviously most relevant ones:

- contraction with respect to (J_i, H): Rewrite the algebra (3.103) in terms of $\bar{P} = \epsilon P_i$, $\bar{K}_i = \epsilon K_i$ and then let $\epsilon \to \infty$. This rescaling can be interpreted as dealing with small length (or space intervals) and small speeds and is called (sl)-contraction in [18]. Because of $[\bar{P}_i, \bar{K}_j] = (\rho/\epsilon^2)\delta_{ij} H$, consistency requires that $\rho \to 0$. Therefore all relative-time algebras contract to their absolute-time partners. Specifically the de Sitter algebras contract to the Newton-Hooke algebras (see Fig. 3.3) and the Poincaré algebra contracts to the Galilei algebra.
- contraction with respect to (J_i, K_i): The limit procedure $H \to \epsilon H$, $P_i \to \epsilon P_i$ corresponds to small space and time intervals and is called a (lt)-contraction. One finds that in this case $\alpha \to 0$. In particular the de Sitter algebras contract to the Poincaré algebra, and $\mathbf{NH}^{\pm} \to \mathbf{Gal}$ (more precisely: their algebras).
- contraction with respect to (J_i, P_i): The limit rescaling of H and K_i corresponds to small speeds and small time intervals and is called a (st)-contraction. One finds $\lambda \to 0$. Specifically the Poincaré algebra contracts to the Carroll algebra.

Aside from the contractions mentioned in the previous list, there are many others, and a full picture is sketched in Fig. 3.4. We are aware about the Poincaré and the Galilei algebras, but what about the other ones arising in this context? The Newton-Hooke groups are for instance of interest as symmetry groups of non-relativistic cosmologies with non-vanishing cosmological constant [217] and for investigating Newton-Cartan spacetimes. Interestingly, the Poincaré algebra has two contractions. The Carroll group was discovered only in the 1960's [341]. Strangely enough, the following seems to have been overlooked: Starting from the Lorentz-boost expressions (3.4), it is usually argued that for $\beta = \mathfrak{v}/c \to 0$, these reduce to the Galilei transformations

$$\Delta x' = \Delta x - \mathfrak{v}\Delta t \qquad \Delta t' = \Delta t.$$

But this conclusion is only valid if, besides $\beta \ll 1$,

$$\beta \frac{\Delta x}{c\Delta t} \ll 1, \qquad \text{and} \qquad \beta \frac{c\Delta t}{\Delta x} \sim 1,$$

which is only true if $\Delta x \ll c\Delta t$. Thus one only is allowed to interpret the Galilei transformations as a $c \to \infty$ limit of the Lorentz boosts if the space intervals are much smaller than time intervals (in units $c = 1$). One could just as well consider the case $\Delta x \gg c\Delta t$ or

$$\beta \ll \frac{\Delta x}{c\Delta t} \qquad \text{and} \qquad \frac{c\Delta t}{\Delta x} \sim \beta \ll 1$$

for which the non-relativistic limit

$$\Delta x' = \Delta x \qquad\qquad \Delta t' = \Delta t - \frac{\mathfrak{v}}{c^2}\Delta x$$

arises. Therefore the Carroll group applies to kinematics for which space intervals are much larger than time intervals. This is physically a little strange, since in a world with this kinematical group all objects would have zero velocity, but their momenta must not vanish[23].

What are the groups belonging to the kinematical algebras? In principle these can be derived by applying the Baker-Campbell-Hausdorff procedure (see Appendix A.2.3). We know them already for the Poincaré and the Galilei algebras. The de Sitter group will be treated in the next section. The Carroll group elements and the group composition law as realized by coordinate transformations is derived in [341], and those for the Newton-Hooke groups in [18].

The aforementioned contraction scheme of the de Sitter algebras to the other kinematical algebras is purely group-theoretical and analytic. This scheme (almost) does not depend on how the Lie algebra generators are interpreted physically. Almost– since meaning was given to the generators J_i for spatial rotations and for the K_i as generators of Lorentz-boosts. If we assign to P_i the role of translation generators and to H as the generator of an evolution in time, the generators acquire a physical dimension in the following sense: In the group elements for the generators $X^a = \{H, \vec{P}, \vec{K}, \vec{J}\}$ written as $g = (\tau, \vec{a}, \vec{v}, \mathbf{R}) = \exp(\tau H + \vec{a}\vec{P} + \vec{v}\vec{K} + \vec{\theta}\vec{J})$, the arguments of the exponentials must be dimensionless. Thus $[H] = T^{-1}$, $[P] = L^{-1}$, $[K] = TL^{-1}$, $[J] = 1$, and the algebra (3.103) fixes the mass dimension of the coefficients as

$$[\alpha] = T^{-2} \qquad [\lambda] = 1 \qquad [\rho] = L^2.$$

[23] Lévy-Leblond realized that this kind of world was already described in the literature. Lewis Carroll [77]: “A slow sort of country,” ≫ said the Queen ≪. “Now, here, you see, it takes all the running you can do, to keep in the same place.”

If we introduce a reference velocity c and a reference length $R = \Lambda^{-1/2}$ the algebra (3.103) together with (3.102)–which because of $[J] = 1$ is already dimensionally correct -

$$
\begin{aligned}
&[J_i, J_j] = \epsilon_{ijk} J_k && [J_i, P_j] = \epsilon_{ijk} P_k && [J_i, K_j] = \epsilon_{ijk} K_k && [J_i, H] = 0 \\
&[H, P_i] = \tilde{\alpha}(c^2 \Lambda)\, K_i && [H, K_i] = \tilde{\lambda} P_i && && \\
&[P_i, P_j] = \tilde{\alpha}\tilde{\rho}\, \Lambda\, \epsilon_{ijk} J_k && [P_i, K_j] = \frac{\tilde{\rho}}{c^2}\, \delta_{ij} H && [K_i, K_j] = -\frac{\tilde{\lambda}\tilde{\rho}}{c^2}\, \epsilon_{ijk} J_k &&
\end{aligned}
$$

with numerical (dimensionless) coefficients $\tilde{\alpha}, \tilde{\lambda}, \tilde{\rho}$. Now we are able to mimic the algebra contraction by "playing" with the physically-motivated reference values c and Λ as depicted in Fig. 3.4.

- For $\Lambda \to 0$ and c finite, the Poincaré algebra results. If additionally $\tilde{\lambda} = 0$, the Carroll algebra results.
- For $\Lambda \to 0$ and $c \to \infty$ we distinguish two cases: (A) keeping $\tau := (c^2 \Lambda)^{-2}$ finite results in the Newton-Hooke algebras, (B) $\tau \to \infty$ yields the Galilei-algebra. The Newton-Hooke algebras involve one parameter τ carrying the dimension of time. They differ from the Galilei algebra only in the commutator $[H, P_i] = \pm \frac{1}{\tau^2} K_i$.

Off course one can still rescale the generators, and it was shown in [73] that this scaling allows to choose the non-vanishing free parameters as $\tilde{\alpha}, \tilde{\lambda}, \tilde{\rho} = \pm 1$.

Kinematical Spacetimes

In his so-called *Erlanger*[24] *Programm*, Felix Klein aimed to systematize all geometries/spaces known at the time. The idea is to investigate groups of transformations of a space onto itself or to adjoin to any geometry a group of transformations that leaves the geometry invariant. Thus the geometry of a manifold is characterized as the theory of invariants of a transformation group of that manifold. This yields a one-to-one relationship between a symmetry group and a geometry/space. The symmetry group associated to a geometry is called the isotropy group or its group of motion.

Bacry and Lévy-Leblond showed that all kinematical groups admit a four-dimensional spacetime interpretation. This is basically true, because by assumptions rotations and boosts form a subgroup of each of the kinematical groups. Thus for every kinematical group one can define a four-dimensional homogeneous space as the quotient of the group by the six-dimensional subgroup generated by $\{J_i, K_i\}$. In the case of the de Sitter groups and the Poincaré group, this six-dimensional group is the Lorentz group. In the cases of the Galilei, the Carroll and the Newton groups, it is - because of the vanishing commutator $[K_i, K_j]$ - the group $\mathbf{SO(3)} \ltimes \mathbb{R}^3$, and therefore specifically

[24] Named after the German city Erlangen, now in the Federal state of Bavaria.

- de Sitter space $dS(4, 1) = \mathbf{SO(4, 1)/Lor}$
- Minkowski space $\mathbb{M}^4 = \mathbf{P/Lor}$
- Aristotelian space = $\mathbf{Gal/(SO(3)} \ltimes \mathbb{R}^3)$.

We further recognize a correlation between the geometry, its symmetry group and adequate tensorial objects: Aristotelian geometry and $\mathbf{SO(3)}$ tensors (e.g. three-vectors), Minkowski geometry and Lorentz tensors. Later we will see that this correlation is also viable in Riemann geometry with diffeomorphism as symmetry group and Riemann tensors.

3.5 *Generalizations of Poincaré Symmetry

The Poincaré group can be generalized in two different ways: On the one hand, we may look for a group of which it is a subgroup (example: conformal group); and on the other hand we may investigate a group of which it is a contraction (de Sitter and anti-de Sitter groups).

3.5.1 Conformal Symmetry

The notion of conformal transformations came from various sources which at first glance are hard to understand in their mutual interrelations. For a historically inspired review, see [310], in which the roots ranging from conformal mappings of two-dimensional surfaces to observations about enlarged symmetry groups for Maxwell's equations (E. Bessel-Hagen), and to ideas about modified geometries for unifying electrodynamics and gravitation (H. Weyl) are carefully scouted.

Conformal Group

Consider a pseudo-Euclidean space $R^{(p,q)}$ of dimension $D = p + q$. By definition, the conformal group **C(p,q)** is the subgroup of coordinate transformations $x \mapsto x'$ which leave the metric $g_{\mu\nu}$ of this manifold invariant up to a scale factor:

$$\bar{g}_{\mu\nu}(\bar{x}) = \Omega(x) g_{\mu\nu}(x).$$

These transformations form the largest symmetry group that leaves the light cone $(d\tau)^2 = 0$ invariant: $g_{\mu\nu}dx^\mu dx^\nu = 0$ implies $\bar{g}_{\mu\nu}d\bar{x}^\mu d\bar{x}^\nu = 0$. They also preserve the angle

$$\frac{v \cdot w}{\left(v^2\, w^2\right)^{\frac{1}{2}}} \qquad \text{with} \qquad v \cdot w = g_{\mu\nu} v^\mu w^\nu$$

between two vectors v and w. The Poincaré group can easily be recognized as a subgroup of the conformal group since it leaves the metric invariant ($\bar{g}_{\mu\nu} = g_{\mu\nu}$ or $\Omega \equiv 1$).

Let us determine the generators of the conformal algebra. On the one hand, the change in the metric can be expressed as

$$\delta g_{\mu\nu} = \bar{g}_{\mu\nu}(\bar{x}) - g_{\mu\nu}(x) = \tilde{\Omega} g_{\mu\nu}$$

with an infinitesimal $\tilde{\Omega} = \Omega - 1$. On the other hand, under infinitesimal general coordinate transformation $x^\mu \mapsto x^\mu + \xi^\mu$, the metric changes as

$$\delta_\xi g_{\mu\nu} = -\partial_\mu \xi_\nu - \partial_\nu \xi_\mu .$$

This follows from

$$g_{\mu\nu} dx^\mu dx^\nu = g_{\mu\nu} \frac{\partial x^\mu}{\partial \bar{x}^\rho} d\bar{x}^\rho \frac{\partial x^\nu}{\partial \bar{x}^\sigma} d\bar{x}^\sigma = g_{\mu\nu} \left(\delta^\mu_\rho - \xi^\mu_{,\rho}\right) \left(\delta^\nu_\sigma - \xi^\nu_{,\sigma}\right) d\bar{x}^\rho d\bar{x}^\sigma = \bar{g}_{\rho\sigma} d\bar{x}^\rho d\bar{x}^\sigma .$$

On equating the two expressions for $\delta g_{\mu\nu}$ and contracting with $g^{\mu\nu}$, one finds

$$\tilde{\Omega}(x) = -\frac{2}{D}(\xi^\lambda{}_{,\lambda}). \tag{3.105}$$

The resulting equation

$$\partial_\mu \xi_\nu + \partial_\nu \xi_\mu = \frac{2}{D}(\xi^\lambda{}_{,\lambda})\eta_{\mu\nu} \tag{3.106}$$

is referred to as the conformal Killing equation. Here $g_{\mu\nu}$ was replaced by the Minkowski metric $\eta_{\mu\nu}$ since in fact we are interested in the conformal group in flat spacetime. To get insight into the consequences of this equation contract it in a first step with $\partial^\mu \partial^\nu$. This amounts to $(D-1)\Box\, \xi^\lambda_{,\lambda} = 0$. Since for $D = 1$ the condition (3.106) is trivially fulfilled for any ξ, we must require

$$\Box\, \xi^\lambda_{,\lambda} = 0 \qquad \text{for} \qquad \text{D} > 1. \tag{3.107}$$

In a second step, contract (3.106) with $\partial_\rho \partial^\nu$, with the result

$$\Box \partial_\rho \xi_\mu + (1 - \frac{2}{D}) \partial_\rho \partial_\mu \xi^\lambda_{,\lambda} = 0.$$

Now add to this the same equation with ρ and μ interchanged, and use (3.106) and (3.107) in order to find

$$(D-2)\partial_\rho \partial_\mu\, \xi^\lambda_{,\lambda} = 0.$$

The case $D = 2$ is dealt with briefly at the end of this subsection. For $D > 2$ the previous condition reads $\partial_\rho \partial_\mu \xi^\lambda_{\ ,\lambda} = 0$. As a last step take the un-contracted derivative $\partial_\rho \partial_\sigma$ of the conformal Killing equation (3.106). The result can be expressed in terms of an object $D_{\rho\sigma\mu\nu} = \partial_\rho \partial_\sigma \partial_\mu \xi_\nu$, and one shows that this tensor vanishes identically. Therefore for $D > 2$, the ξ^μ can be at most quadratic in the coordinates x. Substituting a power expansion of ξ^μ into (3.106) finally yields its most general form

$$\xi^\mu(x) = \epsilon^\mu + \omega^\mu{}_\nu x^\nu + \sigma x^\mu + \gamma^\nu(\delta^\mu_\nu x^2 - 2x^\mu x_\nu), \tag{3.108}$$

where $\{\epsilon^\mu, \omega^\mu{}_\nu, \sigma, \gamma^\mu\}$ are infinitesimal constants and $\omega^\mu{}_\nu$ is antisymmetric in its indices. Here we read off the Poincaré transformations with parameters (ϵ, ω), dilations (or, synonymously, scale transformations[25]) with parameters σ, and so-called special conformal transformations (parameter γ). The number of independent parameters for $\mathbf{C(p,q)}$ is $D + \frac{D(D-1)}{2} + 1 + D$, which is identical to $\frac{(D+1)(D+2)}{2}$. From (3.108), $\xi^\lambda_{\ ,\lambda} = D\sigma + 2D\gamma_\mu x^\mu$ results. Inserting this back into (3.105), we read off the result that $\tilde{\Omega} = 0$ for Poincaré transformations, $\tilde{\Omega} = -2\sigma$ for dilations and $\tilde{\Omega} = -4\gamma_\mu x^\mu$ for special conformal transformations.

In $D = 4$ dimensions, the conformal transformations of $\mathbf{C(3,1)}$ are characterized by 15 parameters. These stand for the $4 + 6$ parameters of the Poincaré group, 1 parameter for dilations and the 4 parameters for special conformal transformations. The algebra $\mathfrak{c}(3,1)$ is spanned by the ten generators $\{T_\mu, M_{\mu\nu}\}$ of the Poincarè algebra and five further generators S and C_μ. Since the coordinate transformations are generated by $X = i\xi^\mu \partial_\mu$, the descriptor (3.108) yields for the generators of the dilations and the special conformal transformations

$$S = i x^\nu \partial_\nu, \qquad C_\mu = i(x^2 \partial_\mu - 2x_\mu x^\nu \partial_\nu). \tag{3.109}$$

As for the finite transformations on the coordinates, T_μ generate spacetime translations $\hat{x}_\mu + a_\mu$, $M_{\mu\nu}$ generate Lorentz rotations $\hat{x}_\mu = \Lambda^\nu_\mu x_\nu$, S generates scale transformations $\hat{x}_\mu = e^\kappa x_\mu$, and the$C_\mu$ generate special conformal transformations

$$\hat{x}_\mu = \frac{x_\mu - c_\mu x^\nu x_\nu}{1 - 2c^\nu x_\nu + (c^\nu c_\nu)(x^\lambda x_\lambda)}.$$

The latter can be obtained by subjecting x_μ successively to an inversion, a translation and a second inversion:

$$\begin{aligned} x_\mu \ \to\ x'_\mu = \frac{x_\mu}{x_\lambda x^\lambda} \ \to\ x''_\mu = x'_\mu - c_\mu \ \to\ x'''_\mu &= \frac{x''_\mu}{x''_\lambda x''^\lambda} \\ &= \frac{x_\mu - c_\mu x^\nu x_\nu}{1 - 2c^\nu x_\nu + (c^\nu c_\nu)(x^\lambda x_\lambda)}. \end{aligned}$$

[25] Also, the term "dilatations" is used; this refers to British English.

Since the Poincaré group is a subgroup of the conformal group, the algebra involving the $\{T_\mu, M_{\mu\nu}\}$ generators is unchanged. The full algebra can for instance be calculated from the explicit realizations (3.98, 3.109). The additional non-vanishing Lie brackets are

$$
\begin{aligned}
[T_\mu, S] &= iT_\mu \\
[T_\mu, C_\nu] &= 2i(M_{\mu\nu} - \eta_{\mu\nu} S) \\
[M_{\mu\nu}, C_\rho] &= i(\eta_{\nu\rho} C_\mu - \eta_{\mu\rho} C_\nu) \\
[S, C_\mu] &= iC_\mu .
\end{aligned}
$$

The commutator $[T_\mu, C_\nu]$ reveals that a Poincaré-invariant theory which is invariant under special conformal transformations is necessarily also scale invariant, and thus has the full conformal group as a symmetry group. The reverse is not true in general: There are theories which are scale invariant, but not invariant under special conformal transformations.

The conformal group **C**(**3,1**) is isomorphic to **SO**(**4,2**) of which **SO**(**3,1**) is a subgroup. If the generators of **SO**(**4,2**) are denoted as $\overline{M}_{AB}$ with $A, B = 0, \ldots 5$ and with $\eta_{AB} = (\eta_{\mu\nu}, -1, 1)$, the isomorphism becomes manifest with the identification

$$
\begin{aligned}
M_{\mu\nu} &\rightleftharpoons \overline{M}_{\mu\nu} & \qquad T_\mu &\rightleftharpoons \overline{M}_{\mu 4} + \overline{M}_{\mu 5} \\
C_\mu &\rightleftharpoons \overline{M}_{\mu 4} - \overline{M}_{\mu 5} & \qquad S &\rightleftharpoons \overline{M}_{45} .
\end{aligned}
$$

Choosing $\gamma^\nu \equiv 0$ in the transformations (3.108) gives rise to a subgroup in the conformal group, namely the *Weyl group*. It is composed of the Poincaré transformations and scale transformations with the only additional non-vanishing Lie bracket $[T_\mu, S] = iT_\mu$.

Conformal Field Theory

Conformal invariance was introduced in physics when in 1909/10, H. Bateman and E. Cunningham described in several articles that Maxwell's equations have a symmetry group even larger than what we call today the Poincaré group. In 1921, E. Bessel-Hagen applied and extended Noether's second theorem in order to find the (fifteen) conservation laws related to this larger symmetry group [40]. He did this–as he writes in his article–on the request by "*Herrn Geheimrat*" F. Klein. For details about the articles by Bateman and Cunningham and for their reception in the article by Bessel-Hagen, see [310]. Instead of following the ingenious calculations by Bessel-Hagen, we can derive his results in a modern notation.

The behavior of fields Q (possible indices suppressed) with respect to the transformations of the conformal group is

$$\bar{\delta}_\epsilon Q = i\epsilon^\nu P_\nu Q \qquad \bar{\delta}_\omega Q = -\frac{i}{2}\omega^{\mu\nu}(L_{\mu\nu} + \Sigma_{\mu\nu})Q$$
$$\bar{\delta}_\sigma Q = i\sigma(S + \hat{S})Q \qquad \bar{\delta}_\gamma Q = i\gamma^\mu(C_\mu + \hat{C}_\mu)Q.$$

Here, it is assumed that the transformations might pick up extra pieces similar to the Σ spin part for Lorentz transformations. And pretty much as $\Sigma_{\mu\nu}$ is a representation of the $L_{\mu\nu}$, $\hat{S}$ and $\hat{C}_\mu$ are to be understood as representations of S and C_μ. Taking commutators of two successive infinitesimal transformations must yield the algebra of the conformal group. This requires $\hat{S} = -id\mathbb{I}$, $\hat{C}_\mu = 2i(\Sigma_{\mu\nu} - \eta_{\mu\nu}d\,\mathbb{I})x^\nu$. The constant d is called the scale dimension, or the 'conformal degree' of the field. It is chosen in such a way that in the action, the kinetic term for Q is scale invariant. By arguments explained in Sect. 5.3.1, for bosonic fields it is related to the spacetime dimension D by $d = (D-2)/2$; and for fermionic fields, it is $d = 3/2$. In a linear representation, the algebra of the conformal group is thus

$$T_\mu = i\partial_\mu \qquad M_{\mu\nu} = x_\mu P_\nu - x_\nu P_\mu + \Sigma_{\mu\nu}$$
$$S = x^\nu P_\nu + id \qquad C_\mu = 2x_\mu S - x^2 P_\mu + 2\Sigma_{\mu\nu}x^\nu.$$

Let me clarify a point here concerning the relationship between the conformal transformations of fields as stated above and their transformation behavior under general coordinate transformations. For instance for a vector field we have the scale transformation

$$\bar{\delta}_\sigma A_\mu = -\sigma x^\lambda A_{\mu,\lambda} - \sigma d A_\mu.$$

On the other hand, under infinitesimal general coordinate transformations a vector field transforms as[26]

$$\bar{\delta}_\xi A_\mu = -\xi^\nu_{\ ,\mu} A_\nu - A_{\mu,\nu}\xi^\nu.$$

Certainly, the conformal transformations (3.108) are specific general coordinate transformations. For the dilation part, we find with $\xi^\mu = \sigma x^\mu$:

$$\bar{\delta}_{\xi(\sigma)} A_\mu = -\sigma A_\nu - \sigma x^\nu A_{\mu,\nu}.$$

[26] More about this in Sect. 7.3.1.

Thus $\bar{\delta}_\sigma A_\mu = \bar{\delta}_{\xi(\sigma)} A_\mu$ only in the case $d = 1$, that is for $D = 4$. In other dimensions one can cure the defect by assuming that the A_μ transform as vector densities with a certain weight W. Then

$$\bar{\delta}_{\xi(\sigma)} A_\mu = W A_\mu \xi^\lambda{}_{,\lambda}(\sigma) - \sigma d A_\mu - \sigma x^\nu A_{\mu,\nu} = \sigma(WD - 1) A_\mu - \sigma x^\nu A_{\mu,\nu},$$

so that conformal transformations on the fields can be identified with general coordinate transformations if one associates the weight $W = \frac{4-D}{D}$. This remark was admittedly rather sketchy. The full story is to define an extended conformal group in a so-called Weyl space; see [205].

Let me come back to the question of conformal invariance of free electrodynamics. The scale transformations of the $F_{\mu\nu}$ as derived from the transformations of A_μ are calculated to be

$$\bar{\delta}_\sigma F_{\mu\nu} = -\sigma \left(x^\lambda \partial_\lambda + \frac{D}{2} \right) F_{\mu\nu}.$$

Thus

$$\begin{aligned} \bar{\delta}_\sigma \mathcal{L}_{ED} &= -\frac{1}{2} F^{\mu\nu} \bar{\delta}_\sigma F_{\mu\nu} = \frac{\sigma D}{4} F^{\mu\nu} F_{\mu\nu} + \frac{\sigma}{2} F^{\mu\nu} x^\lambda \partial_\lambda F_{\mu\nu} \\ &= \frac{\sigma D}{4} F^{\mu\nu} F_{\mu\nu} + \frac{\sigma}{4} \partial_\lambda (F^{\mu\nu} F_{\mu\nu}) = -\partial_\lambda (\sigma x^\lambda \mathcal{L}_{ED}) \end{aligned}$$

which shows that free electrodynamics is scale (quasi-)invariant in any spacetime dimension. As for special conformal transformations the calculations become somewhat more tedious and I simply quote the result from [297]

$$\bar{\delta}_\gamma \mathcal{L}_{ED} = \partial_\mu [\gamma_\sigma (2x^\sigma x^\mu - \eta^{\sigma\mu} x^2) \mathcal{L}_{ED}] + (4 - D) \gamma_\sigma F^{\sigma\rho} A_\rho.$$

Thus the action of Maxwell's theory is invariant under special conformal transformations only for $D = 4$. What about the Noether currents associated to the additional symmetries? Certainly they can be determined directly from the generic expression (3.49). As elucidated before, the currents are not unique: One may add terms which either vanish on-shell or terms whose divergence vanishes identically. By adding appropriate expressions R. Jackiw and S-Y. Pi [297] arrived at the form of the conformal currents as

$$J^\mu_\xi = -B^{\mu\nu} \xi_\nu + \frac{D-4}{2D} \partial_\nu \Big(\xi^\nu F^{\mu\rho} A_\rho \Big).$$

For $D = 4$ we rediscover the structure of eq.(3.65), now extended to conformal transformations. The divergence of this current is

$$\partial_\mu J^\mu_\xi = -B^{\mu\nu} \partial_\mu \xi_\nu + \frac{D-4}{2D} (\partial_\mu \partial_\nu \xi^\nu) F^{\mu\rho} A_\rho + \frac{D-4}{2D} (\partial_\nu \xi^\nu) F^{\mu\rho} \partial_\mu A_\rho,$$

where the property $\partial_\mu B^{\mu\nu} = 0$ of the Belinfante tensor and the field equations $\partial_\mu F^{\mu\rho} = 0$ were used. Because of the symmetry of the Belinfante tensor and because of the conformal Killing condition (3.106) the first term can be written as

$$B^{\mu\nu}\partial_\mu\xi_\nu = \frac{1}{D}B^\mu{}_\mu\partial_\nu\xi^\nu = -\frac{D-4}{4D}(\partial_\nu\xi^\nu)F^{\mu\rho}F_{\mu\rho}$$

and therefore this cancels with the third one. We are left with

$$\partial_\mu J^\mu_\xi = \frac{D-4}{2D}(\partial_\mu\partial_\nu\xi^\nu)F^{\mu\rho}A_\rho.$$

This vanishes for translations, Lorentz rotations, and scale transformations since in these cases $\partial_\mu\partial_\nu\xi^\nu = 0$. And–in agreement with earlier findings–it vanishes for special conformal transformations only in four dimensions.

Previously, we saw that the properties of the Belinfante tensor (no divergence, index symmetry) are explicitly related to invariance under translations and under Lorentz rotations. Pretty much as the Belinfante tensor is an "improved" canonical energy-momentum tensor, the Belinfante tensor can be "improved". C. Callan, S. Coleman and R. Jackiw [66] gave a recipe for constructing an energy-momentum tensor $T^{\mu\nu}_{CCJ}$ which is divergence-free, symmetric and traceless within conformal invariant theories, such that $J^\mu_\xi = -T^{\mu\nu}_{CCJ}\xi_\nu$. We saw that for $D = 4$, the electromagnetic Belinfante tensor is traceless and thus serves as the CCJ-tensor. The recipe of "improving" energy-momentum tensors seems to be somewhat *ad hoc*. A generic differential-geometry-based approach is described in [191]. In Sect. 7.5.2, we will meet still another energy-momentum tensor, namely the Hilbert tensor. As a matter of fact there is a whole family of Hilbert tensors. They coincide on-shell, but reflect conformal symmetry in different ways; see [420].

In order to see that the requirement of scale invariance of a Lagrangian restricts its dependence on the fields, consider the simplest case, the Lagrangian for a scalar field (3.61). From $\bar{\delta}_\sigma\varphi = -\sigma(x^\lambda\partial_\lambda + d)\varphi$, one calculates

$$\bar{\delta}_\sigma\mathcal{L}_\varphi = -\partial_\lambda(\sigma x^\lambda\mathcal{L}_\varphi) + \sigma\left(d(\frac{dV}{d\varphi})\varphi - DV\right).$$

Only if the last term vanishes is the Lagrangian quasi-invariant. And the vanishing of this term entails a condition on the functional dependence of the potential. For $D = 2$ the potential must be identically zero. For $D = 4$ a potential $V \propto \varphi^4$ is admissible, but a mass term being proportional to φ^2 would break scale invariance.

The conformal group is the symmetry group of all physical systems that have no intrinsic masses or dimensional coupling constants. Since at high energies the rest energies (i.e. masses) can be neglected, the conformal group becomes approximately valid also for particles with non-vanishing rest mass. This is one of the reasons for establishing conformal invariance in field theories as a research field in its own.

Infinite Series of Conservation Laws for Electrodynamics

The next point seems to lead a little outside the context of conformal symmetries. In 1964, D.M. Lipkin found "a new conservation law in electromagnetic theory" not related to the energy-momentum tensor, but characterized by what he called the "zilch" tensor. For its definition introduce

$$Z^{\mu}{}_{\nu\rho} = {}^*F^{\mu\sigma} F_{\sigma\nu,\rho} - F^{\mu\sigma *} F_{\sigma\nu,\rho},$$

an expression bi-linear in the electromagnetic field strength, its dual and derivatives thereof. Here ${}^*F^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma}$, $F_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}{}^*F^{\rho\sigma}$. By definition, the Z tensor is symmetric in its first two indices, $Z^{\mu\nu}{}_{\rho} = Z^{\nu\mu}{}_{\rho}$ and also traceless in the same pair, $Z^{\mu}{}_{\mu\rho} = 0$. On solutions of the free Maxwell equations, $Z^{\mu\rho}{}_{\rho} = 0$. Thus all contractions vanish. Further it can be derived that on-shell $Z^{\mu\nu\rho}{}_{,\rho} = 0$. Therefore there are charges

$$Z^{\mu\nu} = \int d^3x \, Z^{\mu\nu 0}.$$

These are the ten components of the zilch tensor. A physical interpretation in terms of the conserved polarization of the electromagnetic field can be established.

In [206] it is shown that together with the conserved currents from conformal symmetry, the currents related to the zilch are just the first members in an infinite series of conserved quantities for the free electromagnetic field. These are bilinear forms of the type $J_{\mu}^{(m)}(D^n F, D^k F)$, where $D^n F$ stands for the n-th derivative of the field strength tensor $F_{\mu\nu}$.

- There are exactly 15 conserved currents of zero order $(n = 0 = k)$ depending on products of two field strength components and which are of the form $J_{\mu}^{(0)} = K^{\nu} B_{\nu\mu}$. Here B is the Belinfante energy-momentum tensor, and K obeys $\partial^{\nu} K^{\mu} + \partial^{\mu} K^{\nu} - \frac{1}{2} g^{\mu\nu} \partial_{\lambda} K^{\lambda} = 0$. This is the conformal Killing equation (3.106) for $D = 4$, and we rediscover the conformal currents.
- There exist exactly 84 conserved currents of first order which can be represented as

$$J_{\mu}^{(0)} = K^{\sigma\nu} Z_{\sigma\nu\mu} + 2\epsilon_{\mu\nu\lambda\sigma}(\partial^{\lambda} K^{\rho\nu}) B^{\sigma}_{\rho}$$

where $(Z_{\sigma\nu\mu})$ is the tensor introduced by Lipkin, and where $K^{\sigma\nu}$ is a Killing tensor obeying

$$\partial^{(\mu} K^{\sigma\nu)} = \frac{1}{3} \partial_{\lambda} K^{\lambda(\mu} g^{\sigma\nu)} \qquad K^{\sigma\nu} = K^{\nu\sigma} \qquad K^{\mu}_{\ \mu} = 0.$$

We saw that the conformal Killing equation led to transformations described by a second order polynomial in x^{μ}. In this case, one is led to a fourth-order polynomial.

The set of conserved currents includes the zilch-induced currents from above plus further "new" currents. These new currents correspond to those transformations which include the fourth powers. For some of the currents it was shown that they can be derived by Noether's theorem with field transformations that depend on second derivatives of the fields.

- For $J_\mu^{(m)}$ with $m > 1$ a generic expression in terms of conformal Killing tensors and generalizations of the energy-momentum and the zilch tensor can be found in the literature[27].

Conformal Algebra in D = 2

In the two-dimensional case the conditions (3.106) on ξ_μ become explicitly

$$\partial_1\xi_1 = \partial_2\xi_2 \quad \partial_1\xi_2 = -\partial_2\xi_1,$$

known as Cauchy-Riemann equations in the theory of complex functions. Introduce complex coordinates $(z, \bar{z} = x^1 \pm ix^2)$ and write $\xi(z) = \xi^1 + i\xi^2$ and $\bar{\xi}(\bar{z}) = \xi^1 - i\xi^2$. Then the two dimensional transformations are none other than analytic coordinate transformations $z \to f(z),\ \bar{z} \to \bar{f}(\bar{z})$. Thus, in $D = 2$ all holomorphic functions ξ and $\bar{\xi}$ are solutions of the conformal Killing equations. In this case we have an infinite number of transformations accompanied by an infinite number of associated conserved quantities.

The algebra of these transformations can be numbered by $n \in \mathbb{Z}$: Expand the solutions for e.g. $\xi(z)$ as a Taylor series

$$\xi(z) = \sum_{n \geqslant -1} a_n z^{n+1}$$

and calculate the variation of a scalar analytic function $F'(z') = F(z)$ as

$$\bar{\delta} F = -a_n z^{n+1}\partial_z F =: a_n L_n F.$$

The $L_n = -z^{n+1}\partial_z$ are generators of the infinite conformal symmetry and satisfy

$$[L_m, L_n] = (m - n)L_{m+n}.$$

This can be extended to $n < -1$. Identical relations with generators $\bar{L}_n$ follow for $\bar{\xi}(\bar{z})$, and the L_n and $\bar{L}_n$ commute. The full algebra $(L_n, \bar{L}_n)$ became known as

---

[27] I'm not aware whether all the infinite conservation laws can be derived from Noether's theorem, how general the symmetry transformations are to be, and which symmetry groups are at work. I'm also not aware of a physical interpretation of the additional symmetries. It seems in any case that most of these conservation laws do not carry over to the case of interacting fields.

Virasoro algebra. It made its appearance in physics for the first time in the dual string model in particle physics[28] because the action for the relativistic string is proportional to its 2D-worlsheet; some more details are given in Sect. 8.4.2.

3.5.2 de Sitter Group

It was H. Minkowski who, in his famous speech "*Raum und Zeit*" [372] emphasized the group-theoretical clash between Newton's and Maxwell's physics, the former being invariant under the Galilei group $\mathbf{G}_\infty$, the latter under the Lorentz group $\mathbf{G_c}$ (where I adopt the naming of the two groups by Minkowski). He pointed out that "..since $\mathbf{G_c}$ is mathematically more intelligible than $\mathbf{G}_\infty$ it looks as though the thought might have struck some mathematician, fancy-free, that after all, as a matter of fact, natural phenomena do not possess an invariance with the group $\mathbf{G}_\infty$ but rather with the group $\mathbf{G_c}$, c being finite ..."[29] In effect, the discovery of special relativity and a natural constant c with the dimension of velocity could have occurred just by mathematical boldness, but it did not. Freeman J. Dyson called this a "missed opportunity" [143] and stressed that even Minkowski was not bold enough. Minkowski certainly was aware that it is not the Lorentz group but the Poincaré group under which electrodynamics is invariant. And this again is mathematically a nasty (or "non-intelligible") entity, being the direct sum of the Lorentz and the translation group. Why did Minkowski not again conjure into existence a mathematically more intelligible group? He could have found the de Sitter[30] groups. These groups are simple and, as shown in Sect. 3.4.5, degenerate into the non-semisimple Poincaré group by contraction. The de Sitter groups, however, were not discovered by a mathematician, but rather came under discussion in the context of general relativity and cosmology: As will be shown in Sect. 7.5.2., $\mathbf{SO(4,1)}$ and $\mathbf{SO(3,2)}$ are the invariance groups of an empty expanding–or contracting–universe whose radius of curvature $\mathcal{R}$ is a linear function of time. This radius is related to the cosmological constant Λ by $|\mathcal{R}| = \sqrt{3/\Lambda}$. On the other hand, in the de Sitter groups the quantity $\mathcal{R}$ is a parameter, just as c is a parameter in the Poincaré group.

Indeed, according to surprising findings by the astronomer teams led by S. Perlmutter, B. P. Schmidt, and A. G. Riess[31] from the late 1990's, our universe is expanding at an accelerating rate. Although a non-vanishing cosmological constant is by far not the only explanation of the expansion rate, we should face the possibility

---

[28] In the quantum case this algebra receives a further term proportional to a central charge. This depends on the dimension of spacetime and vanishes for $D = 26$.

[29] "..da $\mathbf{G_c}$ *mathematisch verständlicher ist als* $\mathbf{G}_\infty$, *hätte wohl ein Mathematiker in freier Phantasie auf den Gedanken verfallen können, daβ am Ende die Naturerscheinungen tatsächlich eine Invarianz nicht bei der Gruppe* $\mathbf{G}_\infty$, *sondern vielmehr bei einer Gruppe* $\mathbf{G_c}$ *mit bestimmtem endlichen .... c besitzen."*

[30] Named after the Dutch astronomer Willem de Sitter (1872–1934)

[31] It seems that this is one of the rare circumstances where in 2011, a Nobel prize was awarded for increasing our nescience. The discovery led to the term "dark energy": At least 70 % of the matter-energy content of the universe cannot yet be explained.

that de Sitter relativity has to replace Minkowski relativity [4]. This would have drastic consequences for our current description of fundamental physics. Since the present formulation of quantum field theories is based on the strict validity of Poincaré invariance, its foundation possibly needs to be rewritten. We may feel "safe" because the cosmological constant which effectively describes the expansion is very small, and thus the deviation from exact Poincaré invariance is negligible.

de Sitter Spacetime

It was mentioned already that the de Sitter spacetime $dS(4,1)$ can be represented as the homogeneous space $dS(4,1) = \mathbf{SO(4,1)}/\mathbf{SO(3,1)}$,, that is as the quotient between the de Sitter and the Lorentz group. What does this mean in geometric terms and in coordinates? Consider a five-dimensional pseudo-Euclidean space $\mathbb{E}^{4,1}$ with cartesian coordinates x^A $(A = 0, 1, 2, 3, 4)$ and metric with signature $(+1, -1, -1, -1, -1)$. The de Sitter space is the hypersurface H_4 embedded in this space by

$$\eta_{AB}x^A x^B = \eta_{\mu\nu}x^\mu x^\nu - (x^4)^2 = -\mathcal{R}^2 \qquad (\mu, \nu = 0, 1, 2, 3)$$

with the de Sitter "pseudo-radius" $\mathcal{R}$. Due to $dx^4 = (\eta_{\mu\nu}x^\mu dx^\nu)/(x^4)$, the line element of H_4 becomes

$$ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu - \frac{(\eta_{\nu\nu}x^\mu dx^\nu)^2}{\mathcal{R}^2 + \sigma^2} \tag{3.110}$$

where $\sigma^2 := \eta_{\mu\nu}x^\mu x^\nu$. Similar relations hold for an anti-de Sitter space $dS(3,2) = \mathbf{SO(3,2)}/\mathbf{SO(3,1)}$:

$$\eta_{\mu\nu}x^\mu x^\nu + (x^4)^2 = +\mathcal{R}^2 \qquad ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu + \frac{(\eta_{\mu\nu}x^\mu dx^\nu)^2}{\mathcal{R}^2 - \sigma^2}.$$

In the following, only the de Sitter case will be treated, with the understanding that the anti-de Sitter case can be described analogously. Instead of using the coordinates $\{x^A\}$, it is often more appropriate to work with conformal coordinates $\{\chi^A\}$. These are defined by the stereographic projection from the hypersurface into a Minkowski spacetime. When the Minkowski plane is located at $x^4 = 0$, the projection is given by (see [249])

$$\chi^\mu(x) = \Omega(\sigma^2)x^\mu \qquad \chi^4(x) = -\mathcal{R}\Omega(\sigma^2)\Big(1 + \frac{\sigma^2}{4\mathcal{R}^2}\Big), \tag{3.111}$$

where

$$\Omega(\sigma^2) = \frac{1}{1 - \sigma^2/4\rho^2}.$$

The inverse transformation to (3.111) is

$$x^\mu(\chi) = \Omega^{-1}(\chi^4)\chi^\mu \qquad \Omega(\chi^4) = \frac{1}{2}\Big(1 - \frac{\chi^4}{\mathcal{R}}\Big).$$

In conformal coordinates, the de Sitter line element becomes $ds^2 = \Omega^2 \eta_{AB} d\chi^A d\chi^B$, revealing that de Sitter space is conformally flat. This agrees with results from differential geometry from which it is known that (i) isotropic and homogeneous spaces are maximally symmetric (exhibit the maximum of Killing symmetries), (ii) maximally symmetric spaces are characterized by a constant curvature, (iii) spaces of constant curvature are conformally flat.

Since for the de Sitter space, in contrast to Minkowski space, there is no preferential choice of coordinates you will find other choices $\{X^A\}$ in the literature (e.g. Beltrami coordinates, pseudo-spherical coordinates) besides the "natural" ($\{X^A\} = \{x^A\}$) and the conformal ($\{X^A\} = \{\chi^A\}$) one. This choice has an influence on how the generators of the de Sitter Lie algebra are realized on fields.

The subalgebra structure of the de Sitter group can be made visible by rewriting the canonical algebra for the ten parameter group **SO(4, 1)**

$$[M_{AB}, M_{CD}] = i(\eta_{AD} M_{BC} - \eta_{AC} M_{BD} + \eta_{BC} M_{AD} - \eta_{BD} M_{AC}).$$

Defining

$$T_\mu = \mathcal{R}^{-1}\, M_{4\mu}$$

the algebra is rewritten as

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\sigma} M_{\nu\rho} - \eta_{\mu\rho} M_{\nu\sigma} + \eta_{\nu\rho} M_{\mu\sigma} - \eta_{\nu\sigma} M_{\mu\rho}) \tag{3.112a}$$

$$\big[M_{\mu\nu}, T_\lambda\big] = i(\eta_{\lambda\nu} T_\mu - \eta_{\lambda\mu} T_\nu) \tag{3.112b}$$

$$\big[T_\mu, T_\nu\big] = i\frac{1}{\mathcal{R}^2} M_{\mu\nu}. \tag{3.112c}$$

The factor $1/\mathcal{R}^2$ in the last commutator reveals in which sense the de Sitter algebra is a deformed Poincaré algebra. The generators of $\mathfrak{so}(4, 1)$ can be realized as

$$M_{AB} = \eta_{AC} X^C \Pi_B - \eta_{BC} X^C \Pi_A$$

where Π_A is the momentum canonically conjugate to the coordinate X^A:$\{X^A, \Pi_B\} = \delta^A_B$. Since, due to (3.112a–3.112c), the Lorentz subalgebra of the de Sitter algebra is identical to the flat-space Lorentz algebra, its generators may be expressed through the 4-dimensional coordinates x^μ and momenta: $p_\mu = i\partial_\mu$, $M_{\mu\nu} = i(x_\mu p_\nu - x_\nu p_\mu)$. The realization of T_μ in (3.112a–3.112c) depends on the specific choice of the coordinates on H_4.

- For the "natural" coordinates

$$\{x^\mu, T_\nu\} = \frac{1}{\mathcal{R}}\{x^\mu, M_{\nu 4}\} = \frac{1}{\mathcal{R}}\{x^\mu, \eta_{4C}\, x^C\, p_\nu\} = -\frac{1}{\mathcal{R}}\delta^\mu_\nu x^4.$$

  In using $(x^4)^2 = (\mathcal{R}^2 + \sigma^2)$ we find the relation between the T_μ and the p_μ as

$$T_\mu = \sqrt{\left(1 + \frac{\sigma^2}{\mathcal{R}^2}\right)}\, p_\mu$$

  exhibiting a nonlinear realization of the generators.
- For the conformal coordinates a similar line of reasoning results in

$$\begin{aligned}\{x^\mu, T_\nu\} &= \frac{1}{\mathcal{R}}\{\Omega^{-1}\chi^\mu, \eta_{44}\chi^4\pi_\nu - \chi_\nu\pi_4\} = \frac{1}{\mathcal{R}}\left[-\Omega^{-1}\chi^4\{\chi^\mu, \pi_\nu\} - \chi^\mu\chi_\nu\{\Omega^{-1}, \pi_4\}\right] \\ &= \frac{1}{\mathcal{R}}\left[-\Omega^{-1}\chi^4\delta^\mu_\nu - \Omega^{-2}\frac{1}{2\mathcal{R}}\chi^\mu\chi_\nu\right] = \left(1 + \frac{\sigma^2}{4\mathcal{R}^2}\right)\delta^\mu_\nu - \frac{1}{2\mathcal{R}^2}x^\mu x_\nu\end{aligned}$$

  so that with the canonically-conjugate momenta $p_\mu = \Omega(\xi^4)\pi_\mu$

$$\begin{aligned}T_\mu &= \left(1 + \frac{\sigma^2}{4\mathcal{R}^2}\right)p_\mu - \frac{1}{2\mathcal{R}^2}x_\mu x^\nu p_\nu \\ &= p_\mu + \frac{1}{4\mathcal{R}^2}\left[\sigma^2\delta^\nu_\mu - 2x_\mu x^\nu\right]p_\nu = p_\mu + \frac{1}{4\mathcal{R}^2}C_\mu.\end{aligned}$$

  Therefore, in conformal coordinates the translations in de Sitter space are composed of the ordinary translations p_μ and a special conformal transformation C_μ.

In de Sitter spacetime, the very fact that the translation generator T_μ is not identical to the Minkowski momentum entails modified notions of energy and momentum (see e.g. [4]). Also, the relation between the special relativistic energy E_R and momentum p_R is no longer the dispersion relation (3.19) but for large values of the de Sitter radius it entails an extra term proportional to $1/\mathcal{R}^2$ according to

$$\frac{E_R^2}{c^2} - p_R^2 = m^2c^2 + \frac{1}{\mathcal{R}^2}\mathcal{D}.$$

As shown in Sect. 7.5.2, Minkowski spacetime is a solution of the gravitational sourceless field equations without a cosmological constant, whereas de Sitter spacetime is a solution of the vacuum field equations with a cosmological constant.

3.6 On the Validity of Special Relativity

Even if SRT is frequently praised as one of the great intellectual achievements of the 20th century, it is not at all sacrosanct but–like any other physical model or theory–must be tested experimentally with ever-increasing precision. So far, it has survived all tests with an astounding accuracy. In [561], a deviation from Lorentz invariance is parametrized in terms of $\delta = (c^{-2} - 1)$, and evidenced as $\delta < 10^{-20}$. But perhaps tomorrow one will find in a deviation of an experimental result from a theoretical prediction of special relativity. This of course would imply that Poincaré symmetry is broken on a certain scale.

Observational hints for this symmetry breaking from ultra-high-energy cosmic ray events may already have been detected. The energy observed for some particles exceeds a theoretical upper limit (the so-called GZK-limit) and points to the possibility that ultra-high-energy photons travel slower than low-energy ones.

Even from a theoretical point of view, there are reasons to doubt the universal validity of special relativity. For one thing, why should Poincaré symmetry be exact, given that most of the other symmetries are broken? Furthermore, from the following chain of arguments, it would even be a surprise if special relativity is valid on all scales: SRT is a limiting case of GR in the sense that, in one version of the equivalence principle, the Riemann geometry of general relativity is locally a Minkowski geometry. However, GR is conceptually not compatible with the principles of quantum physics. If one believes in the unity of physics, one needs to formulate a "quantum gravity". Thus the validity of special relativity–and by this the exactness of Poincaré symmetry–depends on answering the question "What is the semiclassical, flat-space limit of quantum gravity?" [287]. A related argument is that of the existence of a fundamental length parameter–the Planck length L_{Pl} - which is not compatible with the prediction of length contraction from SRT.

On the other hand, the consequences of broken Poincaré symmetry would be severe: As will be shown in the next chapters, Poincaré invariance is a cornerstone in relativistic field theories and thus both in the axiomatic formulation of field theories and in today's theories of the fundamental interactions. One would even no longer know how to define a particle and would have to abandon the CPT theorem in its current form.

There are various approaches for describing or understanding a possible violation of Poincaré invariance:

- SME
  The SME program attempts to describe violations of Lorentz invariance in a model- and theory-independent way [52]: It introduces all Lorentz-violating interaction terms into the Lagrangian of the standard model of particle physics. Indeed, SME stands for "Standard Model Extension".

- DSR
  DSR is an extension of special relativity in the sense that not only the velocity of light, but also another entity, for example the Planck length plays a distinguished role. In the various DSR models there are accordingly two observer-independent scales with dimension velocity and length (or energy, or momentum). The D in DS(pecial)R(elativy) stands either for Doubly (because of the two parameters) or for Deformed (because of the deformation of the Lorentz algebra). A good review on the various DSR models is given in [287].
- de Sitter special relativity
  Another specific way to introduce a further parameter into a modified SRT is based on a de Sitter space approach. As described in the previous subsection, the de Sitter space differs from Minkowski space by having a non-vanishing curvature, characterized by the parameter $\mathcal{R}$ (with the dimension of a length). Since the curvature is related to the cosmological constant and since astrophysical observations have revealed an accelerated expansion of our universe that can be mimicked by the presence of a non-zero cosmological constant in Einstein's field equations, it is very well conceivable, that we need to consider a de Sitter special relativity. This frame could be called a DSR. However, there is a difference to the DSR approaches described previously: In those, it is assumed that the Lorentz symmetry is broken for very high energies, in de Sitter special relativity Lorentz symmetry is untouched and it is the translational part which is modified.
- VSR
  "Very Special Relativity" stands for an approach inspired by A.G. Cohen and S. Glashow [94], in which subgroups of **ISO(3,1)** are investigated, which are possibly conserved as symmetry groups after breaking the full Poincaré symmetry. If one assumes that energy-momentum conservation is strictly valid, one thus investigates subgroups of the Lorentz group. The algebra of the largest proper subgroup (called **SIM(2)**) is spanned by the generators $\{K_x + J_y, K_y - J_x, K_z, J_z\}$. The group **SIM(2)** can be characterized in terms of **SL(2,C)** matrices to which the Lorentz group is isomorphic: **SIM(2)** is represented by matrices

  $$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \quad \text{with} \quad (\mathrm{a,\ b,\ d}) \in \mathbb{C}\,, \quad \text{and} \quad \mathrm{ad} = 1.$$

  Indeed this subgroup of the Lorentz group is sufficient to explain all current experimental bounds. VSR invariance implies Lorentz invariance in the limit of CP conservation. It is able to make definite predictions for CP violations and neutrino masses.

For a thorough review of both theoretical frameworks and the experimental bounds on Lorentz invariance violation, see [365].

3.7 Concluding Remarks and Bibliographical Notes

In this chapter we saw that from faith in symmetries, one of the most important groups in fundamental physics, namely the Poincaré group, made its appearance. It is hard to believe that when Minkowski presented his 4-dimensional formulation of special relativity he met with incomprehension and even aroused opposition (a historic account on the notion of 4-vectors is [530].) Today, every student shows by an exercise that the "classical" Maxwell equations (3.1)–which were not even written by Maxwell in this form–can be encoded in the Lorentz-invariant form (3.30) and (3.31) by the identification as in (3.29). And some other students may be asked to demonstrate that these equations follow from the Lagrangian (3.35). But of course our forefathers did this in a different order. Karl Schwarzschild was the first writing down the action for Maxwell's theory (1903), and in 1909 Max Born gave it its form in the Minkowskian spacetime framework.

We arrived at the symmetry group for non-quantized flat-space physics by following more or less the historical course of discoveries according to Einstein and Minkowski, resulting in a prominent role of the vacuum velocity of light. Today, in retrospect, one can of course take sort of a birds-eye view and reconsider the essentials. Thus people started anew to find all kinematical groups [18] from symmetry arguments by moderate assumptions on space and time. This yields an interesting landscape, wherein the Poincaré and the Galilei group are only subspecies and successive contractions of the de Sitter groups. Given the current picture of an increasingly expanding universe, it may very well be the case that future generations of students first learn about the de Sitter and the Newton-Hooke groups, while the Poincaré and the Galilei groups will be considered to be nothing but historical aberrations.

Although historically special relativity was motivated by electrodynamics, Maxwell's theory can be founded in its own right, and specifically without recourse to a metric and to Poincaré transformations. As shown by F.W. Hehl and Y.N. Obukhov, electrodynamics (in this case in its macroscopic form) as well generalizations thereof can be derived from a set of axioms. I leave details of this approach to the literature, e.g. [266], and sketch the essential points here in terms of differential forms: Charge conservation and magnetic flux conservation (taken as Axiom 1 and Axiom 3) lead to the Maxwell equations $dH = J$ and $dF = 0$, where F is the field strength two-form (which in a (3+1)-foliation of spacetime exhibits the electric field strength E and the magnetic field strength B: $F = E \wedge dt + B$), $H = -\mathcal{H} \wedge dt + \mathcal{D}$ is the excitation two-form with the magnetization $\mathcal{H}$, and the displacement $\mathcal{D}$, and $J = -j \wedge dt + \rho$ is the electric current three-form with the electric current density j and the electric charge density ρ. Axiom 2 in [266] frames the well-established force law for an electric charge in an electromagnetic field, i.e. the force density relation $f_\mu = i_\mu F \wedge J$. In order to derive the energy-momentum law for electrodynamics from this Lorentz force density, f_μ needs to be expressed as an exact form. The authors find good reasons to rewrite $f_\mu = d\Sigma_\mu + X_\mu$ and postulate as Axiom 4 that the energy-momentum three-form Σ_μ of the electromagnetic field reads $\Sigma_\mu = \frac{1}{2}\Big[F \wedge i_\mu H - H \wedge i_\mu F\Big]$. Observe that none of the relations introduced so far make reference to a metric, therefore they are called to define "premetric electrodynamics". Since this is formulated in terms of differential forms, premetric electrodynamics is generally covariant per se. Now Axiom 5 relates the two-forms F and H by a linear and local "constituent relation" $H = \chi F$, i.e. in components $H_{\mu\nu} = \frac{1}{2}\chi_{\mu\nu}{}^{\rho\sigma} F_{\rho\sigma}$. The decomposition of the 6×6 χ-matrix into irreducible pieces results in the split of the previous relation as $H_{\mu\nu} = \frac{1}{2}\kappa_{\mu\nu}{}^{\rho\sigma} F_{\rho\sigma} + 2S^{\lambda}_{[\mu} F_{\nu]\lambda} + \alpha F_{\mu\nu}$ (here I'm using a slightly different notation than the authors). The first term $\kappa_{\mu\nu}{}^{\rho\sigma}$, called *metric-dilation*, has 20 independent components. The *skewon* S^{ν}_{μ} is traceless with 15 components and the *axion* α is a pseudo-scalar. Maxwell electrodynamics is recovered if the skewon and the axion field vanish, and if the metric-dilation field can be expressed in terms of the four-dimensional ϵ-symbol and a metric $g^{\mu\nu}$ as $\kappa_{\mu\nu}{}^{\rho\sigma} = \lambda\sqrt{-g}\epsilon_{\mu\nu\lambda\kappa} g^{\lambda\rho} g^{\kappa\sigma}$. Observe that only in this last step one needs the existence of the metric. And also observe that up to this point the axiomatically stated electrodynamics is richer than Maxwell's theory. Is there any physically meaningful and/or experimentally testable further axiom that uniquely leads to Maxwell's theory? By investigating the propagation of electromagnetic rays in this extended electrodynamics, Hehl and Obukhov find that this Axiom 6 is the quest for the non-existence of birefringence (also called double refraction) in vacuum. This quest technically corresponds to a specific case of the otherwise quartic Fresnel wave surface for electromagnetic waves, namely that this surface degenerates to the light cone of general relativity: "In this sense, the light-cone is a derived concept." Now it is known that the light-cone structure is not sufficient to establish a Riemannian geometry,[32] but that it relates to a Weyl geometry. From the point of symmetry groups it is interesting that the isometry group of the Weyl geometry is the Weyl group (sic!). If, as last step one sets a scale, that is postulates a time or length scale, one eventually arrives at a pseudo-Riemann geometry and at the Maxwell relation $H = \lambda_0 * F$, where λ_0 is to be identified with the admittance of free space and $(*)$ denotes the Hodge duality operation (see Appendix E.2.3).

There is another reason to consider Maxwellian electrodynamics as a toy theory for further considerations of symmetries in fundamental physics, since it exhibits another feature typical for Yang-Mills type theories: It is invariant with respect to local gauge transformations, exactly the feature Emmy Noether dealt with. We derived both Noether theorems, the first one concerned with conservation laws, the second one giving relations among the field equations (Noether identities). The theorems are notable in that they unite important topics in theoretical physics: variational principles, global symmetries and conservation laws, gauge symmetries and model building (for interacting fields).

Remarkably enough is that the work by Noether prompted F. Klein to the comment that the physicists' term "relativity" should more appropriately be called "invariance relative to a group". That is what this book is all about!

---

[32] This is – by general relativists – known by the names of the authors J. Ehlers, F. A. E. Pirani and A. Schild as the EPS axiomatics.

Chapter 4
Quantum Mechanics

Alle Quantenzahlen, mit Ausnahme der sog. Hauptquantenzahl, sind Kennzeichen von Gruppendarstellungen.

As we strongly believe today, there is no classical physics–or more properly–classical physics is only a certain limiting case of quantum physics. As a matter of fact, the theories of the fundamental interactions (except for gravity, alas) are quantum field theories. Thus we must pose and answer the questions concerning symmetries anew in the context of quantum physics. It is astounding that Noether's insights about symmetries become even simpler in quantum physics. This is essentially known since the work of Eugene Wigner, in which far-reaching relations between symmetry transformations, their unitary representations, and observables were derived. That work is essentially the topic of this chapter. Finally the Schrödinger equation is shown to follow largely from the projective representations of the Galilei group.

4.1 Principles of Quantum Mechanics

In this section I will give a very short and concise overview of the basic principles of quantum physics, like it was settled in the book of John von Neumann [525] "*Mathematische Grundlagen der Quantenmechanik*", which appeared in 1932, after essentially in the mid and late twenties of the last century quantum mechanics had gained its ultimate shape.

"All quantum numbers, with the exception of the so-called principal quantum number, are indices characterising representations of groups". (H. Weyl in the introduction of "*Gruppentheorie und Quantenmechanik*" [550].) As a matter of fact because of the SO(4) invariance of the Coulomb potential also the principal quantum number is related to a group representation: W. Pauli, Z. Phys. 36 (1926) 336–363.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_4,

4.1.1 Hilbert Space

The Hilbert space is so to say the arena of Quantum Mechanics.

$$
\begin{aligned}
\text{physical state} &\Longleftrightarrow \text{vector/ray in Hilbert space}\\
\text{observable} &\Longleftrightarrow \text{Hermitean operator}\\
\text{time development} &\Longleftrightarrow \text{unitary transformation.}
\end{aligned}
$$

A Hilbert space is a complex, complete, separable, unitary vector space $\mathcal{V}$. Leaving the technical subtleties requiring completeness and separability aside (details for instance in [302]), for our purposes it suffices to say that a Hilbert space is a vector space over the field of complex numbers (c-numbers) with the additional property of being unitary. If the vectors are denoted as suggested by to P.A.M. Dirac [124], by "kets" $|\psi\rangle \in \mathcal{V}$ and "bras" $\langle\phi|$ (the bras being vectors in a space dual to the ket space), a unitary vector space is defined as having a scalar product $\langle\phi|\psi\rangle \in \mathbb{C}$ with the properties

$$
\begin{aligned}
\langle\phi|\psi\rangle &= \langle\psi|\phi\rangle^*\\
\langle\phi|a\psi\rangle &= a\langle\phi|\psi\rangle, \quad a \in \mathbb{C}\\
\langle\phi|\psi+\chi\rangle &= \langle\phi|\psi\rangle + \langle\phi|\chi\rangle\\
\langle\phi|\phi\rangle \geq 0 \quad &\text{and} \quad \langle\phi|\phi\rangle = 0 \models |\phi\rangle = 0.
\end{aligned}
$$

This allows to define a norm $\|\phi\| := (\langle\phi|\phi\rangle)^{\frac{1}{2}}$ and a distance of two vectors as $d(\varphi,\psi) := \|\varphi-\psi\|$. Two vectors $|\phi\rangle$ and $|\psi\rangle$ are called orthogonal iff $\langle\phi|\psi\rangle = 0$. It is always possible to construct a basis of orthogonal and normalized vectors. In finite dimensional unitary vector spaces every vector can be expanded in an orthonormal basis $|e_k\rangle$ in the form

$$|\phi\rangle = a^k\,|e_k\rangle.$$

Every finite dimensional vector space is a Hilbert space. An example for an infinite-dimensional Hilbert space is the space L^2 of square-integrable functions for which $\langle\phi|\psi\rangle := \int \phi^*(x)\psi(x)dx$.

4.1.2 Operators

Operators[1] $\hat{L}$ map the vector space $\mathcal{V}$ onto itself:

$$|\psi\rangle \in \mathcal{V} \models \hat{L}|\psi\rangle \in \mathcal{V}.$$

[1] As prevailing in the (quantum) physics context, operators are denoted by a hat; in contexts in which the operator nature is obvious I refrain from the hat notation.

In quantum physics in most cases, one is interested in *linear operators*:

$$\hat{L}|\phi + \psi\rangle = \hat{L}|\phi\rangle + \hat{L}|\psi\rangle \qquad\qquad \hat{L}|a\varphi\rangle = a\hat{L}|\varphi\rangle.$$

In the context of symmetries we will encounter a prominent anti-linear operator, namely time-reversal. An *anti-linear operator* $\hat{A}$ complex conjugates any c-number on its right:

$$\hat{A}|a\varphi\rangle = a^*\hat{A}|\varphi\rangle.$$

Further notions are

- Both linear and anti-linear operators themselves constitute a vector space.
  Addition of operators:
  for $\hat{L}_i|\psi\rangle = |\psi_i\rangle$ define $(\hat{L}_i + \hat{L}_j)|\psi\rangle := |\psi_i\rangle + |\psi_j\rangle$.
  Multiplication with a scalar:
  for $\hat{L}|\psi\rangle = |\chi\rangle$ and $\hat{L}' = a\hat{L}$ define $\hat{L}'|\psi\rangle = a\hat{L}|\psi\rangle = a|\chi\rangle$.
- For very linear operator $\hat{L}$ its *adjoint* $\hat{L}^\dagger$ is defined implicitly by

  $$\langle\psi|\hat{L}^\dagger\phi\rangle \equiv \langle\hat{L}\psi|\phi\rangle.$$

  In contrast, the adjoint to an anti-linear operator is defined by

  $$\langle\psi|\hat{A}^\dagger\phi\rangle \equiv \langle\hat{A}\psi|\phi\rangle^*.$$

- An operator $\hat{H}$ is called *Hermitean*[2] if $\hat{H}^\dagger \equiv \hat{H}$.
- An operator $\hat{U}$ is called *unitary* iff

  $$\langle\hat{U}\psi|\hat{U}\varphi\rangle = \langle\psi|\varphi\rangle.$$

  An operator $\hat{T}$ is called *anti-unitary* iff

  $$\langle\hat{T}\psi|\hat{T}\varphi\rangle = \langle\psi|\varphi\rangle^*.$$

  Both unitary and anti-unitary transformations leave the square of the absolute value of the scalar product invariant: $|\langle\varphi|\psi\rangle|^2 = |\langle\hat{U}\psi|\hat{U}\varphi\rangle|^2 = |\langle\hat{T}\psi|\hat{T}\varphi\rangle|^2$. Both unitarity and anti-unitarity can be expressed in the form $\hat{U}^\dagger = \hat{U}^{-1}$, $\hat{T}^\dagger = \hat{T}^{-1}$. A unitary (anti-unitary) operator can be shown to be linear (anti-linear). An anti-unitary operator $\hat{T}$ can be written as $\hat{T} = \hat{U}\hat{K}$, where $\hat{U}$ is unitary and $\hat{K}$ is the operator which complex-conjugates every state on its right. Since $\hat{K}\hat{K} = I$ also $\hat{T}\hat{K} = \hat{U}$. Now compare $\hat{T}i|\varphi\rangle = \hat{U}\hat{K}i|\varphi\rangle = -i\hat{U}|\varphi\rangle^*$ with $i\hat{T}|\varphi\rangle = i\hat{U}|\varphi\rangle^*$, to find that

  $$(\hat{T}i + i\hat{T})|\varphi\rangle^* = 0. \tag{4.1}$$

---

[2] For the purpose of this book it is not essential that strictly speaking a distinction of a self-adjoint and a Hermitean operator has to be made. However, if it comes to the mathematical foundation of quantum mechanics it is mostly the self-adjointness property which counts.

- If $\hat{H}$ is Hermitean, $\hat{U}_H := \exp(i\hat{H})$ is unitary since

$$(\hat{U}_H)(\hat{U}_H^\dagger) = \exp(i(\hat{H} - \hat{H}^\dagger) = \hat{1}$$

(and *vice versa*). This fact is extraordinarily important since it allows to define symmetry for quantum physics in a strikingly simple manner, as shown below in Sect. 4.2.1.
- Eigenvalues and eigenvectors of operators
If $\hat{L}|\psi\rangle = l|\psi\rangle$ with $l \in \mathbb{C}$, $|\psi\rangle$ is called eigenvector of $\hat{L}$ with eigenvalue l. It is easy to prove that (1) eigenvalues of Hermitean operators are real, (2) eigenvectors for different eigenvalues are orthogonal. Eigenvectors of a Hermitean operator span a complete basis. In the finite-dimensional case

$$\hat{A}|\psi_n\rangle = a_n|\psi_n\rangle,$$

and every other vector $|\phi\rangle$ can be expressed in this basis in the form: $|\phi\rangle = \sum c_k|\psi_k\rangle$.
In case $a_n \equiv a_m$ for $|\psi_n\rangle \neq |\psi_m\rangle$, one is dealing with degeneracy.
- Matrix expressions of operators
The action of an operator on a vector belonging to an orthonormal basis

$$\hat{L}|e_k\rangle = \sum_j L_{kj}|e_j\rangle$$

can be expressed by $(N \times N)$-matrices $L_{kj} \in \mathbb{C}$, where $L_{kj} = \langle \hat{L}e_k|e_j\rangle$. The adjoint is then given by $L^\dagger_{kj} = L^*_{jk}$. Performing transformations one after the other amounts to matrix multiplication.

4.1.3 States, Observables, and Measurements

In this subsection the terms in the mathematics of Hilbert space are translated to quantum-mechanical entities. I mainly adopt Dirac's bra-ket notation for abstract states, and only in some specific cases refer to notions such as wave functions.

States

Physical states are "rays" in a Hilbert space. A ray $||\phi\rangle$ is a set $\{|\phi\rangle\}$ of normed vectors ($\langle\phi|\phi\rangle = 1$). Two vectors $|\phi\rangle$ and $|\tilde{\phi}\rangle$ belong to the same ray iff $|\tilde{\phi}\rangle = \xi|\phi\rangle$, where ξ is a complex number with $|\xi| = 1$. In other words: $|\phi\rangle$ and $|\tilde{\phi}\rangle = e^{i\alpha}|\phi\rangle$ represent one and the same physical state. In the following, I do not always painstakingly distinguish between states and vectors in a Hilbert space, because it should become clear from the context what is meant.

If $|\phi\rangle$ and $|\psi\rangle$ are (representatives of) states, also any linear combination $a|\phi\rangle + b|\psi\rangle$ with $a, b \in \mathbb{C}$ is a state. This *superposition principle* in a certain sense renders quantum physics "simple". However, a closer look leads to a refinement of interrelating physically realizable states with a ray in a Hilbert state. The superposition principle would allow for instance the realization of states with different charges, or with integer and half-integer spin, with different baryon numbers, and even with different masses. This can be avoided only by splitting the full Hilbert space into subspaces, and by defining appropriate rays in these subspaces. Then the previous examples of outlandish superpositions are inhibited by so-called superselection rules; more about this below.

Observables and Measurement

Observables are defined to represent measurable quantities and are identified with Hermitean operators. This identification needs a further qualification in case of the decomposition of the Hilbert space mentioned previously in the context of superselection rules, where not all Hermitean operators represent measurable quantities.

The interpretation of measuring the observable $\hat{A}$ in the state $||\psi\rangle$ is as follows: Let $|\psi\rangle$ be a representative of the state. This vector can be expanded in terms of eigenvectors $|a_n\rangle$ of the operator $\hat{A}$:

$$|\psi\rangle = \sum c_n |a_n\rangle, \quad \text{where} \quad \hat{A}|a_n\rangle = \alpha_n |a_n\rangle$$

or, if the Hilbert space is infinite,

$$|\psi\rangle = \int da\, c(a)|a\rangle, \quad \text{where} \quad \hat{A}|a\rangle = a|a\rangle.$$

The probability to measure the value a for the observable $\hat{A}$ is then given by

$$P(a) = \sum_{\{j\}} |c_j|^2 = \sum_{\{j\}} |\langle a_j|\psi\rangle|^2,$$

where $\{j\}$ runs over all states for which $a = \alpha_n$. Thus in the non-degenerate case, $P(\alpha_n) = |c_n|^2$. After the measurement of $a = \alpha_n$, the system is definitely in the state $|\underline{\psi}\rangle = \sum_{\{j\}} c_j |a_j\rangle$, i.e. in case of non-degeneracy $|\underline{\psi}\rangle = |\alpha_n\rangle$. As a matter of fact, these definitions are independent of which representation vector for the state is chosen: For another representative $|\tilde{\psi}\rangle$ holds $|\tilde{\psi}\rangle = \sum \tilde{c}_n |a_n\rangle$ and $P(a) = \sum |\langle a_j|\tilde{\psi}\rangle|^2 = |\langle a_j|\psi\rangle|^2$. If a system is in the state $||\psi\rangle$ the probability to find the observable $\hat{A}$ in a state $||a_k\rangle$ is given by

$$P\big([||\psi\rangle] \to [||a_k\rangle]\big) = |\langle\psi|a_k\rangle|^2.$$

Again, this probability does not depend on which representative one takes from the ray.

Two observables $\hat{A}$ and $\hat{B}$ can be measured simultaneously iff

$$\hat{A}\hat{B}|\psi\rangle - \hat{B}\hat{A}|\psi\rangle =: [\hat{A}, \hat{B}]|\psi\rangle = 0.$$

From this follows that $\hat{A}$ and $\hat{B}$ do have the same eigenstates, and that these states can be written as $|a, b, \sigma\rangle$ with the properties

$$\hat{A}|a, b, \sigma\rangle = a|a, b, \sigma\rangle \quad \hat{B}|a, b, \sigma\rangle = b|a, b, \sigma\rangle.$$

Here σ stands for further parameters (quantum numbers) characterizing the overall state.

The coefficients c_n above, or in the infinite case the function $c(a)$, are determined by the following brief calculation: First consider

$$\langle a|\psi\rangle = \int da' \, \langle a|c(a')|a'\rangle.$$

Then assume the normalization $\langle a|a'\rangle = \alpha\,\delta(a - a')$ with a constant α, from which

$$\langle a|\psi\rangle = \alpha c(a) \qquad\qquad |\psi\rangle = \frac{1}{\alpha}\int da \, |a\rangle\langle a|\psi\rangle.$$

This gives rise to the completeness relation

$$1 = \alpha^{-1} \int da|a\rangle\langle a|, \tag{4.2}$$

for obvious reasons also called “resolution of unity”.

4.1.4 Time Evolution

The dynamical evolution of a state follows the Schrödinger equation

$$i\hbar\frac{d}{dt}|\psi\rangle = \hat{H}|\psi\rangle, \tag{4.3}$$

where $\hat{H}$ is the Hamilton operator (or Hamiltonian), which is Hermitean with real eigenvalues:

$$\hat{H}|\psi_n\rangle = E_n|\psi_n\rangle.$$

It is easily verified that the formal solution of the Schrödinger equation is given by

$$|\psi(t)\rangle = \exp\left\{-\frac{i}{\hbar}\hat{H}t\right\}|\psi(0)\rangle.$$

In defining the unitary operator

$$\hat{U}_t := \exp\left\{-\frac{i}{\hbar}\hat{H}t\right\}$$

we see that the evolution of a state from t_1 to t_2 proceeds as:

$$|\psi(t_2)\rangle = \hat{U}_{(t_2-t_1)}|\psi(t_1)\rangle.$$

The operator $\hat{U}_t$ also mediates between the Schrödinger picture and the Heisenberg picture: Let $\hat{U}_t|\psi_0\rangle = |\psi(t)\rangle$ and define for every operator $\hat{L}$ the associated operator $\hat{L}_H(t) := \hat{U}_t\hat{L}\hat{U}_t^{-1}$. While in the Schrödinger picture the states are time-dependent, in the Heisenberg picture the operators are-time dependent. For the time evolution of the operators one finds

$$i\hbar\frac{d\hat{L}_H(t)}{dt} = [\hat{L}_H(t), \hat{H}].$$

The Schrödinger equation entails the equation of motion for the expectation value of an operator $\hat{L}$ with respect to a state $|\psi\rangle$, defined as $\langle|\hat{L}|\rangle_\psi = \langle\psi|\hat{L}|\psi\rangle$. We derive

$$\frac{d}{dt}\langle\hat{L}\rangle_\psi = \left(\frac{d}{dt}\langle\psi|\right)\hat{L}|\psi\rangle + \langle\psi|\left(\frac{\partial\hat{L}}{\partial t}\right)|\psi\rangle + \langle\psi|\hat{L}\left(\frac{d}{dt}|\psi\rangle\right).$$

In using the Schrödinger equation (4.3), this becomes

$$\begin{aligned}\frac{d}{dt}\langle\hat{L}\rangle_\psi &= -\frac{1}{i\hbar}\langle\psi|\hat{H}\hat{L}|\psi\rangle + \left\langle\left(\frac{\partial\hat{L}}{\partial t}\right)\right\rangle_\psi + \frac{1}{i\hbar}\langle\psi|\hat{L}\hat{H}|\psi\rangle \\ &= \frac{1}{i\hbar}\langle[\hat{L}, \hat{H}]\rangle_\psi + \left\langle\left(\frac{\partial\hat{L}}{\partial t}\right)\right\rangle_\psi\end{aligned}$$

or

$$\frac{d}{dt}\langle\psi|\hat{L}|\psi\rangle = \frac{1}{i\hbar}\langle\psi|[\hat{L}, \hat{H}]|\psi\rangle + \langle\psi|\left(\frac{\partial\hat{L}}{\partial t}\right)|\psi\rangle. \tag{4.4}$$

Because of the ample structural similarity with the classical Hamilton equations of motion (2.15), there seems to exist a "canonical" quantization procedure which amounts to replacing classical phase space objects F, G by operators $\hat{F}, \hat{G}$ in a Hilbert space, and to using the substitution/correspondence rules

$$\{F, H\} \longrightarrow \frac{1}{i\hbar}[\hat{F}, \hat{H}] \qquad [\hat{F}, \hat{G}] = i\hbar\widehat{\{F, G\}}. \tag{4.5}$$

For the classical coordinates x^j and momenta p^k this amounts to postulating the canonical commutation relations

$$[\hat{x}^j, \hat{p}^k] = i\hbar\,\delta^{jk} \tag{4.6}$$

which one may add as an additional requirement to the foundation of quantum mechanics. Observe that this is to be read as a condition on operators in a Hilbert space. It can be fulfilled for instance in configuration space by realizing the momenta as $\hat{p}^k = -i\hbar(\partial/\partial x^k)$, and this is what was done in the very early days of quantum mechanics and what one learns in elementary courses on the subject. But (4.6) can also be realized in the momentum space as $\hat{x}^j = i\hbar(\partial/\partial p^j)$. Problems arise if instead of Cartesian coordinates one switches for instance to spherical coordinates. Given this, modern approaches to the foundation of quantum physics replace canonical commutation relations by more abstract structures such as Heisenberg algebras and Weyl systems; see e.g. [167]. The recipe (4.5)–sometimes jokingly called "Dirac quantization by hatting"–also presupposes that every observable can be written in terms of $\hat{x}$ and $\hat{p}$. Another pitfall is that for a phase space function $F(x, p)$ its quantum substitute $\hat{F}(\hat{x}, \hat{p})$ is not uniquely defined. Take for instance $F(x, p) = xp$ which allows the substitution $\hat{x}\hat{p}$, but also $\hat{p}\hat{x}$, or any linear combination thereof (this is the ordering problem of quantum physics).

In order to later understand the idea of a scattering matrix and of transition amplitudes in field theory, let me mention also the Dirac picture (or interaction representation) of quantum mechanics. Here one splits the overall Hamiltonian into a free and an interaction part as $\hat{H} = \hat{H}_F + \hat{H}_I$ and defines states $|\psi\rangle_I$ and operators $\hat{L}_I$ by the unitary transformations

$$|\psi(t)\rangle_I = e^{(i/\hbar)\hat{H}_F t}|\psi(t)\rangle, \qquad \hat{L}_I(t) = e^{(i/\hbar)\hat{H}_F t}\hat{L}(t)e^{-(i/\hbar)\hat{H}_F t}$$

where $|\psi(t)\rangle$ and $\hat{L}$ are the states and operators in the Schrödinger representation. With the aid of the Schrödinger equation, the time evolution of states and operators in the interaction picture becomes

$$i\hbar\frac{d}{dt}|\psi(t)\rangle_I = \hat{H}_I(t)|\psi(t)\rangle_I, \qquad i\hbar\frac{d\hat{L}_I(t)}{dt} = [\hat{L}_I(t), \hat{H}_F]. \tag{4.7}$$

In the Dirac picture, therefore, the time evolution of states is determined by the interaction Hamiltonian, and the time evolution of operators by the free Hamiltonian.

As you can understand from the previous remark and other comments made in passing, my presentation of the core of quantum theory is traditional and in the spirit of many textbooks. It is presumably outdated if one is directly interested in quantum field theory. There, the foundation of quantum physics in terms of path integrals is more appropriate; for this reason, I have included Appendix D.

4.2 Symmetry Transformations in Quantum Mechanics

As been worked out in previous chapters, in classical physics symmetries are related to conserved quantities. How can these be characterized in quantum mechanics? If at all, only eigenvalues of observables can qualify as conserved quantities. Thus let

$\hat{A}|a\rangle = a|a\rangle$, or $a = \langle a|\hat{A}|a\rangle$, and applying (4.4),

$$\dot{a} = \frac{1}{i\hbar}\langle a|[\hat{A}, \hat{H}]|a\rangle + \left\langle a|\frac{\partial \hat{A}}{\partial t}|a\right\rangle$$

we conclude

- If all eigen-values of the operator $\hat{A}$ are conserved quantities ($\dot{a} = 0$), holds

$$i\hbar\frac{\partial \hat{A}}{\partial t} = [\hat{H}, \hat{A}].$$

- If $\hat{A}$ does not depend on time explicitly, and if $\hat{A}$ commutes with the Hamilton operator $\hat{H}$, all eigen-vectors of $\hat{A}$ are conserved quantities.

4.2.1 Wigner Theorem

Given are two observers $\mathcal{O}$ and $\mathcal{O}'$, who describe a system in states $||\phi\rangle$ and $||\phi'\rangle$, respectively. Evidently a symmetry is present if both observers find the same probabilities:

$$P\big(||\phi\rangle \to ||\phi_k\rangle\big) = P\big(||\phi'\rangle \to ||\phi'_k\rangle\big).$$

Consequently, for arbitrary states we require in the case of a symmetry

$$|\langle\phi'|\psi'\rangle|^2 = |\langle\phi|\psi\rangle|^2. \tag{4.8}$$

E. P. Wigner demonstrated in 1931 [555], that this implies that for every symmetry transformation S, there must exist a symmetry operator $\hat{U}_S$ which transforms a state $|\varphi\rangle \in ||\varphi\rangle$ into a state $\hat{U}_S|\varphi\rangle \in ||\varphi'\rangle$, and which is either unitary and linear

$$\langle\hat{U}_S\phi|\hat{U}_S\psi\rangle = \langle\phi|\psi\rangle \qquad \hat{U}_S\,(\xi|\phi\rangle + \eta|\psi\rangle) = \xi\,\hat{U}_S|\phi\rangle + \eta\,\hat{U}_S|\psi\rangle,$$

or anti-unitary and anti-linear:

$$\langle\hat{U}_S\phi|\hat{U}_S\psi\rangle = \langle\phi|\psi\rangle^* \qquad \hat{U}_S\,(\xi|\phi\rangle + \eta|\psi\rangle) = \xi^*\,\hat{U}_S|\phi\rangle + \eta^*\,\hat{U}_S|\psi\rangle.$$

It is obvious that if $\hat{U}_S$ has these properties, the condition (4.8) is fulfilled. The proof that these conditions are also necessary was given in [555], and has been presented in various versions thereafter. The version by V. Bargmann [28] is close to the original one but more understandable. S. Weinberg closes a gap in Wigner's original proof in Chap. 2 of [536].

Among the symmetry transformations, we also may consider the trivial one to which of course the identity operator $\hat{U} = I$ is associated, and which obviously is linear and unitary. All transformations which are continuously connected to the identity (for instance translations and rotations) are thus also characterized by linear unitary symmetry operators. As a matter of fact, an anti-linear and anti-unitary operator only arises in the case of time reversal. This will be demonstrated in the subsection on discrete symmetries in the next chapter.

Each symmetry operator $\hat{U}_S$ is only fixed up to a phase: if $\hat{U}_S$ obeys the Wigner conditions, also $\hat{U}'_S = e^{i\gamma}\,\hat{U}_S$ fulfills the conditions.

The group property of symmetry transformations S (as transformations of states, i.e. rays in a Hilbert space) is handed over to the associated symmetry operators $\hat{U}_S$, however again up to a phase factor. Let first act S on the ray $||\psi\rangle$, that is S: $||\psi\rangle \to ||\psi'\rangle$ with $\hat{U}_S|\psi\rangle \in ||\psi'\rangle$. Then, let S' act on this state so that $\hat{U}_{S'} \cdot \hat{U}_S|\psi\rangle \in ||\psi''\rangle$. Now $\hat{U}_{S'\circ S}|\psi\rangle$ is also in the ray $||\psi''\rangle$, and the vectors in this ray can differ at most by a phase:

$$\hat{U}_{S'}\hat{U}_S|\psi\rangle = e^{i\phi(S',S,\psi)}\,\hat{U}_{S'\circ S}|\psi\rangle,$$

where, as indicated, the phase ϕ may depend on the state $|\psi\rangle$. To see that this is not the case, take two linearly-independent vectors $|\psi_A\rangle$ and $|\psi_B\rangle$ and consider the state $|\psi_{AB}\rangle = |\psi_A\rangle + |\psi_B\rangle$. Now

$$\begin{aligned} e^{i\phi(S',S,\psi_{AB})}\,\hat{U}_{S'\circ S}(|\psi_A\rangle + |\psi_B\rangle) &= \hat{U}_{S'} \cdot \hat{U}_S|\psi_A\rangle + \hat{U}_{S'} \cdot \hat{U}_S|\psi_B\rangle \\ &= e^{i\phi(S',S,\psi_A)}\,\hat{U}_{S'\circ S}|\psi_A\rangle + e^{i\phi(S',S,\psi_B)}\,\hat{U}_{S'\circ S}|\psi_B\rangle. \end{aligned}$$

Multiplying this relation by the inverse of $\hat{U}_{S'\circ S}$, one gets

$$e^{\pm i\phi(S',S,\psi_{AB})}(|\psi_A\rangle + |\psi_B\rangle) = e^{\pm i\phi(S',S,\psi_A)}|\psi_A\rangle + e^{\pm i\phi(S',S,\psi_B)}|\psi_B\rangle$$

where the $(\pm)$ applies to a unitary and an anti-unitary operator $\hat{U}$, respectively. Since the two states $|\psi_A\rangle$ and $|\psi_B\rangle$ were assumed to be linearly-independent, the phase ϕ is independent of the states, and the previous relation can be written as an operator relation:

$$\hat{U}_{S'} \cdot \hat{U}_S = e^{i\phi(S',S)}\,\hat{U}_{S'\circ S}. \tag{4.9}$$

For $\phi \equiv 0$, the symmetry operators constitute a representation of the symmetry group. This is a favorable case, because then all results known from group representation theory can be applied directly. For non-vanishing phases $\phi(S', S)$, one is dealing in general with a projective (or "ray") representation. These are discussed in Appendix A.3.5.

Strictly speaking, Wigner's rule in the form of (4.9) holds only for an irreducible Hilbert space (which is tacitly assumed in the foundation of quantum mechanics by von Neumann). Otherwise the phases may very well depend on the state, namely by their link to one of the components in the decomposition of the full Hilbert space. Again this relates to superselection rules.

4.2.2 Symmetry Transformations and Observables

Symmetrie operators $\hat{U}_S$ are unitary (except the one associated with time reversal, which is excluded from the subsequent considerations), observables $\hat{A}$ are Hermitean. Is it possible to find observables which belong to symmetry operators?

If the symmetry transformation is discrete, like for instance space inversion P with $\hat{U}_P \hat{U}_P = I$, one can directly choose $\hat{A}_P = \hat{U}_P$; because of $\hat{U}_P = \hat{U}_P^{-1} = \hat{U}_P^{\dagger}$, the unitary operator is already Hermitean.

For continuous symmetry transformations $S(\sigma)$ with parameter σ let us denote $\hat{U}_{S(\sigma)}$ as $\hat{U}_\sigma$. We can associate to this unitary operator a Hermitean operator $\hat{A}_\sigma$ defined by

$$\hat{U}_\sigma = e^{i\sigma \hat{A}_\sigma} \tag{4.10a}$$

$$\hat{A}_\sigma = -i \left(\frac{\partial}{\partial \sigma} \hat{U}_\sigma \right)_{|\sigma=0} . \tag{4.10b}$$

With this step we have established a direct connection between symmetry transformations $S(\sigma)$ and observables $\hat{A}_\sigma$. If we take the eigenvectors of these observables as a base, the Hilbert space immediately reflects these symmetries.

It should be remarked that the relations (4.10), although apparently obvious, are mathematically nontrivial. A theorem by M. H. Stone "On one-parameter unitary groups in Hilbert Space" (with a refinement by J. von Neumann) defines under which circumstances (4.10a) can be taken for granted; see e.g. [291]. In any case, observe the similarity of these expressions to those which hold between Lie group elements and their algebra generators according to (A.2, A.1).

4.2.3 "Noether Theorem of Quantum Mechanics"

Quotation marks are used in this title since Emmy Noether formulated her theorems only in the context of classical physics. In quantum physics there is a close interplay among symmetries, observables, and conserved quantities. In order to discover these interrelations, we next consider the dynamical evolution of the unitary symmetry operators and their Hermitean descendants associated to a symmetry transformation.

1. dynamical evolution of $\hat{U}_S$
   Let $|\psi_S\rangle = \hat{U}_S|\psi\rangle$. The state $|\psi_S\rangle$ evolves in time according to the Schrödinger equation
   $$i\hbar \frac{d}{dt}(\hat{U}_S|\psi\rangle) = \hat{H}(\hat{U}_S|\psi\rangle)$$
   or, term by term:
   $$i\hbar \frac{d\hat{U}_S}{dt}|\psi\rangle + i\hbar \hat{U}_S|\frac{d\psi}{dt}\rangle = [\hat{H}, \hat{U}_S]|\psi\rangle + \hat{U}_S \hat{H}|\psi\rangle.$$

Since $|\psi\rangle$ fulfills the Schrödinger equation itself, we obtain

$$i\hbar\frac{d\hat{U}_S}{dt}|\psi\rangle = [\hat{H}, \hat{U}_S]|\psi\rangle.$$

If $\hat{U}_S$ does not depend on time explicitly we thus derive

$$[\hat{H}, \hat{U}_S] = 0. \tag{4.11}$$

2. dynamical evolution of $\hat{A}$ or $\hat{A}_\sigma$
   If $\hat{U}_S$ is discrete (with the prominent example of space inversions) the relation (4.11) directly applies to $\hat{A}_S$ as $[\hat{H}, \hat{A}_S] = 0$.
   For a continuous symmetry transformation $S(\sigma)$ we find from (4.10b)

$$[\hat{A}_\sigma, \hat{H}] = \left[-i\left(\frac{\partial}{\partial\sigma}\hat{U}_\sigma\right)_{|\sigma=0}, \hat{H}\right] = -i\frac{\partial}{\partial\sigma}[\hat{U}_\sigma, \hat{H}]_{|\sigma=0},$$

   and since $[\hat{U}_\sigma, \hat{H}] = 0$,

$$[H, \hat{A}_\sigma] = 0.$$

In conclusion the so called Noether theorem of quantum mechanics can be stated as

(1) To every time-independent symmetry there is associated a Hermitean operator which commutes with the Hamiltonian;
(2) Each (time-independent) symmetry transformation has an associated observable which is a conserved quantity.

Astoundingly, symmetry is in quantum physics much easier to formulate than in the classical case. This is essentially because of the probability interpretation and the linearity in the quantum state space.

4.2.4 Symmetries and Superselection Rules

The notion of superselection was coined by G.C. Wick, A.S. Wightman, and E.P. Wigner in their attempt to clarify the meaning of the intrinsic parity of elementary particles [553]. Contrary to the postulate by von Neumann to identify the set of all physical observables with the set of all self-adjoint operators on a Hilbert space $\mathcal{H}$, they consider the possibility that not every self-adjoint operator can serve as an operator associated to an observable. Specifically, the operators do not act irreducibly on the full Hilbert space. Thus $\mathcal{H}$ can be written as a direct sum $\mathcal{H} = \oplus_i \mathcal{H}_i$; the $\mathcal{H}_i$ are called coherent subspaces or superselection sectors. Accordingly it makes sense to consider only operators acting within each of the $\mathcal{H}_i$ and with states/ray representations in each of these superselection sectors. In each sector, the superposition principle is valid, but a linear combination, $\alpha|\psi_i\rangle + \beta|\psi_j\rangle$ of states belonging to two different coherent subspaces is not physically realizable. There is a superselection rule that forbids transitions from one sector to the other.

Of course, one feels uncomfortable if these superselection rules are simply imposed on quantum physics. They should somehow be inherent to the theory. One mechanism is tied to unitary projective representations of a symmetry group: As mentioned before, Wigner's rule (4.9) is valid only within each of the superselection sectors.

Consider two Hilbert spaces $\mathcal{H}'$ and $\mathcal{H}''$, together with ray representations U' and U'' of a symmetry group G. Each of the representations is defined up to equivalence as stated in (A.38). Now take the Hilbert space $\mathcal{H} = \mathcal{H}' \oplus \mathcal{H}''$, and ask under what conditions $U = U' \oplus U''$ is a ray representation of G on $\mathcal{H}$. Thus calculate

$$U_{g_1} U_{g_2} = (U'_{g_1} \oplus U''_{g_1})(U'_{g_2} \oplus U''_{g_2}) = e^{i\phi'(g_1,g_2)} U'_{g_1 g_2} \oplus e^{i\phi''(g_1,g_2)} U''_{g_1 g_2}.$$

This can be written as

$$U_{g_1} U_{g_2} = e^{i\phi(g_1,g_2)} U_{g_1 g_2}$$

only iff the phase factors can be made to coincide–indeed it suffices if they are equivalent in the sense of (A.38). Otherwise we are not allowed to superimpose states from the two Hilbert spaces.

A special case is found of course if the symmetry group G is such that all phases can be reabsorbed in redefining the representations by appropriate phases. According to Bargmann [27], this is possible if the group is simply-connected and if all central charges of the Lie algebra associated to the symmetry group can be transformed away. If this is not the case, one may consider a larger group $\bar{G}$ instead of G. If this is labeled by pairs (g, ρ) with elements of G and $\rho \in \mathbb{R}$ and if the multiplication is defined as

$$(g_1, \rho_1)(g_2, \rho_2) = (g_1 g_2, \rho_1 + \rho_2 + \phi(g_1, g_2)),$$

then indeed $\tilde{U}_{(g,\rho)} := \exp(i\rho)U_g$ constitutes an "ordinary" representation. For more group-theoretical considerations of $\bar{G}$ see [219]. Perhaps formal tricks like this led Weinberg to call the issue of superselection rules "a bit of a red herring". He writes: "… it may or it may not be possible to prepare physical systems in arbitrary superpositions of states, but one cannot settle the question by reference to symmetry principles, because whatever one thinks the symmetry group of nature may be, there is always another group whose consequences are identical except for the absence of superselection rules." (p. 90 in Vol. I of [536]).

The character and origin of superselection rules became a significant topic in axiomatic quantum field theory. A comprehensive short overview is given in [223]. For a "historical review of the development of the notion of superselection rule starting from the recognition in 1952 of the charge and the univalence[3] superselection rules" and "applications to environmentally induced superselection rules in the last decade" see [554]. Environmentally induced superselection is more extensively covered in [225]; it is better known under the name 'decoherence'.

---

[3] This refers to "fermion-boson superselection":

4.3 Quantum Physics and Group Representation

4.3.1 Why Group Representation?

We saw already before–and will find it illustrated it in this section on examples–that (irreducible) representations of the symmetry group have a direct bearing on quantum physics.

- Symmetry transformations form a group. Due to the Wigner theorem the transformations are mapped onto unitary operators. According to (4.9), these constitute a projective representation of the symmetry group.
- The states of the system transform into each other according to a representation of the symmetry group.
- Every system can be described as a superposition of those states that transform according to the irreducible representations of its symmetry group.
- The eigenvalue spectrum of the symmetry invariants classifies the irreducible representations.

4.3.2 Galilei Operators

In this subsection, we determine the unitary operators that are associated to the symmetries with respect to the Galilei group.

Space Translations

Consider a translation in x-direction: $S(a)x := x + a$. In case of invariance of a system under this transformation there is according to (4.10a) a unitary operator $\hat{U}_a = \exp[ia\hat{P}_x]$ with a symmetry observable $\hat{P}_x$. Define $|\alpha'\rangle := \hat{U}_a|\alpha\rangle$. The configuration-space wavefunction of a state $|\alpha\rangle$ is given by $\Psi_\alpha(x) := \langle x|\alpha\rangle$, and thus

$$\Psi_{\alpha'}(x) = \langle x|\alpha'\rangle = \langle x|\hat{U}_a|\alpha\rangle = \langle \hat{U}_a^\dagger x|\alpha\rangle.$$

Since $U_a^\dagger = U_a^{-1} = U_{(-a)}$ we find

$$\Psi_{\alpha'}(x) = \langle x - a|\alpha\rangle \equiv \Psi_\alpha(x - a),$$

or in a Taylor expansion of $\Psi_\alpha(x - a)$

$$\Psi_{\alpha'}(x) = \Psi_\alpha(x) - a\frac{d}{dx}\Psi_\alpha + \frac{1}{2}a^2\frac{d^2}{dx^2}\Psi_\alpha + \ldots \equiv \exp\left(-a\frac{d}{dx}\right)\Psi_\alpha(x). \quad (4.12)$$

On the other hand

$$\Psi_{\alpha'}(x) = \langle x|U_a\alpha\rangle = \langle x|e^{ia\hat{P}_x}\alpha\rangle$$

so that the comparison with (4.12) reveals, that (in configuration space) $\hat{P}_x = id/dx$. This is, up to a factor $\hbar$, identical to the x-component of the quantum-mechanical momentum operator. Since translations commute, it is evident that for a generic space translation with $\vec{a}$ the unitary symmetry operator is

$$\hat{U}_{\vec{a}} = \exp[-i\vec{a}\cdot\hat{\vec{P}}]. \tag{4.13}$$

It is instructive to derive this connection between space translations and momentum observables in another way. Denote the eigenstates of the operator $\hat{\vec{P}}$ as $|\vec{p},\lambda\rangle$, where λ stands for a possible degeneracy:

$$\hat{\vec{P}}|\vec{p},\lambda\rangle = \vec{p}|\vec{p},\lambda\rangle.$$

(This has of course to be read component-wise: $\hat{P}^k|\vec{p},\lambda\rangle = p^k|\vec{p},\lambda\rangle$). Each state in the infinite-dimensional Hilbert space can be represented as a linear combination in this basis:

$$|\alpha\rangle = \int d^3p\, f(\vec{p})|\vec{p},\lambda\rangle.$$

The operation of $\hat{U}_{\vec{a}}$ (as defined by (4.13)) on this state yields

$$\begin{aligned}\hat{U}_{\vec{a}}|\alpha\rangle &= \int d^3p\, f(\vec{p})\hat{U}_{\vec{a}}|\vec{p},\lambda\rangle\\ &= \int d^3p\, f(\vec{p})\exp[-i\vec{a}\cdot\vec{p}]|\vec{p},\lambda\rangle := \int d^3p\,[\hat{U}_{\vec{a}}f](\vec{p})|\vec{p},\lambda\rangle.\end{aligned}$$

Now consider the Fourier transformation of $f(p)$

$$\psi(\vec{x}) = N^3\int d^3p\, f(\vec{p})e^{i\vec{p}\vec{x}} \qquad \text{with} \qquad N = (2\pi)^{-\frac{1}{2}}.$$

Applying the symmetry operation–defined as the Fourier transformation of $[\hat{U}_{\vec{a}}f](\vec{p})$ - we obtain

$$[\hat{U}_{\vec{a}}\psi](\vec{x}) := N^3\int d^3p\,[\hat{U}_{\vec{a}}f](\vec{p})e^{i\vec{p}\cdot\vec{x}} = N^3\int d^3\ p\, f(\vec{p})e^{-i\vec{a}\vec{p}}e^{i\vec{p}\vec{x}} \equiv \psi(\vec{x}-\vec{a}).$$

This relation can be generalized to arbitrary symmetry transformations of the spatial coordinates and of time:

$$(\vec{x}',t') = S_a(\vec{x},t) \Leftrightarrow [\hat{U}_a\psi](\vec{x},t) = \psi(S_a^{-1}(\vec{x},t)). \tag{4.14}$$

If one directly applies the momentum operator on $\psi(x)$, one obtains

$$[\hat{\vec{P}}\psi](\vec{x}) = i\frac{d}{d\vec{a}}[\hat{U}_{\vec{a}}\psi](\vec{x})_{|\vec{a}=0} = i\frac{d}{d\vec{a}}\psi(\vec{x}-\vec{a})_{|\vec{a}=0} = -i\frac{d}{d\vec{x}}\psi(\vec{x}),$$

which again is the explicit realization of the momentum operator in the configuration space.

At this place let me define orthonormalizations for the position and momentum eigenstates, as they are helpful for instance in establishing the path integral formulation of quantum physics, described in Appendix D. Write the completeness relation (4.2) for the position and momentum eigenvalues

$$1 = \int dx|x\rangle\langle x| \qquad\qquad 1 = \frac{1}{2\pi\hbar}\int dp|p\rangle\langle p|.$$

Then one verifies

$$\langle x|x'\rangle = \delta(x-x') \qquad \langle x|p\rangle = e^{-ipx/\hbar} \qquad \langle p|p'\rangle = \frac{1}{2\pi\hbar}\delta(p-p') \tag{4.15}$$

for instance by

$$\begin{aligned}\langle x|x'\rangle &= \frac{1}{2\pi\hbar}\langle x|\int dp\,|p\rangle\langle p|x'\rangle = \frac{1}{2\pi\hbar}\int dp\, e^{-ipx/\hbar}e^{ipx'/\hbar}\\ &= \frac{1}{2\pi\hbar}(2\pi)\int\frac{dp}{2\pi}\,e^{-i(px-px')/\hbar} = \delta(x-x').\end{aligned}$$

Time Translations

Although an explicit time dependence was excluded so far, the previous considerations can be directly adopted: A time translation $S(\tau) = t + \tau$ has an associated symmetry operator

$$\hat{U}_\tau = \exp[i\tau\hat{H}_\tau]$$

with an Hermitean operator $\hat{H}_\tau$. Let the eigenvalue equation of $\hat{H}_\tau$ be

$$\hat{H}_\tau|E,\varrho\rangle = E|E,\varrho\rangle.$$

An arbitrary state $|\alpha\rangle$ is expressed in this basis as

$$|\alpha\rangle = \int dE\, g(E)|E,\varrho\rangle.$$

The Fourier transformation of $g(E)$

$$\varphi(t) = N \int dE\, g(E) e^{iEt}$$

obeys (4.14) in the form of

$$[\hat{U}_\tau \varphi](t) = \varphi(S_\tau^{-1}) = \varphi(t - \tau).$$

Finally, the application of the operator $\hat{H}_\tau$ to φ yields

$$[\hat{H}\varphi](t) = -i \frac{d}{d\tau} [U_\tau \varphi](t)_{|_{\tau=0}} = i \frac{d}{dt} \varphi.$$

This is–up to a factor $\hbar$–the Schrödinger equation, and exhibits the fact that the Galilei generator H_τ is represented by the Hamilton operator $\hat{H}$.

It is also possible to place the correct factors $\hbar$ in the operators by observing the energy-momentum relations $E = \hbar\omega$ and $\vec{p} = \hbar\vec{k}$ for photons and matter waves.

Rotations

The arguments given for space translations can be transferred to rotations on a nearly one-to-one basis. Consider a rotation around the z-axis through an angle φ:

$$R_\varphi = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

According to (4.10a), there exists a unitary symmetry operator $\hat{U}_\varphi = \exp[i\varphi \hat{J}_\varphi]$ with an observable $\hat{J}_\varphi$. Define $|\alpha'\rangle := \hat{U}_\varphi |\alpha\rangle$. In this case

$$\Psi_{\alpha'}(\vec{x}) := \langle \vec{x} | \alpha' \rangle = \langle \vec{x} | \hat{U}_\varphi | \alpha \rangle = \langle \hat{U}_\varphi^\dagger \vec{x} | \alpha \rangle.$$

Since $U_\varphi^\dagger = U_\varphi^{-1} = R_{(-\varphi)}$, we obtain infinitesimally

$$\Psi_{\alpha'}(\vec{x}) = \langle (x + y\varphi,\, -x\varphi + y,\, z)^T | \alpha \rangle \equiv \Psi_\alpha(x + y\varphi,\, -x\varphi + y,\, z),$$

or with a Taylor expansion

$$\begin{aligned} \Psi_{\alpha'}(\vec{x}) &= \Psi_\alpha(\vec{x}) + \frac{\partial(x + y\varphi)}{\partial\varphi} \frac{\partial \Psi_\alpha(\vec{x})}{\partial x} + \frac{\partial(-x\varphi + y)}{\partial\varphi} \frac{\partial \Psi_\alpha(\vec{x})}{\partial y} + \dots \\ &\equiv \exp\left[-\varphi \big(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \big) \right] \Psi_\alpha(\vec{x}). \end{aligned} \tag{4.16}$$

On the other hand

$$\Psi_{\alpha'}(\vec{x}) = \langle\vec{x}|\hat{U}_\varphi\alpha\rangle = \langle\vec{x}|e^{[i\varphi\hat{J}_\varphi]}\alpha\rangle.$$

The comparison with (4.16) demonstrate that in configuration space, $\hat{J}_\varphi$ can be identified with the third component of the angular momentum operator

$$J^3 = i(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}).$$

Galilei Boosts

Consider a velocity-dependent translation in x-direction: $S(v)x := x + vt$. In the case of invariance of a system under this transformation there is according to (4.10a) a unitary operator $\hat{U}_v = \exp[iv\hat{K}_x]$ with a symmetry observable $\hat{K}_x$, etc. You may, if you are not weary of these routine steps, complete the story. The result is of course that $\hat{K}_x$ corresponds (up to a factor $\hbar$) to the quantum mechanical operator for Galilei boosts $K_x = i\,t\,{}^{\partial}/_{\partial x}$.

4.3.3 Bargmann Group

As we know from special relativity, the Galilei group is not the accurate space-time symmetry group of the world. According to present-day knowledge, this is the Poincaré group. We also saw that under some quite moderate assumptions, the kinematical group may even be the de Sitter group. Thus the representations of the Galilei group seem not to be relevant at all. And this is not that unfavorable, since compared with the Poincaré group, finding the representations of the Galilei group is rather laborious. This is the reason why there are only a few textbooks which deal with this topic. Here, I only sketch the line of argumentation, leaving out the technical details. This is done in order to indicate how delicate the determination of projective representations of non-simple groups can be; more details in [340]. Another reason for including this topic is to be seen in the fact that the Schrödinger equation can be derived through appropriate representations of the Galilei group, or rather the Bargmann group.

Representations of the Galilei Group

- Space and time translations: The symmetry group is topologically $\mathbb{R}^3 \times \mathbb{R}$. This group is simply-connected, and therefore it is identical to its universal covering group. However, central charges cannot be ruled out and may arise in the form $[\hat{P}^j, \hat{P}^k] = i\epsilon^{jkl}C^l \cdot I$.

- Translations + rotations: The rotation group is connected, but not simply-connected. Thus one must investigate the universal covering group **SU**(**2**), with its half- and integer-valued representations; see A.3.5. Central charges can be transformed away. If one considers the larger group of rotations plus translations, it turns out that the central charges among the rotation subalgebra and among the algebra of translations can be defined away, but that central charges remain in the $[\hat{P}, \hat{J}]$ algebra.
- Translations + rotations + Galilei boosts: On appending the Galilei boosts to the previous symmetry transformations, it turns out that the central charges in the $[\hat{P}, \hat{J}]$ algebra can be eliminated, but that one is left with unavoidable central charges in the commutators $[\hat{P}^k, \hat{K}^j] = im^{kj} \cdot I$. From a Jacobi identity with respect to $[[\hat{P}^k, \hat{K}^j], \hat{J}^l]$ one deduces that m^{kj} must be a symmetric matrix, which by taking appropriate linear combinations of the $\hat{P}^k$ and the $\hat{K}^j$ can be made diagonal and proportional to the unit matrix:

$$[\hat{P}^k, \hat{K}^j] = im\delta^{kj} \cdot I. \tag{4.17}$$

  Thus the central extension of the Galilei algebra (called 'Bargmann algebra') contains an additional generator $m \cdot I$. This is the same term that arose in the previous chapter by the Inönü-Wigner contraction of the Poincaré group and the heuristic modification of the energy operator according to (3.100). The extra generator obviously commutes with all other generators and may serve together with the Casimir operator $\hat{J}^2$ to characterize the irreducible representations of the Galilei group.

If one labels the 11-dimensional Bargmann group as $\tilde{g} = (g, \theta)$, where $g = (\tau, \vec{a}, \vec{v}, \mathbf{R})$ is a group element of the Galilei group and θ the additional element from the extension. Then the group multiplication becomes

$$\tilde{g}' \circ \tilde{g} = (g' \circ g, \theta' + \theta + \varphi_m(g', g)) \qquad \text{with} \qquad \varphi_m(g', g) = \frac{1}{2\,m}(v'^2\tau + \vec{v}' \cdot R'\vec{a}).$$

Here $g' \circ g$ is the group composition (2.72) in the Galilei group.

Schrödinger Equation and the Bargmann Group

We will see in the next chapter that symmetries of a system, together with some further reasonable assumptions, restrict the form of the dynamical equations for this system. This is also true of the Schrödinger equation, which can in principle be derived from the representations of the Bargmann group in a Hilbert space. Leaving out details, it will be shown here how the structure of the Hamiltonian operator follows from the interplay of energy, momentum and boost operators.

Since the operators $\hat{H}$ and $\hat{\vec{P}}$ commute, there are eigenstates[4] $|E, \vec{p}\rangle$ with

$$\hat{H}|E, \vec{p}\rangle = E|E, \vec{p}\rangle \qquad \hat{\vec{P}}|E, \vec{p}\rangle = \vec{p}|E, \vec{p}\rangle.$$

How does a Galilei boost affect the momentum of these eigenstates? We first calculate the commutator

$$[\hat{\vec{P}}, e^{-i\vec{v}\cdot\hat{\vec{K}}}] = m\vec{v}\, e^{-i\vec{v}\cdot\hat{\vec{K}}}$$

where m originates from the central extension of the Galilei algebra according to (4.17). Therefore

$$\hat{\vec{P}}\big(e^{-i\vec{v}\cdot\hat{\vec{K}}}|E, \vec{p}\rangle\big) = (m\vec{v} + \vec{p})\big(e^{-i\vec{v}\cdot\hat{\vec{K}}}|E, \vec{p}\rangle\big).$$

This relation reveals that the boosted state $e^{-i\vec{v}\cdot\hat{\vec{K}}}|E, \vec{p}\rangle$ has the momentum $(m\vec{v}+\vec{p})$ (as could be expected).

In much the same manner the energy of a Galilei-boosted state is determined to be:

$$\hat{H}\big(e^{-i\vec{v}\cdot\hat{\vec{K}}}|E, \vec{p}\rangle\big) = \left(E + \vec{v}\cdot\vec{p} + \frac{1}{2}mv^2\right)\left(e^{-i\vec{v}\cdot\hat{\vec{K}}}|E, \vec{p}\rangle\right).$$

If one starts with a state with momentum $\vec{p} = 0$, energy and momentum after a Galilei boost are given by

$$E' = E + mv^2/2 \qquad \vec{p}' = m\vec{v}.$$

The first relation demonstrates that in non-relativistic physics only energy differences can be determined. For $m \neq 0\,(!)$ the velocity can be expressed by the momentum $\vec{v} = \frac{\vec{p}'}{m}$ so that $E' = E + p'^2/2\,m$. And this leads to the non-relativistic Hamiltonian operator

$$\hat{H} = \hat{P}^2/2\,m + \hat{W}$$

with an internal energy W, which commutes with all representation operators of the Galilei transformations.

4.3.4 Symmetries of the Schrödinger Equation

Galilei Covariance of the Schrödinger Equation

We saw that the equations of classical mechanics are covariant with respect to the Galilei group. In which sense is the constitutive equation of non-relativistic quantum mechanics Galilei-covariant? This question has been dealt with in [340] and I essentially follow this article. Let us for simplicity consider the one-particle

[4] Further quantum numbers completely labeling the state are supressed.

Schrödinger equation with a scalar potential V (the case of external fields such as given by an electromagnetic vector field is treated in [58]):

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2\,m}\nabla^2\psi + V\psi.$$

Can we expect that if this equation is transformed to new coordinates

$$\vec{x}' = \mathbf{R}\vec{x} + \vec{v}t + \vec{a} \qquad t' = t + \tau \tag{4.18}$$

the wave function ψ' solving this transformed equation is the same as the original one? We will see that this is not the case in general, but that the transformed wave function at any point differs from the original wave function at the transformed point by a phase factor

$$\psi'(\vec{x}, t) = e^{-if(\vec{x}', t')}\,\psi(\vec{x}', t'). \tag{4.19}$$

The space-time dependent phase factor will now be determined. From (4.18), one obtains

$$\partial/\partial t = \partial/\partial t' + \vec{v}\cdot\vec{\nabla}' \qquad \vec{\nabla} = \mathbf{R}\vec{\nabla}'$$

so that

$$\begin{aligned} 0 &= i\frac{\partial\psi'}{\partial t} + \left(\frac{1}{2\,m}\nabla^2 - V\right)\psi' \\ &= \left[i\frac{\partial}{\partial t'} + i\vec{v}\cdot\vec{\nabla}' + \left(\frac{1}{2\,m}\nabla^2 - V\right)\right]e^{-if(\vec{x}', t')}\,\psi(\vec{x}', t'). \end{aligned}$$

This condition becomes, after applying the differential operators separately to the functions f and ψ (and after dropping the primes):

$$\begin{aligned} &\left(\frac{\partial f}{\partial t} + \vec{v}\cdot\vec{\nabla}f - \frac{1}{2\,m}\nabla^2 f\right)e^{-if}\psi - i\left(\vec{v} - \frac{1}{m}\vec{\nabla}f\right)e^{-if}\cdot\vec{\nabla}\psi \\ &\quad + e^{-if}\left(i\frac{\partial}{\partial t} + \left(\frac{1}{2\,m}\nabla^2 - V\right)\right)\psi = 0. \end{aligned}$$

The last term vanishes because $\psi(\vec{x}, t)$ satisfies the Schrödinger equation. Therefore, two conditions remain:

$$\vec{v} - \frac{1}{m}\vec{\nabla}f = 0 \qquad\qquad \frac{\partial f}{\partial t} + \vec{v}\cdot\vec{\nabla}f - \frac{1}{2m}\nabla^2 f = 0.$$

These have as solution

$$f(\vec{x}, t) = m\vec{v}\cdot\vec{x} - \frac{1}{2}mv^2 t + f_0 \tag{4.20}$$

showing that a space-time dependent phase factor in non-relativistic quantum physics arises only in the case of Galilei boosts ($\vec{v} \neq 0$). Observe the explicit appearance of the mass–playing the role of a central charge.

Lie Point Symmetries of the Schrödinger Equation

Let me re-derive this result by using the methods of S. Lie to find all point symmetries of the Schrödinger equation. For simplicity this consideration is first restricted to the free Schrödinger equation in one spatial dimension, written as

$$\Delta = \psi_t - \lambda \psi_{xx} = 0 \qquad \text{with} \qquad \lambda = \frac{i}{2\,m}.$$

As explained in Sect. 2.2.4 we can find all point symmetry generators $X = \xi^\mu(t, x, \psi)\partial_\mu + \eta(t, x, \psi)\partial_\psi$ (with $x^\mu = (t, x)$) from the condition

$$0 = \overline{X}\Delta = d_t\chi - \lambda d_x d_x \chi\big|_{\Delta=0} \quad \text{where} \quad \chi = \eta - \xi^\mu \psi_\mu \quad \text{and} \quad d_\mu \chi = \chi_\mu + \chi_\psi \psi_\mu.$$

If written explicitly, this contains terms proportional to $\psi_x\psi_{tx},\ \psi_{tx},\ \psi_x\psi_\mu,\ \psi_\mu$ and a term without any derivatives of ψ. All these are required to vanish separately[5]. This results in the defining equations for the functions ξ^μ and η:

$$\xi^\mu_\psi = 0 \quad \eta_{\psi\psi} = 0 \quad \xi^t_x = 0 \quad 2\xi^x_x = \xi^t_t \quad \xi^x_t + 2\lambda\eta_{x\psi} = 0 \quad \eta_t = \lambda\eta_{xx}.$$

The solution of this system of differential equations is

$$\xi^t = c_2 + 2c_4\,t + 4c_6\,t^2 \qquad\qquad \xi^x = c_1 + c_4\,x + c_5\,t + 4c_6\,xt$$
$$\eta = (c_3^+ + i c_3^- - \frac{1}{2\lambda}c_5\,x - 2c_6\,t - \frac{1}{\lambda}c_6\,x^2)\psi + \varphi \qquad\qquad \text{with} \qquad \varphi_t = \lambda\varphi_{xx}$$

containing eight real constants/parameter c_i and a function φ which is required to fulfill the Schrödinger equation. The independent symmetry generating vector fields can thus be chosen as

$$X_1 = \partial_x \qquad\qquad X_2 = \partial_t \qquad\qquad X_4 = 2t\,\partial_t + x\,\partial_x$$
$$X_5 = t\,\partial_x + mxX_- \qquad\qquad X_6 = 4t^2\partial_t + 4tx\,\partial_x - 2tX_+ + 2m\,x^2X_-$$
$$X_+ = (\psi\,\partial_\psi + \psi^*\,\partial_{\psi^*}) \qquad\qquad X_- = i(\psi\partial_\psi - \psi^*\partial_{\psi^*})$$
$$X_\varphi = \varphi(x, t)\,\partial_\psi + \varphi^*(x, t)\,\partial_{\psi^*} \qquad \text{with} \quad \varphi_t = \lambda\varphi_{xx}.$$

[5] Note that if you replace ψ by a real function $u(x, t)$, $u_t = u_{xx}$ is the heat equation. You can find this as a standard example in all books treating the Lie approach for solving differential equations; e.g. [284], [399], [485].

The finite transformations corresponding to (X_1, X_2) are translations in (x, t). From the discussion of the point particle Lie transformations we know that X_4 generates scalings with $x \to e^{\kappa} x$, $t \to e^{2\kappa} x$. Also X_+ and X_- have the structure of scaling transformations which for X_- amounts to the multiplication of ψ with a phase factor. The generator X_φ amounts to $\psi + \varphi(x, t)$ where $\varphi(x, t)$ is itself a solution of the Schrödinger equation. This is of course a symmetry transformation because the Schrödinger equation is a linear differential equation. We are left with interpreting X_5 and X_6. The first term in X_5 is a Galilei boost $x \to x + vt$. Inserting this into the differential equation for $\hat{\psi}$ this becomes

$$\frac{d\hat{\psi}}{dv} = im(\hat{x} + v\hat{t})\hat{\psi}$$

which is solved by

$$\hat{\psi} = \psi \, \exp i\big\{mvx - \frac{m}{2}v^2 t\big\}.$$

This is the phase which the Schrödinger function picks up by a Galilei transformation; compare with (4.20). Finally, you can verify that X_6 corresponds to the finite transformations

$$\hat{x} = \frac{x}{1 - \epsilon t} \qquad \hat{t} = \frac{t}{1 - \epsilon t} \qquad \hat{\psi} = \psi \sqrt{1 - \epsilon t} \, \exp\left[\frac{im}{2}\frac{\epsilon x^2}{(1 - \epsilon t)}\right]. \tag{4.21}$$

Because of a certain similarity with the special conformal transformations of the conformal group I will call these for convenience 'conformal transformations'; sometimes they are called 'expansions' (e.g. in [389]).

The generators X_i and X_α obey an algebra, where the subalgebra formed by the X_i is a Lie algebra. The commutators among the discrete generators and the continuous one

$$\big[X_i, X_\varphi\big] = X_{\varphi^i} = F_i(x, t, \varphi, \partial\varphi)\partial_\psi$$

are also in the algebra in that the F_i obey the Schrödinger equation, too.

The transformation (4.21) seems a little weird, but it leads to a remarkable result. Remember that Lie symmetries define transformations that send solutions of the equations of motion to solutions. If $\psi = F(x, t)$ is a specific solution of the Schrödinger equation, then also $\psi(\hat{x}, \hat{t}, \; F(\hat{x}, \hat{t})$ is a solution. For the transformation (4.21) one thus generates from a solution $F(x, t)$ a solution

$$\psi = (1 + \epsilon t)^{-\frac{1}{2}} \exp\left\{\frac{im}{2}\frac{\epsilon x^2}{1 + \epsilon t}\right\} F\left(\frac{x}{1 + \epsilon t}, \frac{t}{1 + \epsilon t}\right).$$

Specifically, taking for F simply a constant, the symmetry generates a wave packet.

Noether Point Symmetries of the Schrödinger Equation

The one-dimensional Schrödinger equation can be derived from the Lagrangian

$$\mathcal{L}_{\mathrm{S}} = -\frac{1}{2\,m}\psi_x^*\psi_x + \frac{i}{2}(\psi^*\psi_t - \psi_t^*\psi).$$

You may verify that the Noether condition $X_\alpha\mathcal{L}_{\mathrm{S}} + \mathcal{L}_{\mathrm{S}}D_\mu\xi_\alpha^\mu = D_\mu\Sigma_\alpha^\mu$ is fulfilled by the generators $X_1,\ X_2,\ X_-,\ X_5,\ X_6$ (with surface terms for X_5 and X_6), and by the linear combination $X_+ - 2X_4$ of the scale transformations. The generator X_φ is a Noether symmetry for $\varphi = \text{const}$. For each of the Noether generators $X_a = \xi_a^t\partial_t + \xi_a^x\partial_x + \eta_a\partial_\eta + \eta_a^*\partial_{\eta^*}$ we can calculate the currents

$$J_a^t = \frac{i}{2}\psi^*\chi_a - \frac{i}{2}\psi\chi_a^* + \xi_a^t\,\mathcal{L}_{\mathrm{S}} - \Sigma_a^t \qquad J_a^k = -\frac{1}{2m}\psi_x^*\chi_k - \frac{1}{2m}\psi_x\chi_a^* + \xi_a^k\,\mathcal{L}_{\mathrm{S}} - \Sigma_a^k.$$

The conservation laws and conserved charges (assuming an appropriate fall-off at the boundary of spatial integration) are explicitely

- for the phase transformations generated by X_-:

$$\frac{d}{dt}\mathcal{P} + \frac{d}{dx}\mathcal{P}_x = 0 \qquad \text{with} \qquad \mathcal{P} := m\psi^*\psi \quad \mathcal{P}_x := \frac{i}{2}(\psi_x^*\psi - \psi^*\psi_x).$$

  This is interpreted as the conservation of probability $P = \int dx\,\mathcal{P}$.
- for the space translations generated by X_1:

$$\frac{d}{dt}\mathcal{P}_x + \frac{d}{dx}\mathcal{T} = 0 \qquad \text{with} \qquad \mathcal{T} := \frac{1}{2m}\psi_x^*\psi_x + \frac{i}{2}(\psi^*\psi_t - \psi_t^*\psi).$$

  The conserved Noether charge is the momentum $P_x = \int dx\,\mathcal{P}_x$.
- for the time displacements generated by X_2:

$$\frac{d}{dt}\mathcal{E} + \frac{d}{dx}\mathcal{E}_x = 0 \qquad \text{with} \qquad \mathcal{E} := \frac{1}{2m}\psi_x^*\psi_x \quad \mathcal{E}_x := -\frac{1}{2m}(\psi_x^*\psi_t + \psi_t^*\psi_x).$$

  The Noether charge is the energy $E = \int dx\,\mathcal{E}$.
- for the Galilei boosts generated by X_5:

$$\frac{d}{dt}\big(t\mathcal{P}_x - \mathcal{P}x\big) + \frac{d}{dx}\big(t\mathcal{T} - x\mathcal{P}_x\big) = 0.$$

  The Galilei current conservation can be verified by the conservation of the probability and the momentum current. The associated conserved charge is the Galilei momentum $G_x = \int dx\,t\,P_x - x\mathcal{P}$.

- for the scale transformations generated by $(X_+ - 2X_4)$:

$$\frac{d}{dt}\big(x\mathcal{P}_x - 2t\mathcal{E}\big) + \frac{d}{dx}\big(x\mathcal{T} - 2t\mathcal{E}_x + \frac{1}{4\,m^2}\partial_x\mathcal{P}\big) = 0$$

  with the Noether charge $S = 2tE - \int dx\, x\mathcal{P}_x$.
- for the conformal transformations generated by X_6:

$$\frac{d}{dt}\left(tx\mathcal{P}_x - t^2\mathcal{E} - \frac{1}{2}x^2\mathcal{P}\right) + \frac{d}{dx}\left(tx\mathcal{T} - \frac{1}{2}x^2\mathcal{P}_x - t^2\mathcal{E}_x + \frac{1}{4\,m^2}t\partial_x\mathcal{P}\right) = 0$$

  with $C = tS + t^2E - \frac{1}{2}\int dx\, x^2\mathcal{P}$.
- for the transformations generated by X_φ with constant φ:

$$\begin{aligned}\frac{d}{dt}i(\varphi\psi^* - \varphi^*\psi) - \frac{d}{dx}\left[\frac{1}{2\,m}(\varphi\psi^*_x + \varphi^*\psi_x)\right]\\ = \varphi\left(i\psi^*_t - \frac{1}{2\,m}\psi^*_{xx}\right) - \varphi^*\left(i\psi_t + \frac{1}{2\,m}\psi_{xx}\right) = 0.\end{aligned}$$

  The conservation law is nothing but the Schrödinger equation itself.

Schrödinger Group

The previous results can be generalized to the free equation in three dimensions, for which in addition to the already determined Lie and Noether symmetry generators, vector fields arise which generate spatial rotations. The underlying algebra (Schrödinger algebra) in three spatial dimensions, first derived in [253] and [389], contains the Bargmann group generators together with two further generators S and C:

$$\begin{array}{lll} H = \partial_t & T_i = \partial_i & J_k = \epsilon_{kij}x_iT_j + \hat{J}_k \\ G_i = t\,T_i + Mx_i + \hat{G}_i & M = m\mathbb{I} & \\ S = 2t\,H - x^i\,T_i + \hat{S} & C = tS - t^2H - \frac{1}{2}Mx^2 - \hat{G}_ix^i & \end{array}$$

where m is a numerical constant and $\hat{J}$, $\hat{G}$, $\hat{S}$ are numerical matrices that obey the commutation relations

$$[\hat{J}_i, \hat{J}_j] = \epsilon_{ijk}\hat{J}_k \quad [\hat{G}_i, \hat{J}_j] = \epsilon_{ijk}\hat{G}_k \quad [\hat{G}_i, \hat{G}_j] = 0 \quad [\hat{J}_i, \hat{S}] = 0 \quad [\hat{G}_i, \hat{S}] = -\hat{G}_i.$$

Thus the Schrödinger algebra $\mathfrak{sch}(1,3)$ is a 13-parameter Lie algebra spanned by $\{H, T_i, J_i, G_i, M, C, S\}$. It enlarges the centrally extended Galilei algebra with generators $\{H, T_i, J_i, G_i, M\}$ in a similar manner as the conformal algebra $\mathfrak{c}(1,3)$ enlarges the Poincaré algebra $\mathfrak{p}(1,3)$.

Despite analogies between Poincaré/Conformal and Bargmann/Schrödinger, there are differences: Whereas the Galilei group is a Wigner-Inönü contraction of the Poincaré group, the Schrödinger group is not a group contraction of the conformal group.

4.4 Concluding Remarks and Bibliographical Notes

The core principles of quantum mechanics were sketched in the first section of this chapter, but were probably too meager to be understood by someone who has not attended a class in quantum mechanics. Of course there are textbooks on quantum mechanics abound. One of my favorites is the book by P.A.M. Dirac [124], a classic written by one of the fathers of quantum physics, and thus showing a deep understanding of what worried physicists in the twenties of the last century. Not all of these problems have been solved today, and the remark by N. Bohr "If quantum mechanics hasn't profoundly shocked you, you haven't understood it yet" is still valid. Especially the interpretation of quantum physics and its relation to classical physics could not be mentioned at all here. The book by C. Isham [291] has its focus on fundamental issues. In order to complement this chapter, you might read how Isham derives the canonical brackets (4.6) from Wigners theorem and learn more about the connection between unitary transformations and self-adjoined symmetry operators as in (4.10). The deep connections between classical physics in terms of the Hamilton-Jacobi approach and wave mechanics are extensively treated in [167]. Be aware that I did not at all cover the applications of group theory to specific quantum-mechanical systems like atoms and molecules (sometimes called "dynamical symmetries")[6]. These are dealt with directly in standard textbooks on quantum physics or books that specifically address symmetry aspects of non-relativistic quantum theory, such as [83].

[6] These refer to the group representations which H. Weyl had in mind with his statement "All quantum numbers, with the exception of the so-called principal quantum number, are indices characterizing representations of groups".

Chapter 5
Relativistic Field Theory

I was, at one time, greatly interested in establishing all linear equations which are invariant under the inhomogeneous Lorentz group ...

In Chap. 3, Maxwell's electrodynamics was introduced as a relativistic field theory *per se*, and its properties, as derived from symmetries, were investigated on the classical level.

A relativistic field theory complies (by its very name) with the basic rules of special relativity. Especially the action functional from which the dynamical equations derive must be invariant with respect to Poincaré transformations. By this, all fundamental theories need to be formulated as relativistic field theories. But whereas for instance macroscopic effects of gravity are perfectly described by the classical theory of general relativity, there is no "classical" elementary particle physics. Thus the standard model has all features of a "quantized" relativistic field theory. For this reason you won't find textbooks on classical relativistic field theories, but only on what became established as quantum field theory (QFT).

Relativistic quantum field theory had its origin[1] in the year 1926, when M. Born, W. Heisenberg, and P. Jordan applied the rules of nonrelativistic quantum mechanics to quantize a (nonrelativistic) string, this representing a non-polarized free radiation field. And within only three years the theory of general quantum fields was laid out, culminating in the 1929/30 articles by W. Heisenberg and W. Pauli. Their formalism is essentially the one presented in textbooks (and it is sketched in this chapter, too). By this it had become a point of common understanding that in quantized electrodynamics, photons can be created and destroyed. In the period thereafter it became clear that the basic entities are quantum fields, and particles are lumps of energy

---

E. P. Wigner in "Invariant Quantum Mechanical Equations of Motion", in: Theoretical Physics Lectures presented in Trieste, Italy; International Atomic Energy Agency, Vienna 1963, 59–82.

---

[1] A short and thorough historical reflection is Chap. 1 in [536].

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_5,

of the fields. Therefore, it was recognized already by around 1935 that a quantum field should be interpreted to be a theory of quantized subatomic particles and their interaction.

Quantum field theory has many facets and is by far not an established discipline. Particle theorists–myself included–have in mind theories which are formulated in terms of renormalized perturbation theory. But there are variants of QFT aspiring to be more fundamental, such as axiomatic field theories that start with a handful of Wightman axioms (which mainly concern symmetry aspects!) or algebraic field theory (D. Buchholz, R. Haag) based on so called $C^\star$-algebras–sets of bounded operators of a Hilbert space representing local observables. And while in particle physics, the notion of gauge fields–originating from local internal symmetries–is essential, these have no place at all in the algebraic field theory, and they are hard to implement in axiomatic field theories. One should also be aware that quantum field theory as it stands today, and which successfully describes a quite impressive range of phenomena, is not the only way to reconcile the postulates of special relativity and quantum physics. There is even an example of an alternative in the form of the string models–although momentarily it is far from clear that these in fact describe our world.

This chapter is not about full-fledged quantum field theory. Instead, those aspects of symmetries which can be understood on the classical level are treated here. Indeed, it is a difficult endeavor to remain classical, especially in reference to the standard model. At many places in this chapter therefore, although I will use the terminology and notations from a quantum theory, you should keep in mind that I am not intending to really deal with QFT.

Nevertheless, from the point of symmetries many insights into a field theory can be attained without going "quantum". This is good news. The bad news is that classical symmetries may cease to be symmetries of the quantized theory; for instance by quantum/radiative correction giving rise to "anomalies", or in the case that the vacuum is not symmetric ("spontaneous symmetry breaking"). And a book on "Symmetries in Fundamental Physics" would not deserve its title if it neglects these phenomena. But here I get into hot water as an author, since I must deal with deep-lying quantum field topics, although aiming to stay classical.

The same is true for renormalizability, which as will be seen, serves as a sieve for ruling out certain field-theory actions. This is of course a not-so-easily comprehensible topic. But without it, I could not deliver my message on the running coupling constants (motivating a later section on Grand Unified Theories), nor could I communicate a notion of effective field theories, in which for example the requirement of renormalizability is given up in a controlled way, such that symmetries are still respected.

After these remarks about what this chapter is not–or only partly–about, let me tell you what it *is* about: Since we assume invariance with respect to Poincaré transformations, this chapter deals with the (unitary) representations of the Poincaré group, culminating in Eugene Wigner's classification of particle types, possible field variants and their dynamical equations (encoded in actions). Possible field variants are scalar, spinor, vector, ... fields. Higher-spin fields are mentioned; however it seems that nature stops at spin-2 fields. The basic spinors are Weyl-spinors. These naturally appear in the Lagrangian for the standard model and in supersymmetric theories. Vector fields enter the scene as gauge bosons if one wants to turn a global symmetry into a local one. It is here, that generalizations of Maxwell's theory show up in the form of non-Abelian Yang–Mills theories. It will be demonstrated that the kinetic term for the gauge fields is largely dictated by Klein–Noether identities, and that the interaction terms in the Lagrangian are heavily restricted by a criterion of renormalizability.

Although the symmetries dealt with in this monograph are mainly continuous ones, in particle physics we need to deal also with the discrete symmetries P (space inversion), T (time reversal), and C (charge conjugation). And as seems to be the case specifically for the electroweak sector of the SM (but not only there), nature sometimes breaks a symmetry in a tricky manner, called spontaneous symmetry breaking: Here the symmetry of an action is no longer a symmetry of the "ground state". This is treated in Sect. 5.4. The last section refers to effective field theories, a notion which is becoming more and more accepted. Among other things, it allows to relate theories which apply to phenomena on different levels of granularity or on different length scales.

As mentioned at various places before, the underlying structure of the standard model of elementary particles is that of a relativistic quantum field theory. In essence it is defined by its path integral which in its core contains a classical action. Now as it happens, that there is no "classical" elementary particle physics. But nevertheless if we insist on deriving for instance scattering amplitudes from the path integral, we need to reflect on actions.

Aside from this spacetime symmetry we will also impose a local internal symmetry (called "gauge" symmetry). These symmetry requirements restrict at first the possible variants of fields, and then the terms of the free and the interaction part in any Lagrangian. This is one of the main messages of this chapter. Another quest, originating from the quantized theory, and known as renormalizability, restricts the possible types of action so drastically that only a few relativistic quantum field theories are conceivable.

5.1 Representations of the Poincaré Group

Since the Poincaré group is not compact, all its unitary representations (except the trivial one) are infinite-dimensional. They can, however, be organized by the help of what Wigner called "little groups". These are the subgroups of transformations that are left over after the non-compact transformations (i.e. translations and boosts) are fixed. Before introducing this Wigner trick some properties of the Poincaré group are recalled.

5.1.1 Global Structure of ISO(3, 1)

The Poincaré group $\mathbf{ISO(3,1)}$ is isomorphic to the direct sum of translations and the Lorentz group. As described in Sect. 3.4.1, the Lorentz group $\mathbf{Lor} \cong \mathbf{O(3,1)}$ is composed of four different components, denoted by $\mathbf{Lor}^{\{\uparrow\downarrow\}}_{\{\pm\}}$. Among these only $\mathbf{Lor}^{\uparrow}_{+}$ constitutes a group, since it alone contains the neutral element. However, this group is not simply-connected. Its universal covering group is $\mathbf{SL}(2,\mathbb{C})$; for a derivation see Appendix A.3.5. Therefore, if it is a question of "representations of the Poincaré group", this is essentially about "representations of $\mathbf{SL(2,\mathbb{C})} \ltimes R^4$". What about possible central charges in the generator algebra associated to this group? Allow initially that the Poincaré algebra contains central charges:

$$\begin{aligned}
[M_{\mu\nu}, M_{\varrho\sigma}] &= i\left[\eta_{\mu\sigma}M_{\nu\varrho} - \eta_{\nu\sigma}M_{\mu\varrho} + \eta_{\nu\varrho}M_{\mu\sigma} - \eta_{\mu\varrho}M_{\nu\sigma}\right] + C_{\mu\nu,\varrho\sigma}\\
[M_{\mu\nu}, T_\lambda] &= i\left[\eta_{\nu\lambda}T_\mu - \eta_{\mu\lambda}T_\nu\right] + C_{\mu\nu,\lambda}\\
[T_\lambda, M_{\mu\nu}] &= i\left[\eta_{\mu\lambda}T_\nu - \eta_{\nu\lambda}T_\mu\right] + C_{\lambda,\mu\nu}\\
[T_\mu, T_\nu] &= i\, C_{\mu,\nu}.
\end{aligned}$$

The $C's$ inherit index symmetries from the generators and additionally satisfy

$$C_{\varrho\sigma,\mu\nu} = -C_{\mu\nu,\varrho\sigma} \qquad C_{\lambda,\mu\nu} = -C_{\mu\nu,\lambda} \qquad C_{\nu,\mu} = -C_{\mu,\nu}.$$

The sole restrictions on the $C's$ come from the Jacobi identities, resulting in algebraic conditions. They can be found in detail in e.g. [536], Sect. 2.7. Here I merely sketch the results. It is found that for consistency $C_{\mu,\nu} = 0$. Furthermore,

$$C_{\lambda,\mu\nu} = \eta_{\lambda\nu}\bar{C}_\mu - \eta_{\lambda\mu}\bar{C}_\nu, \qquad \bar{C}_\mu = \frac{1}{3}\eta^{\rho\sigma}C_{\rho,\mu\sigma}$$

$$C_{\mu\sigma,\nu\rho} = \eta_{\rho\mu}\bar{C}_{\nu\sigma} - \eta_{\nu\mu}\bar{C}_{\rho\sigma} + \eta_{\sigma\nu}\bar{C}_{\rho\mu} - \eta_{\rho\sigma}\bar{C}_{\nu\mu}, \qquad \bar{C}_{\nu\sigma} = \frac{1}{2}\eta^{\mu\rho}C_{\nu\mu,\rho\sigma}.$$

Therefore, if the $C's$ are not already zero, they can be eliminated completely by defining the generators $\bar{M}_{\mu\nu} = M_{\mu\nu} + \bar{C}_{\mu\nu}$ and $\bar{T}_\mu = T_\mu + \bar{C}_\mu$. These then satisfy the Poincaré algebra without central charges.

5.1.2 Transformation of the Generators

Like any other symmetry operation also the Poincaré transformations (denoted as $p = (\Lambda, a)$) induce unitary linear transformations on vectors in a Hilbert space: $|\psi\rangle \to U(\Lambda, a)|\psi\rangle$. The operators U reflect the group composition (3.91)

$$U(\Lambda', a')U(\Lambda, a) = U(\Lambda'\Lambda, \Lambda' a + a').$$

Here it is assumed that all possible central charges were removed and thus no phase factor appears on the right-hand side. Infinitesimally, $\Lambda^\mu_\nu = \delta^\mu_\nu + \omega^\mu_\nu$ with

$\omega_{\mu\nu} + \omega_{\nu\mu} = 0,\ a_\mu = \epsilon_\mu$, and the unitary symmetry operator becomes, with the sign convention as in (3.93),

$$U(1+\omega, \epsilon) = 1 - \frac{i}{2}\omega_{\mu\nu} M^{\mu\nu} + i\epsilon_\mu P^\mu$$

with Hermitean operators $M^{\mu\nu}$ and P^μ. In order to derive the transformation behavior of the $M^{\mu\nu}$ and P^μ consider the expression $U(\Lambda, a)U(1+\omega,\epsilon)U^{-1}(\Lambda, a)$. By making use of $U^{-1}(\Lambda, a) = U(\Lambda^{-1}, -\Lambda^{-1}a)$ one finds

$$U(\Lambda, a)U(1+\omega,\epsilon)U^{-1}(\Lambda, a) \equiv U\left(1 + \Lambda\omega\Lambda^{-1}, \Lambda\epsilon - \Lambda\omega\Lambda^{-1}a\right).$$

After expanding both sides in terms of ω and ϵ and keeping track of first order terms only, we obtain

$$\begin{aligned} &U(\Lambda, a)\left(-\frac{1}{2}i\omega_{\mu\nu}M^{\mu\nu} + i\epsilon_\mu P^\mu\right)U^{-1}(\Lambda, a) \\ &\qquad = -\frac{1}{2}\left(\Lambda i\omega\Lambda^{-1}\right)_{\mu\nu} M^{\mu\nu} + (\Lambda i\epsilon - \Lambda i\omega\Lambda^{-1}a)_\mu P^\mu. \end{aligned}$$

By equating the coefficients of the ω and the ϵ on both sides we find[2]

$$U(\Lambda, a)\ iM^{\mu\nu}\ U^{-1}(\Lambda, a) = i\Lambda_\varrho{}^\mu \Lambda_\sigma{}^\nu\ (M^{\varrho\sigma} + a^\varrho P^\sigma - a^\sigma P^\varrho) \tag{5.1a}$$

$$U(\Lambda, a)\ iP^\mu\ U^{-1}(\Lambda, a) = i\Lambda_\varrho{}^\mu\ P^\varrho, \tag{5.1b}$$

with the understanding that $\Lambda_\nu{}^\mu = (\Lambda^{-1})^\mu{}_\nu$. Thus, for Lorentz transformations (for which $a \equiv 0$) $M^{\mu\nu}$ transforms as a tensor and P^μ transforms as a vector. Further, for translations ($\Lambda \equiv 0$) P^μ is translation invariant. By considering in the previous expressions also the $U(\Lambda, a)$ in their infinitesimal form, one confirms that the $M^{\mu\nu}$ and P^μ fulfill the Poincaré algebra.

5.1.3 The "Little Group"

As a first step towards the representations of the Poincaré group on one-particle state vectors in a Hilbert space, consider translations only. Since all P^μ commute amongs themselves, every state can be expanded in terms of their common eigenfunctions:

$$P^\mu |p, \sigma\rangle = p^\mu |p, \sigma\rangle. \tag{5.2}$$

[2] You may wonder why the factors with the imaginary unit are kept in these expressions. Indeed for unitary operators $U(\Lambda, a)$ they cancel on both sides of these relations. Later, however, when dealing with discrete symmetries in Sect. 5.5, we find that these factors are the key for deciding about unitarity and anti-unitarity.

Here σ stands for other possible labels. Under translations these states transform as

$$U(\mathbf{1}, a)|p, \sigma\rangle = e^{ia\cdot p}|p, \sigma\rangle. \tag{5.3}$$

In the next step, we determine how states of the form (5.3) transform under Lorentz transformations $U(\Lambda, 0) =: U(\Lambda)$. Write

$$P^{\mu}U(\Lambda)|p, \sigma\rangle = U(\Lambda)[U^{-1}(\Lambda)P^{\mu}U(\Lambda)|p, \sigma\rangle$$

and take into account the transformation behavior (5.1b) of P^{μ} in order to obtain

$$P^{\mu}U(\Lambda)|p, \sigma\rangle = U(\Lambda)[(\Lambda^{-1})_{\varrho}{}^{\mu}P^{\varrho}]|p, \sigma\rangle = \Lambda^{\mu}{}_{\varrho}\, p^{\varrho}U(\Lambda)|p, \sigma\rangle.$$

This shows that $U(\Lambda)|p, \sigma\rangle$ is an eigenstate of P^{μ} with eigenvalue (Λp). Furthermore, since $|\Lambda p, \sigma\rangle$ by definition spans the eigenspace of ΛP, it must be the case that

$$U(\Lambda)|p, \sigma\rangle = \sum_{\sigma'} C_{\sigma\sigma'}(\Lambda, p)|\Lambda p, \sigma'\rangle. \tag{5.4}$$

In Sect. 2.5 of [536] it is argued that by an appropriate choice of the basis, the coefficients $C_{\sigma\sigma'}$ can be assumed to be block-diagonal. This means that in order to arrive at irreducible representations, it suffices to consider one block of $C(\Lambda, p)$ since this itself constitutes a representation of the Lorentz group. The obvious question is how the coefficients $C_{\sigma\sigma'}$ look in irreducible representations of the Poincaré group.

At this place a rather remarkable trick due to E. Wigner [555] is useful, known under the keyword "little group". This was later generalized by G. W. Mackey and became known as "induced representation". The trick is to reduce the quest for a representation of the Poincaré group to the representation of a smaller group. For this purpose, one introduces "standard momenta" k^{μ} related to the four momentum in a specific way, but obeying $p^2 = k^2$. Then the standard momenta serve to seek directly the representation of the smaller group. For each choice of a standard momentum k^{μ} exists a Lorentz transformation L with

$$p^{\mu} = L^{\mu}{}_{\nu}(p)k^{\nu}. \tag{5.5}$$

The eigenstates of the translation operators defined in (5.2) can now be transformed in states referring to a standard momentum as

$$|p, \sigma\rangle = N(p)\; U(L(p))|k, \sigma\rangle, \tag{5.6}$$

where $N(p)$ is a normalization factor. Acting with an arbitrary Lorentz transformation on this state the resulting state can be written as

$$U(\Lambda)|p, \sigma\rangle = N(p)\, U[\Lambda L(p)]|k, \sigma\rangle = N(p)\, U[L(\Lambda p)]U[L^{-1}(\Lambda p)\Lambda L(p)]|k, \sigma\rangle. \tag{5.7}$$

Observe that in this expression the Lorentz transformation

$$W(\Lambda, p) := L^{-1}(\Lambda p)\Lambda L(p), \tag{5.8}$$

called a *Wigner rotation*, returns the vector k back to k via the sequence $k \Rightarrow L(p)k = p \Rightarrow \Lambda p \Rightarrow L^{-1}(\Lambda p) = k$. Therefore, W belongs to the subgroup of the Lorentz transformations that leave k^μ invariant:

$$W^\mu{}_\nu k^\nu = k^\mu. \tag{5.9}$$

This group is called the "little group" (with respect to the standard momentum k). For every W with this property, we have

$$U(W)|k, \sigma\rangle = \sum_{\sigma'} D_{\sigma\sigma'}(W)|k, \sigma'\rangle.$$

The coefficients $D(W)$ constitute a representation of the little group, since

$$\begin{aligned}
\sum_{\sigma'} D_{\sigma\sigma'}(\tilde{W}W)|k, \sigma'\rangle &= U(\tilde{W}W)|k, \sigma\rangle = U(\tilde{W})U(W)|k, \sigma\rangle \\
&= U(\tilde{W}) \sum_{\sigma''} D_{\sigma\sigma''}(W)|k, \sigma''\rangle \\
&= \sum_{\sigma',\sigma''} D_{\sigma\sigma''}(W) D_{\sigma''\sigma'}(W)\sigma''\rangle
\end{aligned}$$

exhibiting that the $D(W)$ do indeed have the property of a representation:

$$D_{\sigma\sigma'}(\tilde{W}W) = \sum_{\sigma',\sigma''} D_{\sigma\sigma''}(\tilde{W}) D_{\sigma''\sigma}(W).$$

Inserting the expression (5.8) for the Wigner rotation into (5.7), we get

$$U(\Lambda)|p, \sigma\rangle = N(p) \sum_{\sigma'} D_{\sigma\sigma'}(W(\Lambda, p))U(L(\Lambda p)|k, \sigma'\rangle$$

which by virtue of (5.6) becomes

$$U(\Lambda)|p, \sigma\rangle = \frac{N(p)}{N(\Lambda p)} \sum_{\sigma'} D_{\sigma\sigma'}(W(\Lambda, p))|\Lambda p, \sigma'\rangle. \tag{5.10}$$

It is argued in Chap. 2 of [536] that for the choice of the normalization factor as $N(p) = \sqrt{k^0/p^0}$, the scalar product on states becomes simply $\langle p', \sigma'|p, \sigma\rangle = \delta_{\sigma\sigma'}\delta^3(\vec{p}' - \vec{p})$.

Table 5.1 Orbits and little groups

Situation	Orbit	Reference k^μ	Little group	Particle
$p^2 = m^2 > 0$	Mass-shell	$(\pm m, 0, 0, 0)$	$SO(3)$	Massive
$p^2 = 0$	Light-cone	$(\pm E, 0, 0, E)$	$ISO(2)$	Massless
$p^2 = -\mu^2 < 0$	Hyperboloid	$(0, 0, 0, \mu)$	$SO(2, 1)^\uparrow$	Tachyonic
$p^\mu = 0$	Origin	$(0, 0, 0, 0)$	$SO(3, 1)^\uparrow$	Vacuum

5.1.4 Classification of Particles

Case Differentiation

The only (independent) functions of the momentum components p^μ which are invariant under proper orthochronous Lorentz transformations $\mathbf{Lor}_+^\uparrow$ are $p^2 = p_\mu p^\mu$ and $sign(p^0)$, the algebraic sign of p^0.
The little group depends on the choice of the standard momentum k. Obviously one must only consider qualitatively different choices, and these can be read of from the six qualitatively distinct solutions of the mass shell condition $p^2 - m^2 = 0$ with respect to the light cone. A solution p^μ can either be located in the future or the past, and either inside, outside, or directly on the light cone. The six qualitatively different solutions are

(a) $p^2 = m^2 > 0, \quad p^0 > 0$
(b) $p^2 = 0, \quad p^0 > 0$
(c) $p^\mu = 0$
(d) $p^2 = m^2 > 0, \quad p^0 < 0$
(e) $p^2 = 0, \quad p^0 < 0$
(f) $p^2 = -\mu^2 < 0$

and they are also set out in Table 5.1. Below, we will only consider cases (a)–(c), since they allow a physical interpretation. Some comments on the other three cases will be made in the concluding remarks of this subsection. Case (c) is simply the vacuum state, which certainly is invariant under $U(\Lambda)$. The cases (a) and (b) relate to massive and massless objects (with positive energies), which are treated separately in the sequel. In order to determine the little groups for the different cases, we need to consider the Casimir operators of the Poincaré group.

Casimir Operators

As known from the group of rotations, the determination of its irreducible representations is eased by using its Casimir operator (see Appendix A.3.5). A Casimir operator is an operator that commutes with all generators of the Lie algebra associated to the group. In case of the Poincaré algebra, the only two independent ones are

- the operator $P^2 = P_\mu P^\mu$, indeed

$$[P^2, P_\mu] = 0, \qquad\qquad [P^2, M_{\mu\nu}] = 0,$$

where the first commutator relation is obvious because the momentum operators mutually commute, and the commutator with the Lorentz generators can either be calculated explicitly or simply be deduced, since P^2 is a Lorentz-scalar.

- the operator $W^2 = W_\mu W^\mu$ which is constructed from the Pauli-Łubanski vector

$$W_\mu := \frac{1}{2}\epsilon_{\mu\nu\varrho\sigma} M^{\nu\varrho} P^\sigma .$$

The Pauli-Łubanski vector obeys

$$[W_\mu, P_\nu] = 0 \quad [W_\mu, M_{\rho\sigma}] = i(\eta_{\mu\sigma} W_\rho - \eta_{\mu\rho} W_\sigma) \quad [W_\mu, W_\nu] = i\epsilon_{\mu\nu\varrho\sigma} W^\rho P^\sigma$$

and furthermore, $W^\mu P_\mu = 0$. For later purposes the Pauli-Łubanski vector is expressed in terms of the (1+3)-components as

$$W_0 = -\vec{J} \cdot \vec{P} \qquad\qquad W^j = -E J^j - (\vec{K} \times \vec{P})^j . \tag{5.11}$$

Let us convince ourselves that for every choice of a standard momentum k^μ, the components of the vector

$$\overset{(k)}{W}_\mu := \frac{1}{2}\epsilon_{\mu\nu\varrho\sigma} M^{\nu\varrho} k^\sigma \tag{5.12}$$

are generators of the little group: Infinitesimally the Wigner rotation has the form $(\delta^\mu{}_\nu - \overset{(k)}{\omega}{}^\mu{}_\nu)$. The condition

$$(\delta^\mu{}_\nu - \overset{(k)}{\omega}{}^\mu{}_\nu) k^\nu = k^\mu \qquad \text{or} \qquad \overset{(k)}{\omega}{}^\mu{}_\nu \, k^\nu = 0$$

is solved by $\overset{(k)}{\omega}_{\mu\nu} = \epsilon_{\mu\nu\varrho\sigma} n^\varrho k^\sigma$ with an arbitrary four-vector n^ϱ. Therefore, the representation becomes

$$U(W) = I + \frac{i}{2} \overset{(k)}{\omega}_{\mu\nu} \, M^{\mu\nu} = I + i n^\varrho \left[\frac{1}{2}\epsilon_{\mu\nu\varrho\sigma} M^{\mu\nu} P^\sigma \right]_{P \to k} ,$$

that is

$$U(W)|p, \beta\rangle = e^{i n^\mu W_\mu} |p, \beta\rangle .$$

Because of $W^\mu P_\mu = 0$ the little group has the dimension 3 except for case (c). The Casimir operators may now serve to characterize the states as

$$P^2 |p^2, w^2, \gamma\rangle = p^2 |p^2, w^2, \gamma\rangle \qquad\qquad W^2 |p^2, w^2, \gamma\rangle = w^2 |p^2, w^2, \gamma\rangle . \tag{5.13}$$

Representations for Massive Particles

For case (a) with $m^2 > 0$, it seems natural to go into the rest system of the particle and to choose

$$(k^\mu) = (m, \vec{0}) \tag{5.14}$$

as the standard momentum. The only Lorentz transformations from $\mathbf{Lor}_+^\uparrow$, which leave a particle in rest are rotations in three dimensions. Therefore the little group belonging to the standard momentum (5.14) is just $\mathbf{SO(3)}$. This is manifest also in that (5.12) in this case becomes

$$\overset{(k)}{W}_\mu = m\frac{1}{2}\epsilon_{\mu\nu\varrho 0}M^{\nu\varrho} = -m\frac{1}{2}\epsilon_{0\mu jk}M^{jk} = (0, -m\vec{J}) \quad \text{with} \quad J^i = \frac{1}{2}\epsilon^{ijk}M^{jk}.$$

Thus W^2/m^2 is, up to a sign, identical to the Casimir operator J^2 of $\mathbf{SO(3)}$. Now we are in a position to specify states (5.13) which are simultaneously eigenstates of the Casimir operators P^2 and W^2:

$$\begin{aligned} P^2|m, s, \ldots\rangle &= m^2|m, s, \alpha\rangle \\ W^2|m, s, \ldots\rangle &= -m^2\, s(s+1)|m, s, \alpha\rangle. \end{aligned}$$

Which further quantum numbers can serve to characterize the state? Since all components of the four-momentum commute with the Casimir operators (and amongst themselves), one may consider the three-momentum $\vec{p}$ (the fourth component is then already taken into account by the mass shell condition). It is not possible to select any single component of the angular momentum (such as J_z) since $\vec{J}$ and $\vec{P}$ do not commute. However $\vec{P}$ commutes with $\vec{J}\cdot\vec{P}$ ($\propto \bar{W}P$). This is the so-called *helicity* λ, that is the component of the angular momentum in the direction of the momentum. A complete characterization of a state is thus provided by

$$\begin{aligned} P^2|m, s, \vec{p}, \lambda\rangle &= m^2|m, s, \vec{p}, \lambda\rangle \\ \vec{P}\,|m, s, \vec{p}, \lambda\rangle &= \vec{p}\,|m, s, \vec{p}, \lambda\rangle \\ \vec{J}\cdot\vec{P}\,|m, s, \vec{p}, \lambda\rangle &= \lambda\mid\vec{p}\mid|m, s, \vec{p}, \lambda\rangle \\ W^2|m, s, \vec{p}, \lambda\rangle &= -m^2 s(s+1)|m, s, \vec{p}, \lambda\rangle. \end{aligned}$$

As an alternative one may choose as a basis $|E, \vec{p}, s, s_3\rangle$, where $E = \sqrt{\vec{p}^2 + m^2}$ and s_3 is the eigenvalue of $\frac{1}{m}W_3$.

Representations for Massless Particles

Since for massless particles no rest system exists, and since in this case the standard momentum must be a null vector, one may choose

$$(k^\mu) = (E, 0, 0, E) \quad \text{with} \quad E > 0,$$

which according to (5.12) leads to the generators

$$\overset{(k)}{W}_0 = -EJ^3 = \overset{(k)}{W}{}^3 \qquad \overset{(k)}{W}{}^1 = -E(J^1 - K^2) \qquad \overset{(k)}{W}{}^2 = -E(J^2 + K^1).$$

We might for instance take as independent generators the set $\{J^3,\ T^1 := K^1 + J^2,\ T^2 := K^2 - J^1\}$ with the algebra

$$[T^1, T^2] = 0 \tag{5.15a}$$
$$[J^3, T^1] = -iT^2 \tag{5.15b}$$
$$[J^3, T^2] = iT^1. \tag{5.15c}$$

This is isomorphic to the $\mathfrak{iso}(2)$ algebra: T^1 and T^2 serve as commuting generators of "translations" and J^3 as the generator of a rotation in a plane perpendicular to $\vec{k}$. Self-evidently, the "translations" and the rotation are defined in an abstract (T^1, T^2)-plane. The Casimir operator becomes in this case

$$\overset{(k)}{W}{}^2 = -(T^1)^2 - (T^2)^2.$$

Together with one of the generators (T^1, T^2, J^3) it can serve to constitute the irreducible states. In analogy with the construction of representations for $\mathbf{SO(3)}$ a preferred choice is J^3. There is even a further analogy: Defining the linear combinations $L_\pm = T^2 \pm iT^1$, one realizes from

$$[J^3, L_\pm] = \pm L_\pm,$$

that the $L_\pm$ act like ladder operators. A crucial difference from $\mathbf{SO(3)}$ is that in the three-dimensional case, the components of the angular momenta are subject to the restriction $\langle (J^3)^2\rangle \le \langle \vec{J}^2\rangle$, with the consequence that the spectrum of J^3 is bounded. This is not the case in the present situation. However, since the operators T^1 and T^2 commute with one another, they can be simultaneously diagonalized and we may set

$$T^1|k, w^2, \beta\rangle = a|k, w^2, \beta\rangle \qquad T^2|k, w^2, \beta\rangle = b|k, w^2, \beta\rangle.$$

In order to understand that the eigenvalues a and b are not reasonable state labels or–as we will see–make sense only for $a = 0 = b$, we make a short digression: Define functions $f(\theta)$ and $g(\theta)$ by

$$f(\theta) = e^{-i\theta J^3} T^1 e^{i\theta J^3}|k, w^2, \beta\rangle \qquad g(\theta) = e^{-i\theta J^3} T^2 e^{i\theta J^3}|k, w^2, \beta\rangle.$$

The algebra (5.15) yields the differential equations $df/d\theta = g$, $dg/d\theta = -f$ which, together with the initial conditions $f(0) = a|k, w^2, \beta\rangle$, $g(0) = b|k, w^2, \beta\rangle$, are solved by

$$f(\theta) = (a\cos\theta + b\sin\theta)|k, w^2, \beta\rangle \qquad g(\theta) = (b\cos\theta - a\sin\theta)|k, w^2, \beta\rangle.$$

Therefore

$$T^1 e^{i\theta J^3}|k, w^2, \beta\rangle = (a\cos\theta + b\sin\theta)e^{i\theta J^3}|k, w^2, \beta\rangle$$
$$T^2 e^{i\theta J^3}|k, w^2, \beta\rangle = (b\cos\theta - a\sin\theta)e^{i\theta J^3}|k, w^2, \beta\rangle.$$

This shows that the state $\exp(i\theta J^3)|k, w^2, \beta\rangle$ is an eigenstate of T^1 and T^2 for all values θ. Without further restrictions this leads to what Wigner calls "continuous spin" or "infinite tower". The appearance of this infinite tower is directly understandable from the perspective of group representation: The little group in the massless case is isomorphic to the non-compact group of Euclidean motions on the plane. Being a non-compact group, this little group's unitary representations are infinite-dimensional, except when the eigenvalues of the "translations" are zero, in which case it effectively reduces to $\mathbf{SO(2)}$. The infinite-dimensional representations are considered unphysical because we never see particle states in nature labelled by extra continuous parameters. Since the "continuous spin" states correspond to $W^2 = -(a^2 + b^2) \neq 0$ one limits the physical states to $W^2 = 0$ that means to $a = b = 0$ or

$$T^1|k, w^2, \beta\rangle = 0 \qquad\qquad T^2|k, w^2\beta\rangle = 0.$$

Now the three conditions $W^\mu W_\mu|\,\rangle = W^\mu P_\mu|\,\rangle = P^\mu P_\mu|\,\rangle = 0$ require that W_μ and P_μ must be parallel:

$$W_\mu = hP_\mu$$

from which, due to (5.11):

$$W_0 = -\vec{J}\cdot\vec{P} = hP_0 = \pm h|\vec{P}|$$

and the quantity h can be understood as the helicity

$$h = \frac{\vec{P}\cdot\vec{J}}{|\vec{P}|}.$$

Denoting the action of h on its eigenstates as $h|k, \lambda, \gamma\rangle = \lambda|k, \lambda, \gamma\rangle$, we finally find for physical states that for massless particles

$$W^\mu|k, \lambda, \gamma\rangle = \lambda k^\mu|k, \lambda, \gamma\rangle.$$

This result conforms with a pure group theoretical line of argument. After having thrown out the "translations" from the little group, we are left with the group $\mathbf{SO(2)}$. Representations of $\mathbf{SO(2)}$ are one dimensional, labelled by a single eigenvalue, the helicity $|\lambda\rangle$. Algebraically λ could be any real number, but there is a topological constraint: Since the helicity is the eigenvalue of the rotation generator around the z-axis, a rotation by an angle θ around that axis produces a phase $\exp(i\theta\lambda)$ on wave functions. Now, the Lorentz group is not simply connected: while a 2π-rotation cannot be continuously deformed to the identity, a 4π-rotation can. This implies that the phase $\exp(4\pi i\lambda)$ must be one. Therefore the helicity is quantized as $\lambda \in \frac{1}{2}\mathbb{Z}$.

I leave off the discussion of the cases (d)–(f), since for these no counterpart have been observed in nature. Cases (d) and (e) describe objects with negative energies, case (f) is a tachyon. The respective little groups and standard momenta are listed in Table 5.1. Mathematically all the six cases (a)–(f) occur. Where are the further "particles" allowed in principle by Wigner's classification scheme? If only the cases (c), (a) and that part of (b) for which there is no continuous spin are realized in nature, the rest should be excluded/forbidden by some other fundamental principle. If there are no such principles . . . .. we know that if something is not forbidden it is allowed. I will come back to this point in the concluding section.

5.2 Symmetry and Quantum Field Theory

5.2.1 Lorentz Symmetry Rules Field Variants

In a relativistic field theory, the fields occuring necessarily must transform according to representations of the Lorentz group. Or conversely: If one knows all representations of the Lorentz group, one knows all possible variants of fields that may appear in a relativistic field theory. Generically a field Φ_α $(\alpha = 1, \ldots, N)$ transforms under a Lorentz transformation $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ as

$$\Phi'_\alpha(x') = D(\Lambda)_\alpha{}^\beta \Phi_\beta(x) \quad \text{or in matrix form} \quad \Phi'(x') = D(\Lambda)\Phi(x), \tag{5.16}$$

where $D(\Lambda)$ is a $(N \times N)$ matrix representation of the Lorentz group[3]:

$$D(\Lambda)_\alpha{}^\beta = \left[\exp(-\frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu})\right]_\alpha^\beta \tag{5.17}$$

with the $(N \times N)$ spin matrix Σ. Infinitesimally, with $\Lambda = 1 + \omega$,

$$\delta\Phi = \Phi'(x') - \Phi(x) = -\frac{i}{2}\omega_{\mu\nu}\Sigma^{\mu\nu}\,\Phi(x) \tag{5.18a}$$

$$\begin{aligned}\bar{\delta}\Phi &= \delta\Phi - \Phi_{,\mu}\delta x^\mu = \delta\Phi - \Phi_{,\mu}\omega^{\mu\nu}x_\nu \\ &= -\frac{i}{2}\omega_{\mu\nu}(L^{\mu\nu} + \Sigma^{\mu\nu})\Phi(x) = -\frac{i}{2}\omega_{\mu\nu}M^{\mu\nu}\,\Phi(x)\end{aligned} \tag{5.18b}$$

with $L_{\mu\nu} = i(x_\mu\partial_\nu - x_\nu\partial_\mu)$.

As will be derived later in this section, the lowest-dimensional representations of the Lorentz group are scalars, spinors and vectors. Accordingly

- For a scalar field, $\varphi'(x') = \varphi(x)$ holds, i.e. a scalar field transforms according to the one-dimensional trivial representation of the Lorentz group (and thus has/needs no further index for labeling a component.). For scalar fields the spin matrix vanishes.
- For a spinor field, $\xi'_\alpha(x') = M_\alpha{}^\beta\xi_\beta(x)$ holds, where ξ is a two-component complex field and $M \in \mathbf{SL}(2, \mathbb{C})$. Spinors transform according to a two-

[3] Again, the sign in front of Σ is a convention.

dimensional irreducible complex representation of the Lorentz group. Actually, as detailed below when calculating the matrix M, there is another inequivalent two-dimensional representation.

- For a contravariant vector field, $A'^{\mu}(x') = \Lambda^{\mu}{}_{\nu} A^{\nu}(x)$ holds, i.e. the field transforms according to the defining representation of the Lorentz group. Correspondingly, one defines the transformation of a covariant vector field. The spin matrix is found from equating $\omega^{\mu}{}_{\nu}$ with $(-i/2)\omega^{\rho\sigma}(\Sigma_{\rho\sigma})^{\mu}{}_{\nu}$, from which

$$(i\Sigma_{\rho\sigma})^{\mu}{}_{\nu} = \delta^{\mu}_{\sigma}\eta_{\rho\nu} - \delta^{\mu}_{\rho}\eta_{\sigma\nu}.$$

5.2.2 Representations of SL(2, $\mathbb{C}$)

Our starting point is the $\mathfrak{sl}(2, \mathbb{C})$ algebra

$$[M^{\mu\nu}, M^{\varrho\sigma}] = i(\eta^{\mu\sigma}M^{\nu\varrho} - \eta^{\nu\sigma}M^{\mu\varrho} + \eta^{\nu\varrho}M^{\mu\sigma} - \eta^{\mu\varrho}M^{\nu\sigma}).$$

It can be assumed that all central charges were brought to vanish by an appropriate modification of the generators. This is a specific case of the procedure sketched previously for the Poincaré algebra. For the Lorentz algebra, due to a theorem by Bargmann [27], it holds true also for the reason that it is semi-simple. Without loss of generality one can instead of the $M^{\mu\nu}$ directly take the (J^i, K^i) and their algebra

$$[J^j, J^k] = i\epsilon^{jkl}J^l \qquad [J^j, K^k] = i\epsilon^{jkl}K^l \qquad [K^j, K^k] = -i\epsilon^{jkl}J^l.$$

The so-called A- and B-spin operators defined as linear combinations of the (J^i, K^i)

$$\vec{A} := \frac{1}{2}(\vec{J} + i\vec{K}) \qquad \vec{B} := \frac{1}{2}(\vec{J} - i\vec{K}) \tag{5.19}$$

obey the algebra

$$\begin{aligned}
[A^j, A^k] &= i\epsilon^{jkl}A^l \\
[B^j, B^k] &= i\epsilon^{jkl}B^l \\
[A^j, B^k] &= 0.
\end{aligned}$$

The algebra $\mathfrak{sl}(2, \mathbb{C})$ is therefore isomorphic to two commuting $\mathfrak{su}(2)$ algebras.

Since the irreducible representations of $\mathfrak{su}(2)$ are $(2j+1)$-dimensional, the representations of $\mathfrak{sl}(2, \mathbb{C})$ are $(2A+1) \times (2B+1)$-dimensional. Quantum states can be built as eigenstates with respect to the Casimir operators[4] $(\vec{A}^2, \vec{B}^2)$ and the A^3- and B^3-components with the admissible values $(a = -A, -A+1, \ldots, A-1, A)$ and $(b = -B, -B+1, \ldots, B-1, B)$.

[4] Equivalently one could choose the Casimirs $(\vec{J}^2 - \vec{K}^2)$ and $\vec{J} \cdot \vec{K}$.

5.2.3 Field Variants

Scalar Fields

The trivial representation, denoted by $\mathbf{1} = (0, 0)$ stands for real and complex scalar fields. A prominent physical scalar field assumed to exist according to the SM is the Higgs field. However, the role of scalar fields on a fundamental level is unsettled, and their are attempts to understand the Higgs as a composite. Also in gravitational theories scalars appear, for example explicitly in Brans-Dicke theories or hidden in $f(R)$ models, see Sect. 7.6.4.

Spinor Fields

There are two different (non-isomorphic) two-dimensional representations, namely $\mathbf{2} = (\frac{1}{2}, 0)$ and $\mathbf{2}* = (0, \frac{1}{2})$, for reasons sketched below, also designated as $\mathbf{2}_L$ and $\mathbf{2}_R$. The associated fields are Weyl spinors, describing massless fermions and anti-fermions respectively. Up to the 1990' neutrinos were assumed to be massless. Thus they were believed to be properly pictured as Weyl spinors. Since further refined experiments hint at a finite mass, neutrinos are more appropriately described by Dirac- or Majorana-spinors, see Appendix B.1.4. Weyl spinors ξ_α have two components $(\alpha = 1, 2)$. For the case $(\dfrac{1}{2}, 0)$ the representations for the A-spin and the B-spin may be chosen as

$$D_{\mathbf{2}}(\vec{A}) = \frac{\vec{\sigma}}{2} \qquad\qquad D_{\mathbf{2}}(\vec{B}) = 0$$

where $\vec{\sigma}$ is the vector built from the three Pauli matrices (A.12). From this, due to (5.19),

$$D_{\mathbf{2}}(\vec{J}) = \frac{\vec{\sigma}}{2} \qquad\qquad D_{\mathbf{2}}(\vec{K}) = -i\frac{\vec{\sigma}}{2}.$$

For the case $(0, \frac{1}{2})$ we obtain correspondingly

$$D_{\mathbf{2}*}(\vec{J}) = \frac{\vec{\sigma}}{2} \qquad\qquad D_{\mathbf{2}*}(\vec{K}) = i\frac{\vec{\sigma}}{2}.$$

The unitary operator related to a Lorentz transformation $\Lambda = 1 + \lambda$, namely

$$U(\lambda) = \exp\left(-\frac{i}{2}\lambda_{\mu\nu}M^{\mu\nu}\right),$$

is rewritten in terms of $\vec{J}$ and $\vec{K}$ with parameters

$$(\lambda^{23}, \lambda^{31}, \lambda^{12}) := \vec{\theta}\,, \quad (\lambda^{01}, \lambda^{02}, \lambda^{03}) := \vec{\omega}$$

as

$$U(\lambda) = \exp(-i\vec{\theta}\cdot\vec{J} - i\vec{\omega}\cdot\vec{K}).$$

Thus the two different representations $D(\Lambda)$ of the Lorentz group on a two-component spin-($^1/_2$) field are given by

$$(M_{\mathbf{2}})_\alpha{}^\beta = \left[\exp\frac{i}{2}(\vec{\theta}\cdot\vec{\sigma} \; - \; i\vec{\omega}\cdot\vec{\sigma})\right]_\alpha^{\;\beta} \qquad \text{for} \qquad (\frac{1}{2}, 0)$$

$$(M_{\mathbf{2}*})_\alpha{}^\beta = \left[\exp\frac{i}{2}(\vec{\theta}\cdot\vec{\sigma} \; + \; i\vec{\omega}\cdot\vec{\sigma})\right]_\alpha^{\;\beta} \qquad \text{for} \qquad (0, \frac{1}{2}).$$

The unimodular 2×2 representation matrices $(M_{\mathbf{2}})$ and $(M_{\mathbf{2}^*})$ are elements of $\mathbf{SL}(\mathbf{2}, \mathbb{C})$, and–consistent with this attribute–described by six real parameters.

In order to distinguish which of the two non-equivalent representations one is dealing with, it is convenient to introduce two different kinds of spinors. From B. van der Waerden [520] originates a *dot notation*: ξ_α transforms according to the $\mathbf{2}$-representation and $\bar{\xi}_{\dot{\alpha}} := \xi^*_\alpha$ according to the $\mathbf{2}^*$-representation:

$$\xi'_\alpha(x') = M_\alpha^{\;\beta}\xi_\beta(x) \qquad\qquad \bar{\xi}'_{\dot{\alpha}}(x') = M^*{}_{\dot{\alpha}}^{\;\dot{\beta}}\bar{\xi}_{\dot{\beta}}(x).$$

In terms of the $\sigma^{\mu\nu}$ and $\bar{\sigma}^{\mu\nu}$-matrices (defined by the Pauli matrices according to (B.21)–the previous transformations can be written as

$$\xi'_\alpha = \left(I - \frac{1}{2}\lambda_{\mu\nu}\sigma^{\mu\nu}\right)_\alpha^{\;\beta}\xi_\beta \qquad\qquad \bar{\xi}'_{\dot{\alpha}} = \left(I - \frac{1}{2}\lambda_{\mu\nu}\bar{\sigma}^{\mu\nu}\right)_{\dot{\alpha}}^{\;\dot{\beta}}\bar{\xi}_{\dot{\beta}}.$$

The spinor calculus by van der Waerden allows–among other things–the construction of Lorentz tensors and (higher-rank spinors) from the two Weyl spinors as basic units; see Appendix B.1.3. In the appendix, also the four-component spinors traditionally employed in physics (Dirac and Majorana spinors) and their relation to the Weyl spinors are treated.

Vector- and Tensor Fields

Out of the $\mathbf{2}$- and the $\mathbf{2}^*$-representation of $\mathbf{SL}(\mathbf{2}, \mathbf{C})$ one obtains further representations by building products of representations and subsequent reduction, e.g.

$$\mathbf{2}\times\mathbf{2} : (0, \frac{1}{2}) \times (0, \frac{1}{2}) = (0, 0) \oplus (0, 1) = \mathbf{1}\oplus\mathbf{3}$$

$$\mathbf{2}^*\times\mathbf{2}^* : (\frac{1}{2}, 0) \times (\frac{1}{2}, 0) = (0, 0) \oplus (1, 0) = \mathbf{1}\oplus\mathbf{3}^*.$$

Both of these products lead to a singlet and a triplet representation. The singlet cleary belongs to a scalar field. The representation $(1,0) \oplus (0,1)$ has six independent elements and describes an antisymmetric tensor (with two indices). Furthermore

$$\mathbf{2} \times \mathbf{2}^* : \ (0, \frac{1}{2}) \times (\frac{1}{2}, 0) = (\frac{1}{2}, \frac{1}{2}) = \mathbf{4}.$$

The field related to this four-dimensional representation is a vector field and belongs to the defining representation of $\mathbf{SL}(\mathbf{2}, \mathbb{C})$.

By taking further products of the previous lower-dimensional representations, one is led to higher-spin fields. Of these only the spin-$^3/_2$ and the spin-2 fields have been seriously considered. They show up in supergravity (see Sect. 8.3.3).

5.2.4 *Quantum-Field Theoretical Symmetry Transformations*

Lorentz Transformations

Under Lorentz transformations $x'^\mu = \Lambda^\mu{}_\nu x^\nu$, the fields transform according to (5.16), that is

$$\Phi'_\alpha(x') = D(\Lambda)_\alpha{}^\beta \Phi_\beta(x). \tag{5.20}$$

In a quantum field theory, states $|\Psi\rangle$ transform with a unitary matrix (or an anti-unitary matrix in case of time reversal) as $|\Psi\rangle \to |\Psi'\rangle = U(\Lambda)|\Psi\rangle$. Therefore the classical relation (5.20) must be replaced by

$$\langle\Psi|U^{-1}(\Lambda)\Phi_\alpha(\Lambda x)U(\Lambda)|\Psi\rangle = D(\Lambda)_\alpha{}^\beta \langle\Psi|\Phi_\beta(x)|\Psi\rangle.$$

Since this relation must hold for arbitrary states, we obtain

$$U(\Lambda)\Phi_\alpha(x)U^{-1}(\Lambda) = D^{-1}(\Lambda)_\alpha{}^\beta \Phi_\beta(x'). \tag{5.21}$$

Let us explore the meaning of this for an infinitesimal Lorentz transformation $U(\Lambda) \simeq 1 - \frac{i}{2}\omega^{\mu\nu}M_{\mu\nu}$ with a representation $D^{-1}(\Lambda)_\alpha{}^\beta \simeq (1 + \frac{i}{2}\omega^{\mu\nu}\Sigma_{\mu\nu})_\alpha{}^\beta$. By expanding both sides of (5.21) in ω and disregarding terms $\mathcal{O}(\omega^2)$ we initially find

$$\Phi_\alpha(x) - \frac{i}{2}\omega^{\mu\nu}\left[M_{\mu\nu}\Phi_\alpha\right] = \Phi_\alpha(x') + \frac{i}{2}\omega^{\mu\nu}(\Sigma_{\mu\nu})_\alpha{}^\beta \Phi_\beta(x)$$

from which, with $\Phi(x') = \Phi(x^\lambda + \omega^\lambda{}_\nu x^\nu) = \Phi(x) - (\partial_\lambda\Phi)\omega^{\lambda\nu}x_\nu = \Phi(x) + \frac{i}{2}\omega^{\mu\nu}L_{\mu\nu}\Phi(x)$,

$$\bar{\delta}\Phi_\alpha = \frac{i}{2}[\omega^{\mu\nu}M_{\mu\nu}, \Phi_\alpha],$$

where the last expression follows from (5.18).

"Internal" Symmetries

The Noether theorem also holds for an internal/local symmetry group **G**, where **G** is an N-dimensional Lie group. In this case, the transformation does not affect the coordinates, but acts on the symmetry group indices. The classical transformation

$$\Phi'_\alpha(x) = D(g)_\alpha{}^\beta \Phi_\beta(x)$$

in field theory, in analogy with (5.21) becomes

$$U(g)\Phi_\alpha(x)U^{-1}(g) = D^{-1}(g)_\alpha{}^\beta \Phi_\beta(x), \tag{5.22}$$

and particularly a group scalar transforms as $U(g)\Phi_\alpha(x)U^{-1} = \Phi_\alpha(x)$. For $U(g) = \exp(i\sum_a \theta^a C^a)$, we derive, quite analogously to the previous calculation for the Lorentz group,

$$\bar{\delta}\Phi_\alpha = i[\theta^a C_a, \Phi_\alpha].$$

This is the quantum (field) analogue to the classical Poisson bracket relation (3.51).

5.3 Actions

This section deals with the astounding fact that the contingent action functionals in a relativistic quantum field theory are rather restricted, the essential constraints arising from symmetry and renormalization requirements.

5.3.1 Requirements on a QFT Action

The action functionals we are dealing with are of the form

$$S[\varphi] = \int_\Omega d^4x\, \mathcal{L}(\Phi_\alpha(x), \partial_\mu \Phi_\alpha(x)).$$

It obeys the postulate of locality, or–next neighbor interaction, since the Lagrangian $\mathcal{L}$ depends only on the field at a certain point x and the derivative of the field at the same point. In order to avoid possible interpretational and/or technical problems (solution instabilities, states with negative norm/ghosts, tachyon states/superluminal propagators), it is also assumed that at most first derivatives occur in the Lagrangian. (This actually is too strong: second-order derivatives are allowed if they behave as boundary terms in the action. This is typically the case for gravitational theories.) There are further requirements:

- $\mathcal{L}$ is classically real (in order to yield a real-valued energy) and Hermitean in QFT (in order to guarantee a stable time evolution).

- $\mathcal{L}$ is Poincaré invariant, that is $\mathcal{L}$ does not depend on the coordinates explicitly and it transforms like a scalar with respect to Lorentz transformations. Later we will see that in presence of a gravitational field, the action must be a scalar with respect to general coordinate transformations.
- The canonical Hamiltonian must be positive definite, in order to ensure the interpretation of the numerical value as the energy.
- If one requires specific discrete symmetries (e.g. P, C, T-invariance) or symmetries with respect to internal symmetries this leads to further requirements on the Lagrangian.
- renormalizability.

Renormalizability Criterium

Renormalization is the endeavor to tame divergences in a quantum field theory. These divergence issues plagued quantum field theory from its beginning. Since this chapter is not a full-fledged exposition of relativistic quantum field theories, the description of renormalizability here can only be informal.

Quantum field theories are formulated with local field operators, and therefore distances can become arbitrarily small. This leads to divergent integrals in momentum space, integrals typically arising in the calculation of radiation correction via loop diagrams. A theory is called renormalizable if all divergences can be removed by the redefinition of a finite number of physical constants in the theory. So for instance renormalization in QED (quantum electrodynamic) leads to a redefinition of the bare mass and the bare charge of an electron, and yields the experimentally measured mass and charge. If instead in this process infinitely many quantities appear, the theory is non-renormalizable, which was at first understood as being "of no predictive power".

A condition of renormalizability originates from power counting in terms of the so-called mass dimension: In the unit system with $c = [L/T] = 1$ and $\hbar = [E \cdot T] = 1$, length and time have the same dimension, namely the inverse of the energy. Because of $E = mc^2$ (for particles at rest) the energy has the same dimension as the mass; or

$$[E] = 1, \ \ [L] = [T] = -1.$$

The dimension of a derivative is $[\partial_\mu] = +1$, and the dimension of the volume element is $[d^D x] = -D$. The action is dimensionless (remember $\hbar = 1$), and thus the Lagrange density has $[\mathcal{L}] = D$. Now the Lagrangian is built from kinetic terms of the fields and from interaction terms. The kinetic terms introduce a mass dimension of the derivatives and allow to determine the mass dimensions of the fields. The interaction terms contain coupling constants. Their mass dimension is in turn defined by the mass dimensions of the fields.

Now a necessary renormalization condition can be formulated in terms of the *superficial degree of divergence* Δ_I which for each interaction term in the

Lagrangian must not become negative:

$$\Delta_I := 4 - d_I - \sum_f n_{If}(s_f + 1) \geqslant 0. \tag{5.23}$$

Here d_I is the number of derivatives in the interaction term, n_{If} the number of fields of type f, and $s_f = [0, 1/2, 1, 0]$ for [scalars, fermions, massive vector fields, and photons and gravitons], respectively. For a derivation of (5.23) see Chap. 16 of [536], where Weinberg remarks, "For renormalization to work in such theories, it is also usually necessary that all renormalizable interactions that are allowed by symmetry principles should actually appear in the Lagrangian."

We will see that together with symmetry arguments, the class of actions for free and interacting scalar, spinor, and vector fields restricted by (5.23) is rather limited. Nevertheless, due to a "change in attitude" of the conception of renormalizability, the power-counting argument is predominantly understood to lead to the lowest-order terms in a series of (even non-renormalizable) contributions to a Lagrangian. More about this concept of effective field theories will be given in Sect. 5.6.

5.3.2 Scalar Fields

The One-Scalar World

The most simple universe (although we certainly would not exist in this world) is described by just one real scalar field. The Lagrangian can be built in terms of powers of the field itself, and terms with derivatives, which because of the requested Poincaré invariance can only come in powers of $\partial_\mu\varphi\partial^\mu\varphi$. The simplest Lagrangian is

$$\mathcal{L} = +\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - V(\varphi).$$

The first term is of course interpreted as a kinetic term, the second is the potential. The plus sign in front of the kinetic term is chosen such that the energy is positive in the metric convention used in this book. Indeed, the canonical momentum is $p_\varphi = \partial\mathcal{L}/\partial_0\varphi$ and the canonical Hamiltonian becomes

$$\mathcal{H} = p_\varphi\partial_0\varphi - \mathcal{L} = \frac{1}{2}(p_\varphi^2 + \partial_j\varphi\partial_j\varphi) + V(\varphi).$$

The factor $1/2$ in front of the kinetic term is chosen for convenience. A different choice can be reabsorbed in a scaling of the field. In D=4 we derive from $[\mathcal{L}] = 4$ that $[\partial_\mu\varphi] = 2$ and $[\varphi] = 1$. One can think of expanding the potential as a Taylor series. The first constant term is irrelevant for the field equations, a term linear in φ can be removed by a shift of the scalar field, and powers of the field higher than four are not allowed due the quest for renormalizability. Therefore the most general potential is

$$V(\varphi) = \frac{1}{2}a\varphi^2 + \frac{1}{3!}g\varphi^3 + \frac{1}{4!}\lambda\varphi^4$$

where the constants have mass dimensions $[a] = 2,\ [g] = 1, [\lambda] = 0$. This complies with (5.23), which for an interaction term φ^5 would lead to $\Delta = -1$. For $g = 0$ the Lagrangian defines the $\lambda\varphi^4$ theory with the field equation

$$\Box\varphi + a\varphi + \frac{1}{3!}\lambda\varphi^3 = 0 \qquad (\Box := \partial_\nu\partial^\nu).$$

If $a > 0$ we may set $a = m^2$. Specifically for $\lambda = 0$ this is the Klein-Gordon equation

$$(\Box + m^2)\varphi = 0, \tag{5.24}$$

describing a spinless neutral particle with mass m. Interpreting m as a mass is justified in that according to the quantization prescription $\partial_\mu \to ip_\mu$, we obtain $\Box \to -p^2$, and thus the Klein-Gordon equation entails the light-cone relation $(p^2 - m^2)\varphi = 0$ on states. For $\lambda \neq 0$ the field φ is coupled to itself.

At this stage, a remark on field redefinition is appropriate: If you decide to call your scalar field φ and work with the Klein-Gordon Lagrangian as discussed above, someone else might call his scalar field ϕ. Let $\phi = \varphi + F(\varphi)$, then your Klein-Gordon Lagrangian becomes in the other variables

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi(1 + 2\,F' + F'^2) - \frac{1}{2}\,m^2(\varphi^2 + 2\varphi F + F^2)$$

with $F' = dF/d\varphi$. Compared to your theory his looks rather weird (and even mimics an interacting field theory) but, nevertheless, if $\phi(\varphi)$ is an invertible function, we expect that observations derived in one formulation are the same as those derived in the other. Indeed this turns out to be true, as is most easily seen in the path integral formulation, where a change from φ to ϕ is just a change in integration variables.

Quantization of the scalar field equation starts with the quantization of the Klein-Gordon field (the "free" field) and a subsequent perturbation in the parameter λ. A standard procedure (the so-called second quantization) starts from the Fourier expansion of free fields in terms of plane waves and the identification of creation and annihilation operators.

Free-Field Solution

Plane waves are characterized by the wave vector k in

$$k \cdot x = k_\mu x^\mu = (k_0 t - \vec{k} \cdot \vec{x}).$$

In momentum space the Klein-Gordon operator $(\Box + m^2)$ becomes $(k^2 - m^2)$ and thus a possible approach for φ is

$$\varphi(x) = (2\pi)^{-3/2} \int d^4k\, \delta(k^2 - m^2)\Theta(k_0) \left\{ A(k)e^{-ik\cdot x} + A^\dagger(k)e^{ik\cdot x} \right\}.$$

Here the delta function serves to make this a solution of the Klein-Gordon equation, and the presence of the step function Θ singles out positive energy solutions ($k_0 > 0$). In a next step, the expression is simplified by carrying out the k_0 integration. For this purpose one uses the rule $\delta(f(x)) = |f'(a)|^{-1}\delta(x - a)$ which holds for a function $f(x)$ with $f(a) = 0$. Because of the two roots of $k^2 - m^2 = 0$ we have

$$\delta(k^2 - m^2) = \frac{\delta(k^0 - \sqrt{\vec{k}^2 + m^2})}{2k^0} + \frac{\delta(k^0 + \sqrt{\vec{k}^2 + m^2})}{2|k^0|}.$$

Then k_0 can be integrated out of

$$\int d^4k\, \delta(k^2 - m^2)\Theta(k_0) = \int d^3k \int_0^\infty \frac{dk^0}{2k^0}\delta(k^0 - \sqrt{\vec{k}^2 + m^2}) = \int \frac{d^3k}{2\omega_k}$$

with

$$\omega_k = +\sqrt{\vec{k}^2 + m^2}. \tag{5.25}$$

Therefore the Fourier expansion of the scalar field φ and its conjugate momentum $\pi = \partial_0\varphi$ becomes, with $A(k) = \sqrt{2\omega_k}\, a(k)$,

$$\varphi(\vec{x}, t) = (2\pi)^{-3/2} \int \frac{d^3k}{\sqrt{2\omega_k}} \left\{ a(\vec{k})e^{-i(\omega_k t - \vec{k}\cdot\vec{x})} + a^\dagger(\vec{k})e^{i(\omega_k t - \vec{k}\cdot\vec{x})} \right\} \tag{5.26}$$

$$\pi(\vec{x}, t) = (2\pi)^{-3/2} \int \frac{d^3k}{\sqrt{2\omega_k}}(i\omega_k) \left\{ -a(\vec{k})e^{-i(\omega_k t - \vec{k}\cdot\vec{x})} + a^\dagger(\vec{k})e^{i(\omega_k t - \vec{k}\cdot\vec{x})} \right\}. \tag{5.27}$$

The canonical commutators of the fields and their momenta

$$[\varphi(\vec{x}, t), \pi(\vec{x}\,', t)] = [\varphi(\vec{x}, t), \partial_0\varphi(\vec{x}\,', t)] = i\delta^3(\vec{x} - \vec{x}\,')$$

imply for the coefficients $a(\vec{k})$ and $a^\dagger(\vec{k})$ (in the quantum theory interpreted as annihilation and creation operators) that

$$[a(\vec{k}), a^\dagger(\vec{k}')] = \delta^3(\vec{k} - \vec{k}').$$

Before going on with the Klein-Gordon field quantization, I digress briefly by reminding you of the notion of creation and annihilation operators in quantum mechanics.

Creation and Annihilation Operators

Consider the one-particle Lagrangian $L = \dot{q}^2/2 - V(q)$, where the mass has been set equal to 1. With the momentum $p = \dot{q}$ the Hamiltonian becomes

$H = p^2/2 + V(q)$. In the Heisenberg picture $q(t)$ and $p(t)$ are operators obeying the canonical commutation relation[5]

$$[q, p] = i\hbar.$$

In the following also $\hbar = 1$. The position and the momentum operators evolve in time according to

$$\frac{dq}{dt} = i[H, q] = p \qquad \frac{dp}{dt} = i[H, p] = -V'(q).$$

Instead of the pair (q, p) which are Hermitean operators, use operators $(a, a^\dagger)$

$$q = C(a + a^\dagger) \qquad p = -\frac{i}{2C}(a - a^\dagger)$$

with a constant C. The $(a, a^\dagger)$ obey the commutation relation

$$[a, a^\dagger] = 1.$$

Let us further specifically treat the harmonic oscillator with $V(q) = (1/2)\omega^2 q^2$. Choosing $C^2 = 1/2\omega$, the Hamiltonian assumes the form

$$H = (a^\dagger a + \frac{1}{2})\omega. \tag{5.28}$$

The eigenvalue of the operator $(a^\dagger a)$ cannot be negative, since for an arbitrary state $|\lambda\rangle$ the expectation value $\langle\lambda|a^\dagger a|\lambda\rangle$ is just the norm of $a|\lambda\rangle$. The lowest eigenvalue of $a^\dagger a$ is zero, and for its eigenstate $|0\rangle$ (called the ground state),

$$a|0\rangle = 0 \qquad \langle 0|a^\dagger = 0$$

must hold, together with $\langle 0| = |0\rangle^\dagger$ and $\langle 0|0\rangle = 1$. By successively applying the *creation operator* $a^\dagger$ to the ground state one generates the full spectrum of quantum states

$$|n\rangle = (n!)^{-\frac{1}{2}}(a^\dagger)^n|0\rangle.$$

These are normalized ($\langle n|n\rangle = 1$) and orthogonal to the states $\langle n| = |n\rangle^\dagger = (n!)^{-\frac{1}{2}}\langle 0|a^n$. Furthermore, the states $|n\rangle$ are eigenstates of the Hamiltonian and obey the eigenvalue equation

$$H|n\rangle = \omega\left(n + \frac{1}{2}\right)|n\rangle.$$

---

[5] In this Chapter I completely suppress the "hats" for operators. The context should make clear of whether an object is an operator or not.

The previous procedure can immediately be generalized to systems with more than one degree of freedom, where one introduces a finite set of creation and annihilation operators $\{a_i^\dagger, a_i\}$, and to field theories with a continuum of creation and annihilation operators $a^\dagger(\vec{k})$ and $a(\vec{k})$, which as we saw in the previous subsection depend on the wave vector in a Fourier expansion of the fields in question.

Quantization of Klein-Gordon Theory

The Hamiltonian of the Klein-Gordon theory becomes in terms of the operators $a^\dagger(\vec{k})$ and $a(\vec{k})$

$$H = \frac{1}{2}\int d^3x\, \left(\pi^2 + \partial_i\varphi\partial_i\varphi + m^2\varphi^2\right) = \int d^3k\, \omega_k\, \left(a^\dagger(k)a(k) + \frac{1}{2}\right),$$

an expression with a striking analogy to (5.28). However, in contrast to quantum mechanics, this Hamiltonian is not finite because of the presence of the $(1/2)$ in the integrand. This term can be dropped, however, if one opts for "normal ordering" of operators: Consider the product of two operators ϕ and ψ and write it as $(\phi^+ + \phi^-)(\psi^+ + \psi^-)$, where the (+) respectively (-) represent the annihilation and the creation part. The normally-ordered product is defined as

$$:\phi\psi := \phi^+\psi^+ + \phi^-\psi^+ + \phi^-\psi^- + \psi^-\phi^+$$

that is all creation(annihilation) parts are on the left(right). The Hamiltonian is thus

$$H = \frac{1}{2}\int d^3x\, :\left(\pi^2 + \partial_i\varphi\partial_i\varphi + m^2\varphi^2\right) := \frac{1}{2}\int d^3k\, \omega_k\, a^\dagger(k)a(k).$$

Similar to the quantum mechanical case the vacuum (or ground state) $|0\rangle$ is defined by $a(\vec{k})|0\rangle = 0$ for all $\vec{k}$. A single- particle state is created from the vacuum as $|\vec{k}\rangle = a^\dagger(\vec{k})|0\rangle$. Therefore

$$\begin{aligned}\langle 0|\varphi(\vec{x},t)|\vec{k}\rangle &= \langle 0|(2\pi)^{-3/2}\int \frac{d^3k'}{\sqrt{2\omega_{k'}}}\left\{a(\vec{k}')e^{-i(\omega_{k'}t-\vec{k}'\cdot\vec{x})}\right\}a^\dagger(\vec{k})|0\rangle \\ &= (2\pi)^{-3/2}(2\omega_k)^{-1/2}e^{-i(\omega_k t-\vec{k}\cdot\vec{x})}\end{aligned}$$

can be interpreted as the relativistic wave function of a single particle with momentum $\vec{k}$. Further, one defines the N-particle states

$$|k_1, k_2, \ldots, k_N\rangle = a^\dagger(k_1)a^\dagger(k_2)\ldots a^\dagger(k_N)|0\rangle.$$

Many-Scalar Worlds

If there is a collection of scalar fields $\{\varphi_i\}$ $(i = 1, \ldots, N)$ we may start with the free-field Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi_i\partial^\mu\varphi_i - \frac{1}{2}M^2_{ij}\varphi_i\varphi_j$$

with a real symmetric matrix M. This Lagrangian can be diagonalized by introducing mass eigenstates $\hat{\varphi}_i$ using an orthogonal matrix $\bar{O}$ such that

$$\hat{\varphi}_i = \bar{O}_{ij}\varphi_j \qquad\qquad M^2_{ij}\bar{O}_{ik}\bar{O}_{jl} = m^2_k\delta_{kl} \quad \text{(no sum over k)}.$$

Then the Lagrangian is written as

$$\mathcal{L} = \frac{1}{2}\partial_\mu\hat{\varphi}_i\partial^\mu\hat{\varphi}_i - \frac{1}{2}\, m^2_i\hat{\varphi}_i\hat{\varphi}_i,$$

which represents a system of decoupled scalar fields with real masses m^2_i. If all masses are the same, the Lagrangian is invariant with respect to an $\mathbf{O}(\mathbf{N})$ symmetry, and if $K < N$ masses are degenerate this reduces to an $\mathbf{O}(\mathbf{K})$ symmetry.

In the case of a complex (or better: Hermitean) scalar field we have to ensure that the action is a real Lorentz scalar. The analogue of the Klein-Gordon Lagrangian becomes

$$\mathcal{L} = (\partial^\mu\varphi^\dagger)(\partial_\mu\varphi) - m^2(\varphi^\dagger\varphi). \tag{5.29}$$

Notice that in writing φ in terms of two real fields $(\varphi = 1/\sqrt{2}(\varphi_1 + i\varphi_2))$ this Lagrangian becomes identical to an $\mathbf{O}(\mathbf{2})$ Lagrangian. The field equations from (5.29) are $(\Box + m^2)\varphi = 0 = (\Box + m^2)\varphi^\dagger$, and are solved by

$$\varphi(\vec{x}, t) = (2\pi)^{-3/2}\int \frac{d^3k}{\sqrt{2\omega_k}}\left\{a(\vec{k})e^{-i(\omega_k t - \vec{k}\cdot\vec{x})} + b^\dagger(\vec{k})e^{i(\omega_k t - \vec{k}\cdot\vec{x})}\right\} \tag{5.30}$$

if ω_k is related to $\vec{k}$ by (5.25). Observe that unlike the real case the Fourier expansion requires having two independent variants of annihilation and creation operators, namely $(a, a^\dagger)$, $(b, b^\dagger)$. Correspondingly there are two different one particle states: $|\vec{p}\rangle := a^\dagger(\vec{p})|0\rangle$ and $|\vec{q}\rangle := b^\dagger(\vec{q})|0\rangle$. We will see later that they have opposite charges; therefore they might reasonably be called "particle" and "anti-particle".

After expressing the "velocities" by the momenta, that is

$$\pi := \frac{\partial\mathcal{L}}{\partial_0\varphi} = \partial_0\varphi^\dagger \qquad\qquad \pi^\dagger := \frac{\partial\mathcal{L}}{\partial_0\varphi^\dagger} = \partial_0\varphi$$

we calculate from the canonical commutation relations

$$[\varphi(\vec{x}, t), \pi(\vec{x}', t)] = [\varphi(\vec{x}, t), \partial_0\varphi^\dagger(\vec{x}', t)] = i\delta^3(\vec{x} - \vec{x}')$$

(and a similar one for $[\varphi^\dagger, \pi^\dagger]$) the commutators $[a(\vec{k}), a^\dagger(\vec{k}')] = \delta^3(\vec{k} - \vec{k}')$ and $[b(\vec{k}), b^\dagger(\vec{k}')] = \delta^3(\vec{k} - \vec{k}')$; all others vanish.

The Lagrangian (5.29) has a global $\mathbf{U}(\mathbf{1})$ symmetry

$$\varphi(x) \to e^{i\theta}\varphi(x) \qquad\qquad \varphi^\dagger(x) \to e^{-i\theta}\varphi^\dagger(x)$$

leading to a Noether current

$$j_\mu = i(\varphi^\dagger \partial_\mu \varphi - \partial_\mu \varphi^\dagger \varphi)$$

with an associated Noether charge

$$C = i \int d^3x (\varphi^\dagger \partial_0 \varphi - \partial_0 \varphi^\dagger \varphi) = \int d^3k \{a(\vec{k})^\dagger a(\vec{k}) - b(\vec{k})^\dagger b(\vec{k})\}.$$

Now consider the action of C on a "particle" and on an "anti-particle" state:

$$C\, a^\dagger(\vec{p})|0\rangle = \int d^3k\ a(\vec{k})^\dagger a(\vec{k}) a^\dagger(\vec{p})|0\rangle = +a^\dagger(\vec{p})|0\rangle$$

$$C\, b^\dagger(\vec{q})|0\rangle = -\int d^3k\ b(\vec{k})^\dagger b(\vec{k}) b^\dagger(\vec{q})|0\rangle = -b^\dagger(\vec{q})|0\rangle.$$

It makes sense to identify C with the electric charge, as we also will see when the complex scalar field is coupled to the electromagnetic field.[6] With this identification it becomes clear that a real scalar field represents a neutral particle (indeed C is identical to zero in this case). Furthermore, we are permitted to interpret $\varphi^\dagger$ creating a particle and annihilating an anti-particle. On the other hand, φ creates an anti-particle and annihilates a particle.

Finally, in case of various complex fields given by the Lagrangian

$$\mathcal{L} = \partial^\mu \varphi_i^\dagger \partial_\mu \varphi_i - M_{ij}^2 \varphi_i^\dagger \varphi_j$$

one might (in analogy with the real case) diagonalize the Hermitean matrix M by redefining the states by a unitary matrix U as $\hat{\varphi}_i = U_{ij}\varphi_j$.

5.3.3 *Spinor Actions*

In order to construct Lagrangians from spinors, we have as basic building blocks Weyl spinors at our disposal. We start by investigating a world with only one spinor, and then allow for sets of Weyl spinors. Already in the case with two spinors, we will see that various actions can be formed. These differ in their mass terms and exhibit distinguished symmetries.

---

[6] This interpretation is due to W. Pauli and V. Weisskopf [408]. Their work not only showed that the particle/anti-particle notion is inherent to quantum field theory–and not bound to exist only for electrons and positrons–but also explained why previous unsuccessful attempts to interpret the zero component of the Noether current as a probability density had failed.

Weyl Spinor Action

Let us separately investigate possible kinetic and mass terms.

Any action needs a kinetic term containing derivatives of the fields. Here, at first the gradient ∂_μ–considered as a Lorentz vector–is to be expressed by spinors. According to (B.24), there are two options

$$\partial_\mu \to \partial^{\dot{\alpha}\beta} := (\bar{\sigma}^\mu)^{\dot{\alpha}\beta}\partial_\mu \qquad \text{or} \qquad \partial_\mu \to \partial_{\alpha\dot{\beta}} := (\sigma^\mu)_{\alpha\dot{\beta}}\partial_\mu. \tag{5.31}$$

The only Lorentz scalar that can be built with the first option is

$$\bar{\xi}_{\dot{\alpha}}(\bar{\sigma}^\mu)^{\dot{\alpha}\beta}\partial_\mu\xi_\beta = \bar{\xi}\bar{\sigma}^\mu\partial_\mu\xi$$

where the expression on the right-hand-side is to be understood from the summation conventions defined in (B.15) and (B.18). In general, this term is not real. If one adds the Hermitean conjugate $\partial_\mu\bar{\xi}\bar{\sigma}^\mu\xi$, the sum is a total derivative $\partial_\mu(\bar{\xi}\bar{\sigma}^\mu\xi)$ and thus not qualified for an action. Therefore, we subtract the two terms and multiply by the imaginary unit i. After all, a possible kinetic term in a Lagrangian is

$$\mathcal{L}_L^K := \frac{i}{2}(\bar{\xi}\bar{\sigma}^\mu\partial_\mu\xi - \partial_\mu\bar{\xi}\bar{\sigma}^\mu\xi) = i\bar{\xi}\bar{\sigma}^\mu\partial_\mu\xi + b.t. \tag{5.32}$$

Similarly the second option in (5.31) results in a kinetic term built from $\eta\sigma^\mu\partial_\mu\bar{\eta}$ as

$$\mathcal{L}_R^K := \frac{i}{2}(\eta\sigma^\mu\partial_\mu\bar{\eta} - \partial_\mu\eta\sigma^\mu\bar{\eta}) = i\eta\sigma^\mu\partial_\mu\bar{\eta} + b.t.$$

In the following, only the first option is explored; the arguments can easily be transferred to the other one.

Power counting reveals that in $D = 4$ the spinor fields have mass dimension $[\xi] = \frac{3}{2}$. Therefore kinetic terms with higher than first derivatives of the spinor fields are to be ruled out by dimensional renormalizability.

For constructing mass terms, we have at our disposal the Lorentz scalars $\xi^2 = \xi^\alpha\xi_\alpha$ and $\bar{\xi}^2 = \bar{\xi}_{\dot{\alpha}}\bar{\xi}^{\dot{\alpha}}$. Since the mass dimension of these terms is $2 \times (3/2) = 3$, each of them, multiplied by a constant with dimension of mass, can serve as a term contributing to the mass term in the Lagrangian. If we further require the Lagrangian to be a real scalar, we obtain

$$\mathcal{L}_W = i\bar{\xi}\bar{\sigma}^\mu\partial_\mu\xi - \frac{1}{2}m_\xi(\xi^2 + \bar{\xi}^2) \tag{5.33}$$

The Euler derivatives of this Weyl Lagrangian are

$$[\mathcal{L}_W]_{\bar{\xi}} = \frac{\partial\mathcal{L}_W}{\partial\bar{\xi}} = i\bar{\sigma}^\mu\partial_\mu\xi - m_\xi\bar{\xi}$$

$$[\mathcal{L}_W]_\xi = \frac{\partial\mathcal{L}_W}{\partial\xi} - \partial_\mu\frac{\partial\mathcal{L}_W}{\partial(\partial_\mu\xi)} = -m_\xi\xi - i(\partial_\mu\bar{\xi})\bar{\sigma}^\mu.$$

Notice that the second Euler derivative is the conjugate to the first one. At a first glance the Euler derivatives with respect to ξ and to $\bar{\xi}$ appear asymmetric. But remember that we could (and should) add the Hermitean conjugate to the Weyl Lagrangian, which up to a boundary term is the same as (5.33). The spinor field ξ stands for a neutral particle which either is massless (called a *massless Weyl fermion*) or massive (*Majorana fermion*).[7]

If the mass $m_\xi = 0$, the Lagrangian (5.33) is invariant with respect to the global $\mathbf{U(1)}$ transformation $\xi' = e^{i\theta}\xi$, leading to the Noether current $j^\mu = -\bar{\xi}\bar{\sigma}^\mu\xi$.

Mass Diagonalization

When the previous action building is extended beyond one Weyl spinor, we are allowed to consider other variants of mass terms. In the case of just two independent spinors ξ and η the condition of dimensional renormalizability allows a Lagrangian

$$\mathcal{L}_{\xi\eta} = i\bar{\xi}\bar{\sigma}^\mu\partial_\mu\xi + i\bar{\eta}\bar{\sigma}^\mu\partial_\mu\eta - \frac{1}{2}m_\xi(\xi^2 + \bar{\xi}^2) - \frac{1}{2}m_\eta(\eta^2 + \bar{\eta}^2) - \frac{1}{2}m_D(\xi\eta + \bar{\eta}\bar{\xi}).$$

We can now distinguish various symmetries subject to the masses (m_ξ, m_η, m_D). If all masses vanish, the Lagrangian is invariant with respect to transformations

$$\xi' = e^{i\theta}\xi, \qquad \eta' = e^{i\vartheta}\eta, \tag{5.34}$$

mediated by $\mathbf{U(1)}_\xi \times \mathbf{U(1)}_\eta$. This symmetry leads to two Noether currents and associated conserved charges. Of course this case represents just two decoupled Weyl fermions. A mass term with non-vanishing m_D leaves a lesser symmetry intact, namely

$$\xi' = e^{i\theta}\xi, \qquad \eta' = e^{-i\theta}\eta, \tag{5.35}$$

known as vector symmetry, $\mathbf{U(1)_V}$. It entails a conserved charge C_V. Similar to the case of complex scalar fields, one shows that the fields carry the charge $+1$, and the complex conjugates the charge -1. The fields ξ and η are interpreted as two-component Dirac fields. Another case arises for $m_\xi = m = m_\eta$ (and assuming $m_D = 0$). Renaming the two spinors as $\xi = \xi_1, \eta = \xi_2$, there is an additional global $\mathbf{O(2)}$ symmetry:

$$\xi_i \to O_i^{\ j}\xi_j$$

with a matrix obeying $O^T O = I$. To this symmetry, a conserved Hermitean Noether current belongs:

$$j^\mu = i(\bar{\xi}_1\bar{\sigma}^\mu\xi_2 - \bar{\xi}_2\bar{\sigma}^\mu)\xi_1.$$

[7] These are two-component fields and should not be confused with the four-component spinors introduced in the next subsection.

It is instructive to go to a new basis using

$$\chi = \frac{1}{\sqrt{2}}(\xi + i\eta) \qquad\qquad \omega = \frac{1}{\sqrt{2}}(\xi - i\eta).$$

Then the Lagrangian becomes

$$\mathcal{L}_{\xi\eta} \quad \Longrightarrow \quad i\bar{\chi}\bar{\sigma}^\mu\partial_\mu\chi + i\bar{\omega}\bar{\sigma}^\mu\partial_\mu\omega - m(\chi\omega + \bar{\omega}\bar{\chi})$$

mimicking two Dirac fields. The $\mathbf{O(2)}$ symmetry in the (ξ, η)-fields is now realized as an $\mathbf{U(1)}$ symmetry $\chi \to e^{i\Theta}\chi$, $\omega \to e^{-i\Theta}\omega$. The current is diagonal in the new variables:

$$j^\mu = \bar{\chi}\bar{\sigma}^\mu\chi - \bar{\omega}\bar{\sigma}^\mu\omega$$

showing that the spinorial fields χ and ω are eigenstates of a charge operator Q with eigenvalues ± 1.

These findings can be generalized to the case of more than two spinor fields and it can be shown (see e.g. [137]) that a mass diagonalization results in a set of massless Weyl fields, a set of neutral Majorana fields, and a set of massive charged Dirac fields.

Dirac and Majorana Fermions

In physics, spinor fields made their first appearance as four-component objects in the Dirac equation, interpreted to describe electrons and their anti-particles (positrons). Mathematically, they correspond to the representation $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$. These Dirac spinors, although belonging to a reducible representation, enter the stage since space inversions P transform the $\mathbf{2}$-representation into the $\mathbf{2^*}$-representation, and *vice versa*, and since space inversion is a valid symmetry (aside from certain weak interaction processes; see Sect. 6.3).

A Dirac spinor is built from two Weyl spinors in the form

$$\psi(x) := \begin{pmatrix} \xi_\alpha \\ \bar{\eta}^{\dot{\alpha}} \end{pmatrix} \doteq \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix}. \tag{5.36}$$

Thus a Dirac spinor is a spinorial object with four components. Its field equations can be derived from a Lagrangian

$$\mathcal{L}_D = i\eta\sigma^\mu\partial_\mu\bar{\eta} + i\bar{\xi}\bar{\sigma}^\mu\partial_\mu\xi - \frac{1}{2}m_D(\xi\eta + \bar{\eta}\bar{\xi}).$$

Defining the conjugate Dirac spinor as in (B.30) by $\overline{\psi} = (\bar{\xi}_{\dot{\alpha}}\ \eta^{\alpha})$ and the Dirac matrices as (compare (B.12))

$$\gamma^{\mu} = \begin{pmatrix} 0 & \sigma^{\mu} \\ \bar{\sigma}^{\mu} & 0 \end{pmatrix}$$

the Lagrangian becomes

$$\mathcal{L}_D = \overline{\psi}(i\gamma^{\mu}\partial_{\mu} - m_D)\psi = \overline{\psi}(i\not{\partial} - m_D)\psi. \tag{5.37}$$

Here, I have introduced the Feynman "slash"-notation: For any four-vector u_{μ}: $\not{u} := \gamma^{\mu}u_{\mu}$. The Lagrangian (5.37) is invariant under the transformations $\psi \to e^{i\theta\psi}$, $\overline{\psi} \to e^{-i\theta}\bar{\psi}$, giving rise to a conserved Noether current

$$j^{\mu} = -\overline{\psi}\gamma^{\mu}\psi. \tag{5.38}$$

The Euler-Lagrange equations for of the Lagrangian (5.37) are the Dirac equations

$$(i\not{\partial} - m_D)\psi = 0. \tag{5.39}$$

Acting on the Dirac equation with $(i\gamma^{\mu}\partial_{\mu} + m)$, we derive

$$(i\gamma^{\mu}\partial_{\mu} - m)(i\gamma^{\mu}\partial_{\mu} + m)\psi = (-i\gamma^{\mu}\gamma^{\nu}\partial_{\mu}\partial_{\nu} - m^2)\psi = -(\Box + m^2)\psi = 0$$

where use has been made of the relation $\gamma^{\mu}\gamma^{\nu} + \gamma^{\nu}\gamma^{\mu} = 2\, g^{\mu\nu}$, defining the gamma matrices. Thus each of the Dirac spinor components obeys the Klein-Gordon equation.

Notice that the Dirac equation was derived from postulates of special relativity and quantum mechanics. Perhaps here is the place to recapitulate the story of how to get from the Schrödinger to the Dirac equation as it is told in elementary treatises: The free Schrödinger equation $i\hbar\dot{\psi} = \hat{H}\psi = \frac{\hbar^2}{2m}\triangle$ is obviously incompatible with special relativity, since time is treated differently from space. In observing that the Hamiltonian operator derives from the Newtonian energy expression as $\hat{H} = \frac{\hat{p}^2}{2m} \Rightarrow \frac{\hbar^2}{2m}\triangle$ the initial idea is to replace the Newtonian energy expression by the relativistic one (3.19)

$$\hat{H}^2 = \hat{p}^2c^2 + m^2c^4 \Rightarrow c^2\hbar^2\triangle + m^2c^4.$$

The quadratic Hamiltonian arises in the time derivative of the Schrödinger equation: $i\hbar\ddot{\psi} = \hat{H}\dot{\psi} = \frac{1}{i\hbar}\hat{H}^2\psi$. Inserting the relativistic expression for the Hamiltonian operator one arrives at the Klein-Gordon equation $(\hat{p}^2 - m^2c^4)\psi = 0$. Because of interpretational problems (since for example the energy operator enters the Klein-Gordon equation quadratically), one can think of taking the square root:

$$\begin{aligned}\hat{p}^2 - m^2c^4 &= (\gamma\hat{p} + mc^2)(\gamma\hat{p} - mc^2) = \gamma^{\mu}\gamma^{\nu}\hat{p}_{\mu}\hat{p}_{\nu} - m^2c^4 \\ &= \frac{1}{2}(\gamma^{\mu}\gamma^{\nu} + \gamma^{\nu}\gamma^{\mu})\hat{p}_{\mu}\hat{p}_{\nu} - m^2c^4.\end{aligned}$$

This can not be realized using ordinary numbers γ_μ; these coefficients must obey $(\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu) = 2\eta^{\mu\nu}$. And these are just the defining equations for the Dirac matrices.

It is shown in Sect. 5.5.4 that together with ψ, also its charge conjugate $\psi^C = i\gamma^2\psi^*$ fulfills the Dirac equation (5.39). Spinors obeying the Dirac equation with a nonvanishing mass m_D are *Dirac fermions*. The pair (ψ, ψ^C) constitutes a pair of (massive) particle and its anti-particle, the electron and the positron being historically the first representatives. If $m_D = 0$ the field is called a (four-component) *Weyl fermion*, propagating according to the field equation $i\gamma^\mu\partial_\mu\psi = 0$. Splitting ψ as $\psi = \psi_L + \psi_R$ and then acting with the projection operators $P_{\{L,R\}}$ (defined by (B.27) on the field equation

$$P_{\{L,R\}} i\gamma^\mu\partial_\mu(\psi_L + \psi_R) = \pm i\gamma^\mu\partial_\mu\Psi_{\{L,R\}} = 0$$

reveals that the Weyl equation holds separately for the chiral components of the field. This relates to the observation that the Lagrangian for Weyl fields exhibits another global symmetry, $\psi \to e^{i\phi\gamma^5}\psi$. By Noether's theorem this gives rise to the conserved axial current $J^\mu := -\bar{\psi}\gamma^\mu\gamma^5\psi$.

As will be extensively explained in the next chapter, nature is chiral: She distinguishes between left-handed and right-handed fermions. Specifically, left-handed neutrinos are favored and right-handed neutrinos are suppresed in weak interaction processes. Therefore the Lagrangians and field equations are most appropriately written in the projected components

$$\psi = P_L\psi + P_R\psi = \psi_L + \psi_R = \begin{pmatrix} \xi_\alpha \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \bar{\eta}^{\dot{\alpha}} \end{pmatrix}.$$

Furthermore, $\overline{\psi}_L = \overline{\psi}\, P_R$, $\overline{\psi}_R = \overline{\psi}\, P_L$ and thus $\overline{\psi}_L\psi_L = 0 = \overline{\psi}_R\psi_R$. The unbundling of any Dirac bilinear $\overline{\psi}\Gamma\psi$ can be computed by evaluating the identity $\overline{\psi}\psi = \overline{\psi}(P_L + P_R)\,\Gamma\,(P_L + P_R)\psi$ using the rules

$$P_L\gamma^\mu = \gamma^\mu P_R \qquad P_R\gamma^\mu = \gamma^\mu P_L \qquad\qquad P_L\gamma^5 = -\gamma^5 P_R \qquad P_R\gamma^5 = \gamma^5 P_L.$$

Thus, specifically the scalar term and the vector term become

$$\begin{aligned} \overline{\psi}\psi &= \overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L \\ \overline{\psi}\gamma^\mu\psi &= \overline{\psi}_L\gamma^\mu\psi_L + \overline{\psi}_R\gamma^\mu\psi_R. \end{aligned}$$

So a mass term mixes R, L-fermions whereas a current term does not. The Dirac Lagrangian obtains the form

$$\mathcal{L}_D = i\overline{\psi}_L\partial\!\!\!/\psi_L + i\overline{\psi}_R\partial\!\!\!/\psi_R - m_D(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$$

leading to the separate field equations

$$i\partial\!\!\!/\psi_L = m_D\psi_R \qquad\qquad i\partial\!\!\!/\psi_R = m_D\psi_L. \tag{5.40}$$

These decouple for Weyl fermions, as seen previously.

From two Weyl spinors (ξ, η) one can also construct the Lorentz invariant Lagrange density

$$\mathcal{L}_M = i\eta\sigma^\mu\partial_\mu\bar{\eta} + i\bar{\xi}\bar{\sigma}^\mu\partial_\mu\xi - \frac{m_\eta}{2}(\eta^2+\bar{\eta}^2) - \frac{m_\xi}{2}(\xi^2+\bar{\xi}^2) \tag{5.41}$$

leading to the field equations

$$\frac{\partial\mathcal{L}_M}{\partial\eta^\alpha} = i\sigma^\mu_{\alpha\dot{\alpha}}\partial_\mu\bar{\eta}^{\dot{\alpha}} - m_\eta\eta_\alpha = 0 \qquad\qquad \frac{\partial\mathcal{L}_M}{\partial\bar{\xi}^{\dot{\alpha}}} = i\bar{\sigma}^{\mu\dot{\alpha}\alpha}\partial_\mu\xi_\alpha - m_\xi\bar{\xi}^{\dot{\alpha}} = 0.$$

For $m_\xi = m_M = m_\eta$ this set of field equations can be rewritten in terms of the four-component spinor $\psi = \begin{pmatrix}\xi^\alpha \\ \bar{\eta}^{\dot{\alpha}}\end{pmatrix}$ with its charge conjugate $C\overline{\psi}^T = \begin{pmatrix}\eta^\alpha \\ \bar{\xi}^{\dot{\alpha}}\end{pmatrix}$ in the form

$$i\partial\!\!\!/\psi = m_M\psi^C. \tag{5.42}$$

This is the *Majorana equation*, which was discovered in 1937 by Ettore Majorana as a field equation which is "just as" Lorentz invariant as the Dirac equation. The distinguishing feature from the Dirac equation is the appearance of both the Majorana fermion and its charge conjugate within the same equation. By taking the complex conjugate of (5.42) and multiplying with γ^2, one finds that the Majorana equation can likewise be written as $i\partial\!\!\!/\psi^C = m_M\psi$. From this and the original equation one further derives that $-\gamma^\mu\partial_\mu\gamma^\nu\partial_\nu = m_M$, yielding the Klein-Gordon equation for each of the components and justifying to interpret m_M as a mass. The (four-spinor) Lagrangian that leads to the Majorana equation is

$$\mathcal{L}_M = \overline{\psi}i\partial\!\!\!/\psi - \frac{1}{2}m_M(\psi^T C\psi + \overline{\psi}C\overline{\psi}^T). \tag{5.43}$$

The mass term can be rewritten substantially by starting for instance from the identity $\psi^T C\psi = \psi^T(P_L+P_R)C(P_L+P_R)\psi$ and

$$\begin{aligned}\psi^T C\psi &= \psi^T P_L C P_L\psi + \psi^T P_L C P_R\psi + \psi^T P_R C P_L\psi + \psi^T P_R C P_R\psi \\ &= \psi_R^T C\psi_R + \psi_L^T C\psi_L,\end{aligned}$$

by using the relation $P_{\{R,L\}}C = CP_{\{L,R\}}$, the feature $P_L P_R = 0$ of the projection operators, and $\psi^T P_{\{R,L\}} = \psi^T_{\{L,R\}}$. If we again allow different masses for the terms in R and L, the Majorana mass term amounts to

$$\mathcal{L}_{Mmass} = -m_R\psi_R^T C\psi_R - m_L\psi_L^T C\psi_L = -m_R\overline{\psi}_L^T\psi_R - m_L\overline{\psi}_R^T\psi_L. \tag{5.44}$$

The Majorana field cannot be coupled to the electromagnetic field, since ψ and ψ^C have–by construction–opposite charges. Therefore Majorana fields represent massive neutral particles that are their own anti-particles. As of today, no experimental findings had shed light on whether Majorana particles exist in nature and on whether neutrinos are to be described by Dirac or by Majorana fields.[8]

Spinor Quantization

In the following, I will give only a few details of the canonical quantization of the Dirac field starting from the field equations $(i\partial\!\!\!/ - m)\psi = 0$. As in the scalar case, it is advantageous to work with Fourier components of ψ, that is in momentum space. The plane wave *ansätze* $u(k)e^{-ikx}$ and $v(k)e^{ikx}$ imply

$$(\gamma^\mu k_\mu - m)u(k) = 0 \qquad\qquad (\gamma^\mu k_\mu + m)v(k) = 0.$$

In the rest frame $(k = (m, \vec{0}))$ these equations are simply $(\gamma^0 - 1)u(0) = 0$ and $(\gamma^0 + m)v(0) = 0$. In the Dirac representation (B.10), their independent solutions are

$$u_1(0) = \begin{pmatrix}1\\0\\0\\0\end{pmatrix} \qquad u_2(0) = \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \qquad v_1(0) = \begin{pmatrix}0\\0\\1\\0\end{pmatrix} \qquad v_2(0) = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}.$$

The momentum-dependent solutions are obtained by applying Lorentz boosts on $u_\alpha(0)$ and $v_\alpha(0)$. Since in the context of symmetries we are not that much interested in these solution, I refer to any textbook on QFT. We need, however, certain completeness or orthogonality conditions. At first one observes that the spinors u_α and v_α transform under a Lorentz transformation in the same way as ψ. Therefore expressions of the form $\bar{u}u$, $\bar{v}v$ are Lorentz scalars if one defines $\bar{u} = u^\dagger\gamma^0$. One finds specifically

$$\bar{u}_\alpha(k)u_\beta(k) = \left(\frac{k\!\!\!/ + m}{2m}\right)_{\alpha\beta} \tag{5.45a}$$

$$\bar{v}_\alpha(k)v_\beta(k) = \left(\frac{k\!\!\!/ - m}{2m}\right)_{\alpha\beta} \tag{5.45b}$$

Furthermore, $\bar{u}_\alpha v_\beta = 0$ and $\bar{v}_\alpha u_\beta = 0$. Arguing along the line of reasoning for the scalar field, the Fourier expansion of the four components ψ of the Dirac spinor field is

$$\psi(x) = (2\pi)^{-3/2} \int d^3k \sqrt{m/\omega_k} \sum_{\alpha=1,2} \left\{ b(\vec{k})u_\alpha(\vec{k})e^{-ik\cdot x} + d^\dagger(\vec{k})v_\alpha(\vec{k})e^{ik\cdot x} \right\}.$$

[8] Conclusive results should come from experiments on double beta decay; see 6.4.2.

Don't be disturbed here by the appearance of a different normalization constant compared to the scalar case. The extra factor $(2m)$ can be reabsorbed into the normalization of the spinor functions u_α, v_α.

The next step is to express the Hamiltonian in terms of the creation and annihilation operators $(b^\dagger, d^\dagger)$ and (b, d). The canonical momentum to ψ is derived from $\mathcal{L}_D = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi$ as $\pi = \partial \mathcal{L}_D / \partial(\partial_0 \psi) = i\bar{\psi}\gamma^0$. Therefore, with the use of the relations (5.45)

$$\begin{aligned} H &= \int d^3x \, (\pi \partial_0 \psi - \mathcal{L}) = \int d^3x \, (i\bar{\psi}\gamma^0 \partial_0 \psi) \\ &= \int d^3k \; \omega_k \sum_\alpha \left[b_\alpha^\dagger(\vec{k}) b_\alpha(\vec{k}) - d_\alpha(\vec{k}) d_\alpha^\dagger(\vec{k}) \right]. \end{aligned}$$

A priori, nothing prevents the energy given by this Hamiltonian to become negative. This situation can be circumvented, however, by a far-reaching assumption: If the creation and annihilation operators obey anti-commutation relations

$$\{b_\alpha(\vec{k}), b_\beta^\dagger(\vec{k}')\} = \delta_{\alpha\beta}\delta^3(\vec{k} - \vec{k}') = \{d_\alpha(\vec{k}), d_\beta^\dagger(\vec{k}')\}$$

instead of the commutator relation for scalar fields, the Hamiltonian becomes (again after normal ordering to get rid of the infinite zero-point energy)

$$H = \int d^3k \; \omega_k \sum_\alpha \left[b_\alpha^\dagger(\vec{k}) b_\alpha(\vec{k}) + d_\alpha^\dagger(\vec{k}) d_\alpha(\vec{k}) \right].$$

This allows for a Fock space interpretation of states generated by successively applying operators $b_\alpha^\dagger(\vec{k})$ and $b_\alpha^\dagger(\vec{k})$ to the vacuum state $|0\rangle$.

Having anti-commutators instead of commutators amounts to imposing equal-time anti-commutators for the fields and their conjugates:

$$\{\psi_i(\vec{x}, t), \psi_j^\dagger(\vec{x}', t)\} = \delta_{ij}\delta^3(\vec{x} - \vec{x}').$$

Yukawa Interaction

With spin-0 and spin-$^1\!/_2$ fields at our disposal, we are able to write down an interaction Lagrangian for these two varieties of fields. The terms in this Lagrangian are generically composed from two spinors and one scalar

$$\mathcal{L}_Y = g_Y \bar{\psi}\varphi\psi.$$

Notice that this is the only possible interaction term compatible with Lorentz invariance and the renormalizability criterion: In four dimensions the coupling g_Y has zero mass-dimension. $\mathcal{L}_Y$ is called after H. Yukawa who aimed to describe the nuclear force among protons and neutrons by the exchange of pi mesons.

5.3.4 Gauge Vector Fields

The next in the sequence of fields to be investigated is the action for a vector field. A prime example has been known since Maxwell and Minkowski as the electromagnetic field A_μ, in a quantum field situation representing a photon. One of the preeminent achievements in particle physics is the rationale that not only the electromagnetic, but also the weak and the strong interactions are mediated by vector fields, generically named gauge bosons, and that, despite of the different strengths and phenomenological appearances,[9] the dynamics of gauge bosons is described by gauge theories. These gauge theories are of Yang-Mills type, and each Yang-Mills theory obtains its form through a symmetry group.

The U(1) Gauge Field: Electromagnetism

The origin of introducing gauge fields was the observation that the Lagrange densities for the charged scalar and the Dirac fields, generically denoted $\mathcal{L}_S(\Phi, \partial\Phi)$, are invariant under global, that is spacetime independent, $\mathbf{U(1)}$ -transformations $\Phi'(x) = e^{i\theta}\Phi(x)$, giving rise to a conserved Noether current

$$j^\mu = i\Phi\frac{\partial\mathcal{L}_S}{\partial(\partial_\mu\Phi)} + h.c.$$

However, such a global transformation all over spacetime[10] appears to be non-physical and even not in the spirit of special relativity. More appropriate is a transformation behavior

$$\Phi'(x) = e^{i\theta(x)}\Phi(x)$$

with a spacetime dependent phase $\theta(x)$. But, in order to save the symmetry, one now has to modify the kinetic terms in the Lagrangian $\mathcal{L}_S$, since the derivatives of the fields Φ transform as

$$\partial_\mu\Phi(x) \to \partial_\mu\Phi'(x) = \partial_\mu(e^{i\theta(x)}\Phi(x)) = e^{i\theta}\partial_\mu\Phi + i(\partial_\mu\theta)e^{i\theta}\Phi.$$

The extra term with derivatives of the phase prevents the invariance of the theory. This observation (and as we will see later–a certain analogy to general relativity) suggests the definition of a modified derivative (called a *covariant derivative*) with the aid of an extra field A_μ as

$$D_\mu\Phi(x) := (\partial_\mu + ieA_\mu(x))\Phi(x). \tag{5.46}$$

---

[9] This ranges from the dynamo on your bicycle to fusion processes in the sun and to the atomic bomb.

[10] from the theatre of Pompey at the moment of Caesar's death to the black hole horizon of the Andromeda galaxy at the moment of last scattering, say.

(Don't be bothered by the constant (ie) appearing together with the new field in (5.46); this is merely a convention, and the motive for its introduction will be explained soon.) The transformation behavior of A_μ is to be derived from

$$(D_\mu \Phi(x))' \stackrel{!}{=} e^{i\theta(x)} D_\mu \Phi(x).$$

leading to

$$A'_\mu = A_\mu - \frac{1}{e}\partial_\mu \theta. \tag{5.47}$$

The expression (5.46) looks familiar: It was used simply as an abbreviation (3.37) in the example of the charged scalars interacting with an electromagnetic field. Now we see that the *raison d'être* of the electromagnetic field hinges on the quest for local gauge invariance. This astounding finding is due to H. Weyl [551]. His recipe amounts to modifying the originally globally-invariant Lagrangian into a locally-invariant one according to

$$\hat{\mathcal{L}}_S := \mathcal{L}_S(\partial_\mu \to D_\mu).$$

The substitution $\partial_\mu \to D_\mu$ is known as the *minimal coupling*. It gives rise to a term in the Lagrangian describing the interaction of Φ with the gauge field A_μ. Minimal coupling is a realization of what also is called the *gauge principle* or the *principle of local covariance*. Whatever you call it, the minimal coupling recipe is a procedure to enlarge matter field theories, which exhibit global symmetries, into theories with local symmetries by coupling the matter fields to gauge fields.

In order that the gauge potential A_μ become dynamic fields themselves, they need their own kinetic term $\mathcal{L}_G$ containing derivatives $\partial_\nu A_\mu$. Of course we know since H. Minkowski and M. Born that the derivatives of the fields enter only in the form $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Indeed, this field strength tensor is invariant with respect to the gauge transformations (5.47):

$$F'_{\mu\nu} = \partial_\mu A'_\nu - \partial_\nu A'_\mu = \partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu}.$$

We note in passing that the field strength arises naturally from the commutator of covariant derivatives

$$(D_\mu D_\nu - D_\nu D_\mu)\phi = ieF_{\mu\nu}\phi. \tag{5.48}$$

We also know the Lagrangian for the electromagnetic field

$$\mathcal{L}_G = \mathcal{L}_{ED} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{5.49}$$

The full fledged theory of the electromagnetic field and the source fields is thus

$$\mathcal{L} = -\frac{1}{4}\, F^{\mu\nu} F_{\mu\nu} + \mathcal{L}_S(\partial_\mu \to D_\mu). \tag{5.50}$$

If the source field is the Dirac field, we recover the defining Lagrangian of QED (QuantumElectroDynamics)

$$\mathcal{L}_{QED} = -\frac{1}{4}\,F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\not{D} - m)\psi = -\frac{1}{4}\,F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\not{\partial} - m)\psi + ej^{\mu}A_{\mu} \tag{5.51}$$

with the Dirac Noether-current $j^{\mu} = -\bar{\psi}\gamma^{\mu}\psi$. Observe the rôle of symmetries in this case: Under global symmetries we can have charge, but only under local symmetries do we get light/photons.

When introduced by Weyl, the covariant derivative had in its definition the elementary charge (the constant ie in (5.46)). In the literature the symbol "*e*" has two different meanings: If *-e* is the charge of the electron, *+e* is the strength of the electromagnetic coupling, and this is meant here. In any case, covariant derivatives are geometric entities which are of course free of any reference to physical coupling constants whatsoever. Therefore, some authors for puristic reasons choose to define their covariant derivative as $\partial_{\mu} + A_{\mu}$. But where has the coupling constant gone? Its tracks can be followed for instance in the QED equation. Let someone define her covariant derivative as $\partial_{\mu} + \lambda\tilde{A}_{\mu}$. No one can forbid her to do so, and therefore you find various λ's in the literature; favorites being $\{\pm ie, \pm i, 1\}$. The QED Lagrangian (5.51) rewritten in terms of the fields $\tilde{A}_{\mu}$ becomes

$$\mathcal{L}_{QED} = -\frac{1}{4}\left(\frac{\lambda}{ie}\right)^{2}\tilde{F}^{\mu\nu}\tilde{F}_{\mu\nu} + \bar{\psi}(i\not{\partial} - m)\psi - i\lambda j^{\mu}\tilde{A}_{\mu}.$$

Thus for instance in the mathematical proper case with $\lambda = 1$, there is no coupling constant in the last term, but it enters the kinetic term of the gauge potential as $(1/e^2)$.

Despite the successful confirmation of QED, we dare to ask the question as to whether the gauge Lagrangian (5.49) is unique. One can show that the second Noether theorem requires that the derivatives of the gauge field occur in terms of $F_{\mu\nu}$ only, and that there is no dependence on the gauge fields themselves. Actually a proof will be given below considering the more general non-Abelian gauge theory. Observe that there cannot be a mass term $M^2 A_{\mu}A^{\mu}$ in the Lagrangian, although it is allowed by dimensional arguments. But a mass term would destroy the local gauge invariance– in fact: $\delta(A_{\mu}A^{\mu}) \propto A^{\mu}\partial_{\mu}\phi$. The gauge Lagrangian must be a Lorentz scalar. As building blocks we have at our disposal the two invariant tensors of the Lorentz group $(\eta_{\mu\nu}, \epsilon_{\mu\nu\rho\sigma})$ and the field strengths $F_{\mu\nu}$. The only linear invariant that can be built is $\eta^{\mu\nu}F_{\mu\nu}$ but this vanishes identically. There are two possible quadratic terms, namely $F^{\mu\nu}F_{\mu\nu}$ and $\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$. The former is our favorite choice, the latter is parity-breaking because of the presence of the completely antisymmetric ϵ-tensor, and this is excluded in electromagnetic processes. It can be written as a total derivative, in any case,

$$\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} = 4\epsilon^{\mu\nu\rho\sigma}(\partial_{\mu}A_{\nu})(\partial_{\rho}A_{\sigma}) = \partial_{\mu}(4\,\epsilon^{\mu\nu\rho\sigma}A_{\nu}\partial_{\rho}A_{\sigma})$$

and thus does not affect the field equations. The mass dimension of the field strengths is $[F_{\mu\nu}] = [\partial_\mu] + [A_\mu] = 2$, since $[A_\mu] = [\partial_\mu] = 1$. Thus in four dimensions the quadratic term has the proper mass dimension in order not to spoil naive renormalization. This dimensional argument also rules out choosing a gauge field Lagrangian $\mathcal{L}_G = f(F)$ other than $f(F) \propto F^2$. What about other coupling terms between photons and fermions? Local symmetry allows a "Pauli term" $g_P \bar{\psi} \left[\gamma^\mu, \gamma^\nu\right] F_{\mu\nu}$ with a coupling g_P. But for this interaction part $\Delta = 4 - 1 - (2 \times (\frac{3}{2}) - 1) = -1$, so it does not survive the criterion (5.23); the coupling g_P has mass-dimension $1/M$. In conclusion, symmetry arguments and the quest for renormalizability fix the QED Lagrangian completely.

The procedure of gauging a globally invariant theory will be exposed in another manner. This starts from the observation that the QED Lagrangian (5.51) can be written in two equivalent forms

$$\mathcal{L} = \mathcal{L}_G + \hat{\mathcal{L}}_D \qquad \text{with} \quad \hat{\mathcal{L}}_D := \mathcal{L}_D(\partial_\mu \to D_\mu) \tag{5.52}$$

$$= \mathcal{L}_G + \mathcal{L}_D + e j^\mu A_\mu \qquad \text{with} \quad j^\mu = i \frac{\partial \mathcal{L}_D}{\partial(\partial_\mu \Phi)} \Phi \tag{5.53}$$

with the Noether current j^μ arising from the global symmetry. Thus replacing the partial derivative by a covariant derivative in the globally- invariant Lagrangian, with the appropriate transformation behavior of the gauge potential, is equivalent to coupling the gauge potential to the Noether current. This can be formulated generically as: Start from an action

$$S_0 = \int d^4x \, \mathcal{L}_S(Q, \partial Q)$$

that is invariant under global phase transformations, infinitesimally given by $\delta Q = i\theta Q$, and which brings about a Noether current j^μ. Allowing the θ's to become spacetime dependent, this action transforms infinitesimally as

$$\delta_\theta S_0 = \int d^4x \, j^\mu(x) \partial_\mu \theta(x).$$

In order to restore local invariance, introduce a compensating field A_μ and add an action term $S_1 = \int d^4x \, g \, j^\mu A_\mu$. Now

$$\delta_\theta(S_0 + S_1) = \int d^4x \, \left\{ j^\mu(x) \partial_\mu \theta(x) + g j^\mu \delta A_\mu + g(\delta j^\mu) A_\mu \right\}.$$

Since the gauge field A_μ transforms as $\delta A_\mu(x) = -\frac{1}{g} \partial_\mu \theta(x)$, the two first terms cancel. In the case of QED, $\delta j^\mu = 0$ and $(S_0 + S_1)$ is an invariant action. In general, one must–owing to the circumstances–modify the transformations and the action unless the procedure yields a locally-invariant action. This technique is sometimes called the "Noether coupling method" or "Noetherization". The procedure has been employed in supergravity as will be described in Sect. 8.3.

To complete the story line, let me state the field equations for the Lagrangian (5.51). The Euler-Lagrange derivatives are

$$[\mathcal{L}]^\mu = \frac{\delta L}{\delta A_\mu} = \partial_\nu F^{\mu\nu} + e\,\bar{\psi}\gamma^\mu\psi \tag{5.54a}$$

$$[\mathcal{L}]_\alpha = \frac{\delta L}{\delta \psi_\alpha} = i\partial_\mu(\bar{\psi}\gamma^\mu)_\alpha + m\,\bar{\psi}_\alpha - e(\bar{\psi}A_\mu\gamma^\mu)_\alpha \tag{5.54b}$$

$$[\bar{\mathcal{L}}]_\alpha = \frac{\delta L}{\delta \bar{\psi}_\alpha} = i\partial_\mu(\gamma^\mu\psi)_\alpha - m\,\psi_\alpha + e(A_\mu\gamma^\mu\psi)_\alpha. \tag{5.54c}$$

The respective Noether identity due to the local symmetry of (5.51) is found to be

$$[\mathcal{L}]^\mu_{\ ,\,\mu} + ie\left([\mathcal{L}]_\alpha\psi_\alpha + \bar{\psi}_\alpha\,[\bar{\mathcal{L}}]_\alpha\right) \equiv 0.$$

Previously, I stressed that a massive vector field ruins gauge invariance. Nevertheless, we will see massive vector fields arising in the Glashow-Salam-Weinberg model of electroweak interactions. Therefore, here some words about what is known as the *Proca theory*, defined by the Lagrangian

$$\mathcal{L}_P = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}M^2\,A^\mu A_\mu \tag{5.55}$$

or the field equations

$$\partial_\mu F^{\mu\nu} + M^2\,A^\nu = 0.$$

In taking partial derivatives ∂_ν of this equation we obtain, because of the antisymmetry of the field strength, the Lorenz equation $\partial_\nu A^\nu = 0$. Observe that here it arises as an equation of consistency, whereas in Maxwell electrodynamics it may serve as a gauge-fixing condition. Therefore in contrast to the Maxwell theory with its two degrees of freedom (or in another manner of speaking: two polarization directions), the Proca field has three degrees of freedom: two transverse ones and a longitudinal one. Using the Lorenz equation we can write the Proca field equations as $(\Box + M^2)A^\mu = 0$, justifying that M can be interpreted as the mass of the vector field. Let me point out here that massive electrodynamics can be made invariant by introducing a scalar field, sometimes called the Stückelberg trick; I will get back to this point in the concluding section of this chapter.

All of the above concerned the classical theory. Attempts to quantize electrodynamics go back as far as the late 1920's and revealed early on that there is a price to pay for gauge invariance in the theory. These difficulties originate in that, depending on the choice of the time variable, certain components of the gauge field do not have a kinetic term. This gives rise to constraints in phase space[11]. For example, the canonical momenta to the A_μ in the Maxwell theory are (in the instant form) given by $\Pi^\mu = \partial\mathcal{L}/(\partial_0 A_\mu) = F^{\mu 0}$ and therefore $\Pi^0 = 0$. This not only prevents making a Legendre transformation classically, but quantum-mechanically it turns out to be incompatible with the canonical commutation relation $[A_\mu(x), \Pi^\nu(y)] = i\delta^\nu_\mu\delta^3(\vec{x} - \vec{y})$.

[11] More about "constrained dynamics" in Appendix C.

The Fermi "trick" of breaking the gauge invariance by starting from the Lagrangian

$$\widehat{\mathcal{L}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{\lambda}{2} (\partial_\mu A^\mu)^2$$

that is, by adding a Lorentz-invariant term with a multiplier λ at first looks promising, in that the momenta for the vector fields

$$\Pi^\nu = F^{\nu 0} - \lambda g^{\nu 0} (\partial_\mu A^\mu)$$

are no longer zero. It is true that the Euler derivatives with respect to λ result in the Lorenz condition $\partial_\mu A^\mu = 0$—which is nice. But the canonical commutation relations are not compatible with the Lorenz condition, since this is proportional to the momentum, and the commutators for the momenta should vanish. (Furthermore, in this version the momentum canonically conjugate to the multiplier field is zero.)

Instead of introducing gauge conditions on the Lagrangian level one should impose them on the surface in phase space that is defined by the constraints. But for many gauge choices, manifest Lorentz invariance is lost. One can alternatively choose a gauge in which Lorentz invariance is guaranteed, but then one needs a procedure to get rid of negative norm ("ghost" states[12]). This procedure goes by the name Gupta-Bleuler quantization, invented independently by S. N. Gupta and by K. Bleuler in 1950. It starts from the decomposition of the fields in terms of the Fourier modes

$$A_\mu(x) = \int \frac{d^3k}{2\omega_k} \left\{ a^\lambda(k) \epsilon^\lambda_\mu(k) e^{-ikx} + a^{\dagger\lambda}(k) \epsilon^\lambda_\mu(k) e^{+ikx} \right\}$$

with the identification of the creation and annihilation operators as $a^{\dagger\lambda}(k)$ and $a^\lambda(k)$ respectively. The canonical commutation relations for the fields and the momenta lead to

$$[a^\lambda(k), a^{\dagger\lambda'}(k')] = -\eta^{\lambda\lambda'} \delta^3(\vec{k} - \vec{k}').$$

The presence of the indefinite metric η signals that in the non-restricted theory the norm of states $\langle | a^\lambda(k) a^{\dagger\lambda}(k') | \rangle$ might become negative. This is the price of a covariant formulation of the theory. But there is still the Lorenz condition. Can it be imposed as $\partial_\mu A^\mu |\psi\rangle = 0$? Yes it can, but this condition is so strong that no states survive (as shown for instance by F. J. Belinfante in 1948). Gupta found a less stringent condition

$$(\partial_\mu A^\mu)^{(+)} |\psi\rangle = 0$$

where the $(+)$ refers to the positive frequency part. And he showed that this condition on states removes the negative norm states from the spectrum. Full QED can be treated in a largely similar manner.

[12] not to be confused with the ghosts in the Faddeev-Popov and BRST quantization.

Non-Abelian Gauge Theory or Yang-Mills Theory

In non-Abelian gauge theories[13], instead of $\mathbf{U(1)}$ one deals with some other Lie group as the local symmetry group. In 1954, C. N. Yang and R. Mills [570] derived the generic framework for the group $\mathbf{SU(2)}$. The interest for this group arose from Heisenberg's introduction of isotopic spin in nuclear physics. R. Utiyama [513] extended the gauge procedure to arbitrary non-Abelian symmetry groups[14]. In the article [513] he describes the following program: "The main purpose of the present paper is to investigate the following problem. Let us consider a system of fields $Q^A(x)$, which is invariant under some transformation group G depending on parameters $\epsilon_1, \epsilon_2, \cdots \epsilon_n$. Suppose that the aforementioned parameter-group G is replaced by a wider group G', derived by replacing the parameters ϵ's by a set of arbitrary functions $\epsilon(x)$'s, and that the system considered is invariant under this wider group G'. Then, can we answer the following questions by using only the postulate of invariance stated above? (1) What kind of field, $A(x)$, is introduced on account of the invariance? (2) How is this new field A transformed under G'? (3) What form does the interaction between the field A and the original field Q take? (4) How can we determine the new Lagrangian $L'(Q, A)$ from the original one $L(Q)$? (5) What type of field equations for A are allowable?". This program will essentially be detailed on the following pages.

Given a Lagrange density $\mathcal{L}_S(\Phi_\alpha, \partial_\mu \Phi_\alpha)$ with "matter/source" fields Φ_α, $(\alpha = 1, \ldots K)$, which can be charged scalar fields or Dirac fields. Assume that $\mathcal{L}_S$ is globally-invariant under an N-dimensional symmetry (Lie)-group $\mathbf{G}$, that is with respect to transformations

$$\Phi' = e^{i\Theta}\Phi = e^{i\theta^a T^a}\Phi, \tag{5.56}$$

or infinitesimally and with all indices:

$$\delta\Phi_\alpha = i\theta^a (T^a)_\alpha^{\ \beta} \Phi_\beta.$$

Here the θ^a are constant infinitesimal parameters and T^a are K-dimensional representation matrices of the Lie algebra

$$[X^a, X^b] = i f^{abc} X^c \qquad (a = 1, \ldots N). \tag{5.57}$$

Global invariance under (5.56) means that

$$\delta_\Theta \mathcal{L}_S = \Big([\mathcal{L}_S]^\Phi (i\Theta\Phi) + h.c.\Big) + \partial_\mu \Big(\Pi^{\phi\mu}(i\Theta\Phi) + h.c.\Big) = 0$$

---

[13] In those cases where I use the term YM-like theories, I assume a theory with a quadratic term in the gauge field strengths, as it is derived below. In this sense, also electrodynamics is a YM-like theory.

[14] Utiyama investigated among others the Lorentz group as a local symmetry group. More about this in Sect. 7.6.3.

giving rise to weakly conserved Noether currents

$$j^{\mu}_{\Phi_a} = i\frac{\partial \mathcal{L}_S}{\partial(\partial_\mu \Phi)} T_a \Phi + h.c.. \tag{5.58}$$

As in the Abelian case, the Lagrangian $\mathcal{L}_S$ can be made locally invariant by the introduction of gauge fields. In this case, one introduces N vector fields A^a_μ ("gauge bosons") in covariant derivatives

$$D_\mu \Phi := (\partial_\mu + igA^a_\mu T^a)\Phi. \tag{5.59}$$

In this expression g is a constant describing the strength of the coupling between the vector fields and the Φ_α. If the symmetry group decomposes into different subgroups ($G = G_1 \times G_2 \times \ldots$), the algebra decomposes into disjunct subalgebras, and for each of them the constants g_i are different. We will meet this situation in the standard model of particle physics, where the gauge group is $\mathbf{SU(3)} \times \mathbf{SU(2)} \times \mathbf{U(1)}$.

Now making the replacement $(\partial_\mu \to D_\mu)$ in $\mathcal{L}_S$ according to the gauge principle, the modified Lagrangian $\hat{\mathcal{L}}_S(\Phi, D\Phi)$ is invariant with respect to local transformations

$$\Phi' = \Omega(x)\Phi \quad \text{with} \quad \Omega(x) = e^{i\Theta} = e^{i\theta^a(x)T^a} \tag{5.60}$$

if the gauge potentials transform appropriately, namely such that $(D\Phi)' \stackrel{!}{=} \Omega(D\Phi)$. Explicitly this requirement reads

$$(\partial\Phi)' + (igA\Phi)' = \Omega\partial\Phi + (\partial\Omega)\Phi + ig(A\Phi)' \stackrel{!}{=} \Omega\partial\Phi + \Omega igA\Phi$$

from which one obtains the transformation behavior of the gauge fields as

$$A'_\mu = \Omega A_\mu \Omega^{-1} - \frac{1}{ig}(\partial_\mu \Omega)\Omega^{-1}. \tag{5.61}$$

Infinitesimally this is

$$\delta_\Theta A_\mu = -\frac{1}{g}\partial_\mu \Theta + i[\Theta, A_\mu]$$

or in components

$$\delta_\theta A^a_\mu = -\frac{1}{g}\partial_\mu \theta^a(x) - \theta^b(x) A^c_\mu f^{bca} \tag{5.62}$$

which is generalization of the Abelian (5.47) to the non-Abelian case.

Thus far we are able to take a covariant derivative of a "matter" field Φ by (5.59). Correspondingly it is possible to define the covariant derivative for objects G^a with indices in the Lie algebra. Starting with the *ansatz* $D_\mu G^a := \partial_\mu G^a + \lambda A^c_\mu f^{cba} G^b$ with some constant λ, we now try to obtain an expression

$$D_\mu G^a = \partial_\mu G^a + igA^c_\mu (\hat{T}^c)^{ab} G^b$$

where the $\hat{T}$ are an N-dimensional representation of the Lie algebra. Choosing $\lambda = g$ the comparison of both expressions yields

$$(\hat{T}^c)^{ab} = -i f^{cba},$$

revealing that $\hat{T}$ is the adjoint representation of the Lie algebra. Thus

$$D_\mu G^a := \partial_\mu G^a + g f^{bca} A^b_\mu G^c. \tag{5.63}$$

With this covariant derivative, (5.62) becomes

$$\delta_\theta A^a_\mu = -\frac{1}{g} D_\mu \theta^a.$$

We already know from the $\mathbf{U}(\mathbf{1})$ case that covariant derivatives do not commute. A short calculation yields

$$\begin{aligned}(D_\mu D_\nu - D_\nu D_\mu) G^a &= i g f^{bca} G^c \big((\partial_\mu A^b_\nu - \partial_\mu A^b_\nu) - g A^b_\mu A^c_\nu f^{bca}\big) \\ &=: i g f^{bca} G^c \tilde{F}^b_{\mu\nu}.\end{aligned} \tag{5.64}$$

Here, for the moment, $\tilde{F}^b_{\mu\nu}$ is simply an abbreviation. With the aid of the identities following from the second Noether theorem we will later recognize it as the Yang-Mills field strength.

Kinetic Energy Term for the Gauge Fields

The next calculations serve to identify a kinetic term for the gauge fields. For this purpose, we need to find an invariant expression containing the derivatives of the A^a_μ. Knowing about the field strength from the Abelian theory, you may try $\check{F}^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\mu A^a_\nu$, but will discover, that this is not invariant, thus we suspect additional terms. Yang and Mills found these for $\mathbf{SU}(\mathbf{2})$ more or less by trial and error. We follow another line of reasoning by which Noether's second theorem is inverted to discover in which combinations the gauge fields and their derivatives must enter the gauge Lagrangian in order to safeguard the symmetry achieved by minimal coupling. Our starting point is again the invariance condition $0 \equiv [\mathcal{L}]_A \bar{\delta} Q^A + \partial_\mu J^\mu$ for the transformations

$$\delta\Phi = i\theta^a(x) T_a \Phi \qquad\qquad \delta A^a_\mu = -f^{bca} A^c_\mu \theta^b(x) - \frac{1}{g}\theta^a_{,\mu} = -\frac{1}{g} D_\mu \theta^a, \tag{5.65}$$

which in this case becomes

$$0 \equiv \big([\mathcal{L}]_\Phi (i\theta^a T_a \Phi) + h.c.\big) - \frac{1}{g}[\mathcal{L}]^\mu_a (D_\mu \theta^a) + \partial_\mu J^\mu.$$

Here it is assumed that the coordinates are not involved in the symmetry transformation; hence $\bar{\delta}Q^A = \delta Q^A$ for all fields. Furthermore

$$J^\mu = \Big(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\Phi)}\delta\Phi + h.c.\Big) + \frac{\partial\mathcal{L}}{\partial(\partial_\mu A^a_\nu)}\delta A^a_\nu = \Big(\Pi^{\Phi\mu}(i\theta^a T_a\Phi) + h.c.\Big) - \frac{1}{g}\Pi^{\nu\mu}_a D_\nu\theta^a$$

–assuming that no surface terms arise. Here we introduced again the "pseudo momenta"

$$\Pi^{\Phi\mu} = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\Phi)} \qquad \Pi^{\nu\mu}_a = \frac{\partial\mathcal{L}}{\partial(\partial_\mu A^a_\nu)}.$$

according to (3.23). If the Lagrangian is invariant with respect to the transformations (5.65), the Noether and the Klein-Noether identities (3.72) are identically fulfilled *(sic)* and do not provide any new information. As mentioned before, we will instead use them as conditions imposed on the Lagrangian in order to be invariant with respect to (5.65):

$$k^{\mu\nu}_a + k^{\nu\mu}_a \overset{!}{=} 0 \tag{5.66a}$$

$$-\frac{1}{g}[\mathcal{L}]^\mu_a + j^\mu_a - \partial_\nu k^{\mu\nu}_a \overset{!}{=} 0 \tag{5.66b}$$

$$([\mathcal{L}]_\Phi(iT_a\Phi) + h.c.) - [\mathcal{L}]^\mu_a f^{acb}A^c_\mu + \partial_\mu j^\mu_a \overset{!}{=} 0, \tag{5.66c}$$

with the immediate consequence of the Noether identity condition

$$([\mathcal{L}]_\Phi(iT_a\Phi) + h.c.) + \frac{1}{g}D_\mu[\mathcal{L}]^\mu_a \overset{!}{=} 0.$$

In (5.66), the

$$j^\mu_a = \big(\Pi^{\Phi\mu}(iT_a\Phi) + h.c.\big) - \Pi^{\nu\mu}_b A^c_\mu f^{acb} \qquad\qquad k^{\mu\nu}_a = -\frac{1}{g}\Pi^{\nu\mu}_a$$

are the coefficients in the Noether current

$$J^\mu_a = j^\mu_a\theta^a + k^{\mu\nu}_a\partial_\mu\theta^a = \frac{1}{g}[\mathcal{L}]^\mu_a + \frac{1}{g}\partial_\nu(\Pi^{\mu\nu}_a\theta^a).$$

Let us now make the *ansatz* for the Lagrangian as

$$\mathcal{L}(\Phi, \partial\Phi, A, \partial A) = \hat{\mathcal{L}}_S(\Phi, D\Phi) + \mathcal{L}_G(A, \partial A)$$

where $\hat{\mathcal{L}}_S$ is related to the ungauged Lagrangian by $\hat{\mathcal{L}}_S(\Phi, D\Phi) = \mathcal{L}_S(\Phi, \partial\Phi)$. Since therefore

$$\frac{\partial\mathcal{L}}{\partial(\partial_\mu\Phi)} = \frac{\partial\hat{\mathcal{L}}_S}{\partial(D_\mu\Phi)},$$

the first part in the current j^μ_a can be identified with the Noether current $j^\mu_{\Phi a}$ for the matter fields from the globally symmetric theory; see (5.58). The identities (5.66a) require that the derivatives of the gauge fields arise in a combination $\check{F}^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\mu A^a_\nu$. Thus we find that the gauge-field part of the Lagrangian is of the form $\mathcal{L}_G(A, \partial A) = \check{\mathcal{L}}_G(A, \check{F}(\partial A))$. The identities (5.66b) provide differential equations for the dependencies of $\check{\mathcal{L}}_G$ on the gauge fields themselves: First write the Euler derivatives $[\mathcal{L}]^\mu_a$ explicitly as

$$\begin{aligned}
[\mathcal{L}]^\mu_a &= \frac{\partial \mathcal{L}_G}{\partial A^a_\mu} - \partial_\nu \frac{\partial \mathcal{L}_G}{\partial A^a_{\mu,\nu}} + \left[\frac{\partial \hat{\mathcal{L}}_S}{\partial (D_\lambda \phi)} \frac{\partial (D_\lambda \phi)}{\partial A^a_\mu} + h.c. \right] \\
&= \frac{\partial \check{\mathcal{L}}_G}{\partial A^a_\mu} - \partial_\nu \Pi^{\mu\nu}_a + \left(i g T_a \Phi \Pi^{\Phi\mu} + h.c. \right). \qquad (5.67)
\end{aligned}$$

Inserted into (5.66b), this reads

$$-\frac{1}{g} \frac{\partial \check{\mathcal{L}}_G}{\partial A^a_\mu} - \Pi^{\nu\mu}_a A^c_\nu f^{cab} \overset{!}{=} 0.$$

Since

$$\Pi^{\mu\nu}_a = \frac{\partial \mathcal{L}_G}{\partial A^a_{\mu,\nu}} = \frac{\partial \check{\mathcal{L}}_G}{\partial \check{F}^c_{\rho\sigma}} \frac{\partial \check{F}^c_{\rho\sigma}}{\partial A^a_{\mu,\nu}} = -2 \frac{\partial \check{\mathcal{L}}_G}{\partial \check{F}^c_{\mu\nu}}$$

the previous condition becomes the differential equation

$$\frac{\partial \check{\mathcal{L}}_G}{\partial A^a_\mu} = -2\, g \frac{\partial \check{\mathcal{L}}_G}{\partial \check{F}^c_{\mu\nu}} A^b_\nu f_{ab}{}^c.$$

Defining $\hat{\mathcal{L}}_G(F(A, \partial A)) = \check{\mathcal{L}}_G(A, \check{F}(\partial A))$, this is solved by

$$F^a_{\mu\nu} := \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g A^b_\mu A^c_\nu f^{bca}. \qquad (5.68)$$

And here is the sought-after result: The Klein-Noether identities enforce that in a Lagrangian of a theory which is locally symmetric with respect to a group with a Lie algebra (5.57), the gauge potentials A^a_μ can only be molded in a field strengths according to (5.68). At the same time we can identify the previously introduced tensor $\tilde{F}^a_{\mu\nu}$ resulting from the commutator of covariant derivatives with $F^a_{\mu\nu}$. The field strength can easily be shown to transform as

$$F'_{\mu\nu} = \Omega F_{\mu\nu} \Omega^{-1} \quad \text{or} \quad \delta_\Theta F_{\mu\nu} = i[\Theta, F_{\mu\nu}] \quad \text{i.e.} \quad \delta_\theta F^a_{\mu\nu} = -\theta^b F^c_{\mu\nu} f^{bca}. \qquad (5.69)$$

Furthermore, from $D_\mu(G^a F^a) = (\partial_\mu G^a) F^a + G^a (\partial_\mu F^a)$ we derive

$$D_\mu F^a := \partial_\mu F^a + f^{abc} A^b_\mu F_c.$$

Because of the Jacobi identity for the covariant derivatives, the relation (5.64) immediately leads to

$$D_{[\lambda} F_{\mu\nu]} = 0 \tag{5.70}$$

which is the non-Abelian version of Maxwell's $rot\,\vec{E} = -\frac{\partial \vec{B}}{\partial t}$ and $div\,\vec{B} = 0$.

As in electrodynamics, we obtain a kinetic term for the gauge field with the renormalization-adjusted dimension, if we consider expressions quadratic in the field strength:

$$\mathcal{L}_G = F^a_{\mu\nu} F^b_{\rho\sigma} Q^{\mu\nu\rho\sigma}_{ab}. \tag{5.71}$$

Its variation under infinitesimal gauge transformations is

$$\delta_\Theta \mathcal{L}_G = -\Big[F^a_{\mu\nu}\theta^b F^c_{\rho\sigma}(f^{cbd} Q^{\mu\nu\rho\sigma}_{ad} + f^{bad} Q^{\mu\nu\rho\sigma}_{dc}) + F^a_{\mu\nu} F^b_{\rho\sigma} \delta_\Theta Q^{\mu\nu\rho\sigma}_{ab}\Big].$$

The kinetic term for the gauge field Lagrangian must be a Lorentz scalar and a scalar with respect to the gauge group. So the previous expression is bound to vanish. Since we do not want–physically unmotivated–functional constraints among the field strengths, we are looking for conditions on the coefficients Q. These should not depend on the fields. The simplest choice is

$$Q^{\mu\nu\rho\sigma}_{ab} = G_{ab}\gamma^{\mu\nu\rho\sigma}$$

with constant tensors G and γ. For the Lorentz group indices the quantity γ can be constructed from the invariant tensors $\eta^{\mu\nu}$ and $\epsilon^{\mu\nu\rho\sigma}$, leading to the Lagrangian

$$\mathcal{L}_G = g_{ab}\eta^{\mu\rho}\eta^{\nu\sigma} F^a_{\mu\nu} F^b_{\rho\sigma} + g'_{ab}\epsilon^{\mu\nu\rho\sigma} F^a_{\mu\nu} F^b_{\rho\sigma} \tag{5.72}$$

with real coefficients g_{ab} and g'_{ab} to be determined from the algebra of the gauge group in question. The second term in (5.72) is a parity-violating expression and we will not incorporate it any longer in these considerations. Like the related term in the Abelian case, it can (for symmetric g'_{ab}) be written as a total derivative. Thus it does not influence the classical field equations. However, this may change in the quantum theory and indeed it crops up at various places as in the questions of chiral anomalies and strong CP-violation, topics dealt with in Sect. 6.4. Under these more or less enforced assumptions the requirement $\delta_\theta \mathcal{L}_G = 0$ leads to the condition

$$f^{cbd} g_{ad} + f^{bad} g_{dc} \stackrel{!}{=} 0. \tag{5.73}$$

This relates the structure constants of the gauge group with the metric allowed in (5.72). It is void for Abelian groups for which one can obviously choose $g_{ab} = \delta_{ab}$. It is also fulfilled in semisimple groups if for g_{ab} one takes the Cartan-Killing metric. The relation (5.73) can be expressed in three equivalent ways[15]:

[15] Here I cite directly [536], Sect. 15.2, adapted to my notation.

- There exists a real symmetric positive-definite matrix g_{ab} that satisfies the invariance condition (5.73)
- There is a basis for the Lie algebra (that is, a set of generators $\tilde{X}^a = S^{ab} X_b$ with a non-singular matrix S) for which the structure constants $\tilde{f}^{abc}$ are antisymmetric not only in the first two indices a and b but in all three indices a, b, c.
- The Lie algebra is the direct sum of commuting compact simple and $\mathbf{U(1)}$ subalgebras.

In Appendix A of Chap. 15 in [536], it is proven that these three propositions are indeed equivalent, and that at the very end g_{ab} can be assumed to be the unit matrix. By normalizing the generators of the Lie algebra, or its representation matrices as $Tr(T^a T^b) = \frac{1}{2}\delta^{ab}$ the gauge field Lagrangian becomes

$$\mathcal{L}_{YM} = -\frac{1}{2} Tr(F_{\mu\nu} F^{\mu\nu}) = -\frac{1}{4}\, F^a_{\mu\nu} F^{\mu\nu}_a. \tag{5.74}$$

This Yang-Mills Lagrangian contains a term $(\partial_\mu A^a_\nu - \partial_\mu A^a_\nu)^2$ quadratic in the derivatives of the gauge potentials, a term $f^{abc} A^{a\mu} A^{b\nu} (\partial_\mu A^c_\nu - \partial_\mu A^c_\nu)$, and a quartic term $(f^{abc} A^{a\mu} A^{b\nu})^2$. The last two terms are absent in an Abelian theory. As in Maxwells theory, the first term describes the propagation of massless vector bosons, and the other two terms give rise to the self interaction of the gauge bosons. Again it is to be stressed that terms of the form $\alpha_{ab} A^a_\mu A^b_\mu$ cannot be made invariant by any choice of the α's. Specifically the gauge bosons are massless objects.

Field Equations

The field equations for the gauge and the matter fields follow from the complete Lagrangian

$$\mathcal{L} = \mathcal{L}_{YM}(F(A, \partial A) + \hat{\mathcal{L}}_S(\phi, D(A)\phi).$$

The field equations for the gauge potentials were already previously calculated in the form (5.67), that is

$$[\mathcal{L}]^\mu_a = \frac{\partial \check{\mathcal{L}}_G}{\partial A^a_\mu} - \partial_\nu \Pi^{\mu\nu}_a + g\, j^\mu_{\Phi a}.$$

with the matter Noether currents $j^\mu_{\Phi a}$. The first term is

$$\frac{\partial \check{\mathcal{L}}_G}{\partial A^a_\mu} = \frac{\partial \mathcal{L}_{YM}}{\partial F^b_{\rho\sigma}} \frac{\partial F^b_{\rho\sigma}}{\partial A^a_\mu} = g F^{b\mu\nu} A^c_\mu f_{acb}.$$

For the second term one finds

$$\Pi^{\nu\mu}_a = \frac{\partial \mathcal{L}}{\partial A^a_{\mu\nu}} = \frac{\partial \mathcal{L}}{\partial F^b_{\rho\sigma}} \frac{\partial F^b_{\rho\sigma}}{\partial A^a_{\mu\nu}} = 2 \frac{\partial \mathcal{L}}{\partial F^a_{\mu\nu}} = -F^{\mu\nu}_a$$

and thus the field equations for the gauge potentials are

$$\partial_\mu F_a^{\mu\nu} = -g(j^\nu_{\Phi a} - F^{b\mu\nu} A^c_\mu f_{acb}) = -g j^\nu_a. \tag{5.75}$$

The currents j_a are conserved (on-shell): $\partial_\nu j^\nu_a \doteq 0$. So they may serve to define conserved quantities

$$C_a = g \int_V d^3x \,(j^0_{\Phi a} - F^{b\mu 0} A^c_\mu f_{acb}) = g \int_{\partial V} F_a^{0k} dS_k.$$

However, these are gauge dependent: Under a gauge transformation with Ω they transform because of (5.69) as

$$C_a = g \int_{\partial V} \Omega F_a^{0k} \Omega^{-1} dS_k.$$

Only if one can find "natural" gauge conditions on the boundary ∂V–which may be taken to be the sphere S^2 at spatial infinity–the charges can be given a physical meaning. So for instance C. N. Yang and R. Mills suggested in their by now classic [569] to restrict the gauge transformations to those for which Ω becomes a constant at infinity. The other option to define conserved currents is offered by (3.79) which in the case of a Yang-Mills theory becomes

$$C_\theta = g \int_{\partial V} (F_a^{0k} \theta^a) dS_k.$$

These charges are gauge invariant, on the cost of depending on the $\theta(x)$. These are to be restricted by the requirement that the variation of the fields A^a_μ and Φ must vanish on ∂V. From (5.65) this requires

$$D_\mu \theta^a{}_{|\partial V} = 0 \qquad\qquad \theta^a T_a \Phi_{|\partial V} = 0.$$

With the covariant derivative, the field equations (5.75) can be written in terms of the matter currents as

$$D_\nu F_a^{\nu\mu} = -g j^\mu_{\Phi a}.$$

Since $D_\mu D_\nu F_a^{\nu\mu}$ is proportional to $F_{\mu\nu} f^{abc} F_c^{\mu\nu}$, and since this vanishes for compact and semi-simple groups, the matter currents obey covariant divergence relations $D_\mu j^\mu_{\Phi a} \doteq 0$. But these, in general, are not eligible to define charges which are conserved and gauge invariant at the same time. Only if there are asymptotic symmetries, one can for instance get interpretable charges; see [2] for the idea to define the charges with respect to a background gauge field, a technique by now standard in the definition of energy-momentum in generally covariant gravitational theories.

The Yang-Mills-Utiyama Recipe

Finally, let me recapitulate the essence of this subsection in the form of a recipe. At the same time I introduce numerical (complex) coefficients in order to reflect the fact that different people use different definitions of structure coefficients, gauge covariant derivatives etc. Let there be a Lagrangian field theory of fields Φ which is invariant under phase transformations

$$\Phi' = \Omega\Phi = e^{i\alpha\Theta}\Phi \qquad \delta\Phi = i\alpha\Theta\Phi$$

where $\Theta = \theta^a T^a$ and T^a are representation matrices of a N-dimensional Lie algebra $\mathfrak{g}$ defined in terms of its generators $[X^a, X^b] = \beta f^{abc} X_c$ $(a = 1, \dots N)$. Define a covariant derivative

$$D_\mu \Phi := (\partial_\mu + \gamma A_\mu)\Phi$$

in terms of Lie-algebra valued gauge fields $A_\mu = A^a_\mu T^a$. This object transforms as

$$A'_\mu = \Omega A_\mu \Omega^{-1} - \gamma^{-1}(\partial_\mu \Omega)\Omega^{-1},$$

or, infinitesimally

$$\delta_\Theta A_\mu = -\frac{i\alpha}{\gamma}\partial_\mu\Theta + i\alpha[\Theta, A_\mu] \qquad \delta_\theta A^a_\mu = -\frac{i\alpha}{\gamma}\partial_\mu\theta^a + i\alpha\beta\theta^b A^c_\mu f_{bc}{}^a. \quad (5.76)$$

Defining the covariant derivative for components of Lie-algebra valued objects as

$$D_\mu G^a := \partial_\mu G^a - \beta\gamma A^b_\mu G^c f_{bc}{}^a,$$

the variation (5.76) becomes

$$\delta_\theta A^a_\mu = -\frac{i\alpha}{\gamma} D_\mu \theta^a.$$

The field strength

$$F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu + \gamma[A_\mu, A_\nu] \qquad F^a_{\mu\nu} := \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \beta\gamma A^b_\mu A^c_\nu f_{bc}{}^a$$

results from the commutator of covariant derivatives

$$[D_\mu, D_\nu] = \gamma F_{\mu\nu}.$$

Under gauge transformations the field strength transforms as

$$F'_{\mu\nu} = \Omega F_{\mu\nu}\Omega^{-1} \quad \text{or} \quad \delta_\Theta F_{\mu\nu} = i\alpha[\Theta, F_{\mu\nu}] \quad \text{i.e.} \quad \delta_\theta F^a_{\mu\nu} = \alpha\beta\theta^b F^c_{\mu\nu} f_{bc}{}^a.$$

Finally, the YM-field Lagrangian is

$$\mathcal{L}_{YM} = \frac{\gamma^2}{4\, g^2} F^a_{\mu\nu} F_a^{\mu\nu}.$$

5.3.5 Higher-Spin Fields

In general, the previous procedure of constructing free-field action functionals for fields with spin-$^3/_2$ and higher just by writing down the most general Lagrangian compatible with Lorentz invariance and dimensional power counting is not immediately feasible.

Nevertheless, the wave equations (that is the dynamical equations) for free half-integer and integer spin fields fields are known. For half-integer spin fields, early pioneers like P. A. M. Dirac [125] and M. Fierz [185] used two-component Weyl spinors in the van der Waerden notation. As then shown by W. Rarita and J. Schwinger [434], it is more advantageous to introduce four-component spinors which additionally carry Lorentz indices. For massive spin-k fields represented by totally symmetric tensors $\Phi_{\mu_1 \dots \mu_k}$ and massive spin-$(k{+}1/2)$ fields represented by totally symmetric $\Psi_{\mu_1 \dots \mu_k}$, their findings are

$$(\Box + m^2)\phi_{\mu_1 \dots \mu_k} = 0 \qquad (i\gamma_\mu \partial^\mu - m)\psi_{\mu_1 \dots \mu_k} = 0 \tag{5.77a}$$

$$\partial^{\mu_1}\phi_{\mu_1 \dots \mu_k} = 0 \qquad \partial^{\mu_1}\psi_{\mu_1 \dots \mu_k} = 0 \tag{5.77b}$$

$$\eta^{\mu_1\mu_2}\phi_{\mu_1 \dots \mu_k} = 0 \qquad \gamma^{\mu_1}\psi_{\mu_1 \dots \mu_k} = 0. \tag{5.77c}$$

(Here for the half-integer spin fields only the Lorentz indices are exhibited, but the spinor character becomes apparent in the presence of γ-matrices.) The equations (5.77) evidently recover the dynamical equations faithfully for a scalar field φ, a Maxwell and a Proca field where ϕ_μ corresponds to A_μ, and a four-component Dirac field for which the ψ in (5.77) is directly identified with the Dirac spinor. Notice that in the massive spin-1 case, the two (applicable) conditions from (5.77) directly are equivalent to the Proca equations and that in the massless case they are equivalent to the Maxwell equations in the Lorenz gauge. Generically, the first set of equations (5.77a) describes free propagation of the fields, (5.77b) eliminates ghosts from the spectrum to guarantee the positiveness of energy, while the last set of conditions (5.77c) guarantees that one is dealing with a spin-k field and that no lower spins are present.

With the work of V. Bargmann and E. P. Wigner [29] it became clear that the free relativistic field equations are in a one-to-one correspondence with the unitary representations of the isometry group. This insight allows one to adapt the Bargmann-Wigner program to other kinematical groups (and to other dimensions) as well; more about this in the concluding remarks to this chapter. In [390] you can find in detail why and how the different covariant wave-functions (Dirac, Fierz, Rarita-Schwinger, Bargmann-Wigner) can be used to carry the same physical representation of the Poincaré group.

However, even for free higher-spin fields, it turned out to be non-trivial to construct Lagrangians that yield the equations (5.77). In an article in 1939, M. Fierz and W. Pauli [186] considered explicitly spins $s = 2$ and $s = {}^3\!/_2$, both massive and massless. For all these cases they were able to find Lagrangians. They needed to introduce extra ("auxiliary") fields in the Lagrangian which at the very end drop out due to their field equations. In the meantime, Lagrangians for both massive and massless free higher-spin fields are known; for an introduction to this field of research see [54]. But the next difficulty arises if one wishes to construct consistent Lagrangian interaction terms with higher-spin fields. Indeed, there are even no-go theorems which prohibit, in Minkowski spacetime, minimal coupling of a massless $s \geqslant {}^3\!/_2$ particle to the electromagnetic field ([521]), and of massless spin $s \geqslant {}^5\!/_2$ particles to the gravitational field [12].[16] Given that in supergravity spin-${}^3\!/_2$ gravitino(s) and a spin-2 graviton do appear, I will spend a few lines on these cases.

Spin-${}^3\!/_2$ Fields

In the spirit of deriving the field types from the representations of $\mathbf{SL}(\mathbf{2}, \mathbf{C})$, a possible definition of a spin-${}^3\!/_2$ field starts from the product of three spin-${}^1\!/_2$ representations

$$\left(\frac{1}{2},0\right) \otimes \left(\frac{1}{2},0\right) \otimes \left(\frac{1}{2},0\right) = \left(\frac{3}{2},0\right) \oplus \left(\frac{1}{2},0\right) \oplus \left(\frac{1}{2},0\right).$$

Here the spin-${}^3\!/_2$ field is represented by a field totally symmetric in the L-spinor indices. In a similar manner one derives an R-spin-${}^3\!/_2$ field from the product of three spin-${}^3\!/_2{}^*$ representations. This corresponds to the original Fierz construction, which as mentioned before, consequently works with Weyl spinors. But there is another option for defining a spin-${}^3\!/_2$ field. Take

$$\left(\frac{1}{2},\frac{1}{2}\right) \otimes \left(\frac{1}{2},0\right) = \left(\frac{3}{2},0\right) \oplus \left(0,\frac{1}{2}\right)$$

by which the field carries a four-vector and a Weyl spinor index. In order to project out the extra term $\left(0, \frac{1}{2}\right)$ one imposes the (Lorentz invariant) condition

$$\sigma^{\mu}\Psi_{\mu L} = 0 \qquad\qquad \text{(respectively} \quad \overline{\sigma}^{\mu}\Psi_{\mu R} = 0).$$

This group product presentation and decomposition underlies the alternative formulation for particles with half-integral spin by W. Rarita and J. Schwinger. The parity eigenstate of the four-component Rarita-Schwinger field is

$$\Psi_{\mu} = \begin{pmatrix} \Psi_{\mu L} \\ \Psi_{\mu R} \end{pmatrix} \qquad \text{with} \quad \gamma^{\mu}\Psi_{\mu} = 0.$$

Rarita and Schwinger also gave a Lagrangian for spin-${}^3\!/_2$ fields. They mention that in the absence of external fields, the Lagrangian is not unique but in any case it leads to

---

[16] Its quite remarkable that these no-go theorems can be circumvented in (A)dS spacetime.

(5.77) for the Rarita-Schwinger field Ψ_μ, and one does not need the acrobatics with additional auxiliary fields of Fierz and Pauli. Today, if one is discussing the Rarita-Schwinger Lagrangian (especially in the supergravity literature), one has in mind

$$\mathcal{L}_{RS} = -\frac{i}{2}\bar{\Psi}_\mu \epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\nu\partial_\rho\Psi_\sigma - m_{3/2}\bar{\Psi}_\mu\sigma^{\mu\nu}\Psi_\nu. \tag{5.78}$$

Contracting the field equations

$$-\frac{i}{2}\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\nu\partial_\rho\Psi_\sigma - m_{3/2}\sigma^{\mu\nu}\Psi_\nu = 0 \tag{5.79}$$

with ∂_μ yields, in the case of non-vanishing mass, $\left[\gamma^\mu\partial_\mu, \gamma^\nu\right]\Psi_\nu = 0$. Furthermore, contracting (5.79) with γ_μ and using the previous result gives

$$\gamma^\mu\Psi_\mu = 0. \tag{5.80}$$

Continuing from the previous two findings:

$$\partial^\mu\Psi_\mu = \frac{1}{2}\{\gamma^\nu\partial_\nu, \gamma^\mu\}\Psi_\mu = -\frac{1}{2}\left[\gamma^\nu\partial_\nu, \gamma^\mu\right]\Psi_\mu = 0. \tag{5.81}$$

Therefore the field equations (5.79) simplify to

$$i\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\nu\partial_\rho\Psi_\sigma = m_{3/2}\Psi_\mu \tag{5.82}$$

and by contraction with $\gamma_\mu\gamma_\lambda$ finally to

$$(\gamma^\lambda\partial_\lambda - m_{3/2})\Psi_\mu = 0 \tag{5.83}$$

with the use of the identity

$$\gamma^{[\mu}\gamma^\nu\gamma^{\rho]} = -i\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\sigma. \tag{5.84}$$

Thus indeed (5.83, 5.81, 5.80) reproduce (5.77) for a spin-$^3/_2$ field in the case of non-vanishing mass $m_{3/2}$. Let us count the number of independent fields: The spinor-vector Ψ_μ has (4×4) components, there are (2×4) conditions (5.81, 5.80) on the fields and thus there are eight degrees of freedom. These correspond to the four possible helicity states $\left(\pm\frac{3}{2}, \pm\frac{1}{2}\right)$.

If $m_{3/2} = 0$ (called now a *gravitino*) the field equations are invariant with respect to gauge transformations

$$\psi_\mu \to \Psi_\mu + \partial_\mu\chi. \tag{5.85}$$

We may impose a gauge choice $\gamma^\nu\Psi_\nu = 0$ which leaves a residual gauge symmetry with $\gamma^\nu\partial_\nu\chi = 0$. Thus what came out as a consistency condition (5.80) in the case of massive spin-$^3/_2$ fields is for massless fields introduced as a gauge condition. (We saw a similar phenomenon in the treatment of the Proca field theory as compared to the Maxwell theory.)

In writing the field equations according to (5.82) and (5.84) as $i\gamma^{[\mu}\gamma^{\nu}\gamma^{\rho]}\partial_\nu\Psi_\rho=0$, with the chosen gauge they read $\gamma^\nu\partial_\nu\Psi^\mu - \gamma^\mu\partial_\nu\Psi^\nu = 0$. Contracting this with γ_μ results in (5.81), and the gravitino Rarita-Schwinger field equations become $\gamma^\nu\partial_\nu\Psi^\mu = 0$. Again all relations (5.81) are derived from the Lagrangian (5.78).

The choice of the Lagrangian (5.78) is not obvious at all (and as mentioned is not the original Rarita-Schwinger Lagrangian). However in the massless case the Lagrangian can be derived from the requirement of invariance with respect to the gauge transformations (5.85): Assume that in analogy with the Weyl fermions, the Lagrangian for the gravitino has the structure $\mathcal{L}_g = \bar{\Psi}_\mu D^{\mu\nu}\Psi_\nu$ with an operator $D^{\mu\nu}$. For dimensional reasons this can only be linear in the derivatives ∂_λ. Further four-vectors at our disposal are the matrices γ^λ. The most general expression that we can write for $D^{\mu\nu}$ is

$$D^{\mu\nu} = a\,\gamma^\mu\partial^\nu + b\,\gamma^\nu\partial^\mu + c\,\eta^{\mu\nu}(\gamma^\lambda\partial_\lambda) + d\,\gamma^\mu\gamma^\nu(\gamma^\lambda\partial_\lambda)$$

with coefficients (a, b, c, d). These are determined from the quest of invariance of the action and the free field equations with respect to (5.85), which requires $a + b = 0$, $a + d = 0$, $b + c = 0$. Therefore only one overall coefficient is left in the Lagrangian $\mathcal{L}_g$. Finally, the terms in $D^{\mu\nu}$ can be arranged with the aid of (5.84) as

$$D^{\mu\nu} \propto \gamma^{[\mu}\gamma^{\rho}\gamma^{\nu]}\partial_\rho.$$

Spin-2 Fields

A spin-2 field can also be defined in various ways, namely in terms of (2, 0), (0, 2), or (1, 1). Opting for the last possibility, this appears in

$$\left(\frac{1}{2},\frac{1}{2}\right) \otimes \left(\frac{1}{2},\frac{1}{2}\right) = [(0,0) \oplus (1,1)]_S \oplus [(0,0) \oplus (1,1)]_A.$$

Here the spin-2 field is represented by a second-rank symmetric tensor $h_{\mu\nu}$. The scalar component (0, 0) is projected out by imposing the zero-trace condition

$$\eta^{\mu\nu}h_{\mu\nu} = 0.$$

Since the massless spin-2 field relates to gravity, a Lagrangian for the $h_{\mu\nu}$ will be derived in Sect. 7.6.1.

5.4 Spontaneous Symmetry Breaking

This section seems to lead away from the endeavor to exploit the power of symmetries in fundamental physics. The term “spontaneous symmetry breaking” refers to a situation in which the Lagrangian and the Hamiltonian still retain those symmetries that were described before, but that some of them are not symmetries of the "envisaged" ground state of the quantum field theory. That is, the "envisaged" ground state can turn out to be the wrong or inappropriate vacuum state.

Spontaneous symmetry breaking is a phenomenon which occurs in statistical mechanics, solid state physics and in elementary particle theory. It appeared already in work by W. Heisenberg on ferromagnets and was investigated around 1960 by Y. Nambu and J. Goldstone in the context of phase transitions in solid state physics. Nambu also pursued ideas how it could be applied to particle physics.[17] So for instance the non-observed (global) chiral symmetry of QuantumChromoDynamics is explained by spontaneous symmetry breaking. Around 1964, P. Higgs and others realized that the spontaneous breaking of a local symmetry is qualitatively different from the spontaneous breaking of a global symmetry. By this mechanism it became possible to formulate the electroweak processes as an originally unbroken Yang-Mills theory; more about this in Sect. 6.3.

5.4.1 Goldstone Bosons

Consider a theory that is defined through an action which is invariant with respect to a global symmetry group **G**. For each symmetry generator there is, according to Noether's first theorem, a conserved current $\partial^\mu j^a_\mu \doteq 0$ and–under appropriate boundary conditions–a conserved charge

$$C^a = \int d^3x \; j^a_0(x)\,. \tag{5.86}$$

In the quantum theory we need to distinguish on whether the Hamiltonian or the vacuum displays this symmetry. In the first case–and this is the one solely dealt with previously–the charges commute with the Hamiltonian:

$$[C^a, H] = 0.$$

This implies that transformations $U = \exp(i\lambda^a C^a)$ leave the Hamiltonian invariant: $UHU^\dagger = H$.

In the second case, we are defining the symmetry on the vacuum. A symmetry is defined as leaving the vacuum $|0\rangle$ invariant if

$$C^a|0\rangle = 0.$$

According to a 1-page article by S. Coleman, we know that "The Invariance of the Vacuum is the Invariance of the World" [95]. But it is not generally true that the symmetries of the "World"–by which Coleman understands the Hamiltonian–is the symmetry of the vacuum. We need to distinguish two cases:

---

[17] Indeed in 2008 Nambu received the Nobel prize in physics "for the discovery of the mechanism of spontaneous broken symmetry in subatomic physics".

- Wigner realization
  In the case

$$C^a|0\rangle = 0 \qquad \text{for all} \quad a$$

  we have the so-called Wigner realization or the unbroken phase. As we saw, two states which are related by a symmetry transformation are degenerate in energy: If $\phi' = U^\dagger \phi U$,

$$\langle 0|\phi'^\dagger H\phi'|0\rangle = \langle 0|U^\dagger \phi^\dagger U H U^\dagger \phi U|0\rangle = \langle 0|\phi'^\dagger H\phi'|0\rangle.$$

  All the states/particles can be decomposed into irreducible multiplets of **G**.
- Goldstone realization
  This is the phase of a spontaneously broken symmetry in which

$$C^a|0\rangle \neq 0 \qquad \text{for at least one} \quad a.$$

  Two states related by a symmetry transformation involving the charges for the broken symmetry are no longer degenerate. Now, there is a theorem by J. Goldstone that in this case massless particles must arise in the spectrum of the theory. These are called Goldstone bosons; with supersymmetry, Goldstone fermions may appear. Here is a sketch of a proof (more details in Chap. 19 of [536]):
  If for a certain conserved charge $C|0\rangle = \int d^3x\, j_0(x) \neq 0$ (here I drop the index a) the state $C|0\rangle$ has infinite norm:

$$\langle 0|CC|0\rangle = \int d^3x\, \langle 0|j_0(x)C|0\rangle = \int d^3x\, \langle 0|j_0(0)C|0\rangle \to \infty.$$

  (The second identity comes about because of translational invariance.) This indicates that the operator C is not well-defined. One way to proceed is to cure/regularize the previous calculation. Another way is to circumvent the problem by taking an operator O that is not invariant under the symmetry transformation generated by C: $[j_0(x), O(y)] \neq 0$. Current conservation allows us to derive

$$\frac{\partial}{\partial x^0}\int d^3x\, \langle 0|\,[j_0(x), O(y)]\,|0\rangle = -\int d^3x\, \langle 0|\left[\partial^i j_i(x), O(y)\right]|0\rangle.$$

  Think of converting the volume integral on the RHS into a surface integral. By choosing the separation between the point x on the surface and the point y sufficiently large, the RHS can be brought to vanish. Hence

$$\frac{\partial}{\partial x^0}\mathcal{O} = 0 \qquad \text{with} \qquad \mathcal{O} = \int d^3x\, \langle 0|\,[j_0(x), O(y)]\,|0\rangle. \quad (5.87)$$

Now insert a complete set of states into the expression for $\mathcal{O}$:

$$\mathcal{O} = \int d^3x \sum \{\langle 0|j_0(x)|n\rangle\langle n|O(y)|0\rangle - \langle 0|O(y)|n\rangle\langle n|j_0(x)|0\rangle\}\,.$$

By using $\langle 0|j_0(x)|n\rangle = \langle 0|j_0(0)|n\rangle e^{-ip_n x}$, the right hand side can be integrated with the result

$$\begin{aligned}\mathcal{O} &= (2\pi)^3 \sum \delta^3(\vec{p}_n)\\ &\quad\times\left\{\langle 0|j_0(0)|n\rangle\langle n|O(y)|0\rangle e^{-iE_n x^0} - \langle 0|O(y)|n\rangle\langle n|j_0(0)|0\rangle e^{iE_n x^0}\right\}.\end{aligned}$$

The delta function selects states $|n\rangle$ with vanishing momenta, that is states with $E_n = m_n$. Since due to (5.87) $\mathcal{O}$ is independent of the time variable, consistency with the last expression requires on the one hand

$$\langle 0|j_0(0)|n\rangle = 0 \qquad \text{for} \qquad m_n \neq 0$$

and on the other hand since, $\mathcal{O} \neq 0$ that there must be states (at least one) in the sum for which $m_n = 0$, but $\langle 0|j_0(0)|n\rangle\langle n|O(y)|0\rangle \neq 0$. These are the Goldstone bosons.

Assume that of the originally $g = \dim \mathbf{G}$ symmetries a number sb are spontaneously broken, and a number u are unbroken. Denote the generators belonging to the unbroken symmetries by U^a. Since the commutator of two unbroken symmetries must be an unbroken symmetry itself, the generators U^a constitute a subalgebra $\mathfrak{u}$ in the Lie algebra $\mathfrak{g}$ of $\mathbf{G}$ with $u = \dim \mathbf{U}$. There are exactly as many Goldstone bosons as there are broken symmetries, namely $sb = \dim \mathbf{G} - \dim \mathbf{U}$.

5.4.2 Nambu-Goldstone Model

A very simple example for spontaneous symmetry breaking is the Nambu-Goldstone model, defined by the Lagrange density of a $\lambda\varphi^4$ theory

$$\begin{aligned}\mathcal{L}_{NG} &= (\partial_\mu\varphi)(\partial^\mu\varphi^*) - V(\varphi,\varphi^*)\\ V(\varphi,\varphi^*) &= a(\varphi\varphi^*) + \frac{\lambda}{2}(\varphi\varphi^*)^2 \cong \frac{\lambda}{2}\,[\varphi\varphi^* - v^2]^2 \quad \text{with} \quad v^2 := -\frac{a}{\lambda}.\end{aligned} \tag{5.88}$$

Here λ and a are free parameters. The parameter λ must be positive, because otherwise the energy is not bounded from below. This theory has a global $\mathbf{U(1)}$ symmetry $\varphi(x) \to e^{i\omega}\varphi(x)$ independent of the actual numerical values of the parameters (a, λ). The kinetic part of the Hamiltonian for this scalar theory was shown to be positive. Thus the ground state is defined by the field value φ_{min} for which the classical potential becomes zero. The minimum of the potential derived

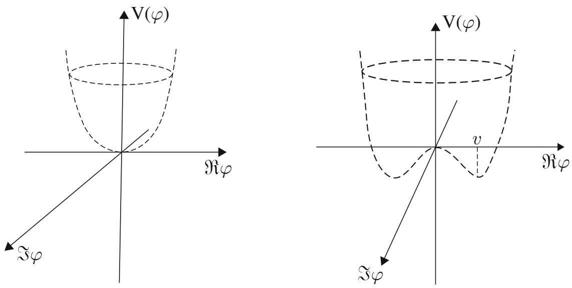


Fig. 5.1 Potentials

from the condition

$$\frac{\partial V}{\partial \varphi} = a\varphi^* + \lambda\varphi^*(\varphi\varphi^*) = 0$$

is located for $a > 0$ at $\varphi_{min} = \varphi^*_{min} = 0$, and for $a < 0$ at

$$| \varphi_{min} |^2 = -\frac{a}{\lambda} = v^2. \tag{5.89}$$

In the former case the potential is a paraboloid with the absolute minimum at $(\Re\varphi, \Im\varphi) = (0, 0)$. In the latter case, the potential has the form of a Mexican hat in the $(\Re\varphi\text{-}\Im\varphi)$-plane with minima on a circle with radius $|\varphi_{min}| = v$ and a maximum at $\varphi_{max} = 0$; see Fig. 5.1. The "ordinary" case with $a = m^2 > 0$, where m can be the interpretated as a mass, is well understood (see Sect. 5.3.2). The free field (Klein-Gordon field) is quantized with creation and annihilation operators, and single-particle states are built by applying creation operators to the ground state $(|0\rangle \cong \varphi = 0)$. All fields are excitations above this vacuum.

In the case of the Mexican-hat type potential, things would go wrong if the same vacuum is chosen. Now $\varphi = 0$ is an unstable point in the potential. The correct vacuum expectation value is determined by the value of $v = \sqrt{-a/\lambda}$:

$$\langle\varphi\rangle_0 := \langle 0|\varphi|0\rangle = \varphi_{min} = v.$$

Here we recover the situation which was described in the previous section: The Lagrangian (and thus the Hamiltonian) is invariant under the symmetry transformation, but the vacuum is not invariant. So where is the Goldstone boson? We need to take into account that the vacuum is displaced. Therefore, instead of the original complex field φ, introduce shifted fields, written in terms of two real fields ϕ_1, ϕ_2 as

$$\varphi(x) = v + \frac{\phi_1(x) + i\phi_2(x)}{\sqrt{2}}. \tag{5.90}$$

Expressed in these fields, the Lagrangian (5.88) becomes

$$\mathcal{L}_{NG} = \frac{1}{2}(\partial_\mu \phi_1)^2 + \frac{1}{2}(\partial_\mu \phi_2)^2 - \lambda v^2 \phi_1^2 + \text{(cubic and quartic terms)}.$$

Here, only ϕ_1 appears quadratically with a non-vanishing coefficient of the correct sign: the field ϕ_1 acquires the mass $m_1 = v\sqrt{2\lambda}$. The field ϕ_2 is the massless Goldstone boson. The original symmetry of the theory disappeared.

There is another route to establish this result: Introduce the real fields (ρ, G) and write

$$\varphi(x) = \frac{1}{\sqrt{2}}(v + \rho(x))e^{iG(x)/v}, \tag{5.91}$$

a parametrization introduced by T. Kibble. This gives

$$\partial_\mu \varphi = \frac{1}{\sqrt{2}} e^{iG(x)/v} \left(\partial_\mu \varrho + \frac{i}{v}(v + \varrho)\partial_\mu G \right)$$

$$(\partial_\mu \varphi)(\partial^\mu \varphi)^* = (1/2)(\partial_\mu \varrho)(\partial^\mu \varrho) + (1/2)\left(\frac{v + \varrho}{v}\right)^2 (\partial_\mu G)(\partial^\mu G)$$

and $\varphi\varphi* = (1/2)(v + \varrho)^2$. Thus the Lagrangian becomes

$$\mathcal{L}_{NG} = \frac{1}{2}(\partial_\mu \varrho)(\partial^\mu \varrho) + \frac{1}{2}\left(\frac{v + \varrho}{v}\right)^2 (\partial_\mu G)(\partial^\mu G) - \frac{\lambda}{4}[2v\varrho + \varrho^2]^2.$$

The potential term produces a mass to the field from the part proportional to ϱ^2, yielding $m_\varrho = v\sqrt{2\lambda}$. The field G does not appear in the potential; it is the massless Goldstone boson.

The appearance of massless Goldstone bosons in case of spontaneous symmetry breaking can be made apparent independent of the previous explicit re-definitions of the fields. Following the argumentation in [412], consider a set of scalar fields φ^α with a Lagrangian $\mathcal{L} = \text{(kinetic terms)} - V(\varphi)$. Denote by φ_0^α the field configuration minimizing $V(\varphi)$:

$$\frac{\partial}{\partial \varphi^\alpha} V \Big|_{\varphi(x) = \varphi_0} = 0.$$

Next expand $V(\varphi)$ around its minimum

$$V(\varphi) = V(\varphi_0) + \frac{1}{2}(\varphi - \varphi_0)^\alpha (\varphi - \varphi_0)^\beta \left(\frac{\partial^2}{\partial \varphi^\alpha \partial \varphi^\beta} V \right)_{\varphi_0} + \ldots$$

The coefficient of the quadratic term is a symmetric matrix

$$m^2_{\alpha\beta} := \left(\frac{\partial^2}{\partial \varphi^\alpha \partial \varphi^\beta} V \right)_{\varphi_0}$$

whose eigenvalues give the masses of the fields. The eigenvalues cannot be negative since φ_0 constitutes a minimum. Next assume that the model has a symmetry with respect to (infinitesimal) transformations $\varphi^\alpha \to \varphi^\alpha + \epsilon \mathcal{A}^\alpha(\varphi)$. Specializing to constant fields, the kinetic part of the Lagrangian becomes invariant and the invariance of the potential manifests itself by

$$V(\varphi) = V(\varphi + \epsilon \mathcal{A}(\varphi)) \qquad \text{or explicitly} \qquad \mathcal{A}^\alpha \frac{\partial}{\partial \varphi^\alpha} V = 0.$$

Now differentiate with respect to φ^γ:

$$0 = \left(\frac{\partial \mathcal{A}^\alpha}{\partial \varphi^\gamma}\right)\left(\frac{\partial V}{\partial \varphi^\alpha}\right) + \mathcal{A}^\alpha \left(\frac{\partial^2 V}{\partial \varphi^\alpha \partial \varphi^\gamma}\right).$$

Restricting this to φ_0^γ, the first term vanishes at the minimum and the vanishing of the second term amounts to the condition

$$0 = \mathcal{A}^\alpha(\varphi_0)\, m^2_{\alpha\gamma}.$$

If the ground state is invariant, we have $\mathcal{A}^\alpha(\varphi_0) = 0$ for all α, and the condition is trivially fulfilled. In case $\mathcal{A}^{\bar{\alpha}}(\varphi_0) \neq 0$ for an index $\bar{\alpha}$, this is an eigenvector of m^2 with eigenvalues zero, leading to as many massless gauge bosons as the number of spontaneous broken symmetries.

5.4.3 Higgs Mechanism

As discovered by P. Higgs and some others, gauge theories do have the capability to “digest” a massless Goldstone boson and to “excrete” a massive boson. This will be explained first for an Abelian gauge theory, then for a specific non-Abelian theory, and finally extended to a general non-Abelian gauge theory.

Abelian Gauge Theory

Consider the coupling of the $\lambda\varphi^4$ theory (5.88) to an electromagnetic field. The Lagrange density is

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (D_\mu \varphi)(D_\mu \varphi)^* - \frac{\lambda}{2} [\varphi\varphi^* - v^2]^2.$$

For $v^2 < 0$ the vacuum is again determined by (5.89). Thus, instead of the original complex field φ introduce the real fields (ϱ, G) in the form (5.91). From

$$D_\mu \varphi = (\partial_\mu - ieA_\mu)\varphi = \frac{1}{\sqrt{2}} e^{iG(x)/v} (\partial_\mu \varrho - ie(v + \varrho) B_\mu)$$

with

$$B_\mu := A_\mu - \frac{1}{ev}\partial_\mu G \tag{5.92}$$

we rewrite the Lagrangian in the new fields as

$$\begin{aligned}
\mathcal{L} &= -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \frac{1}{2}(\partial_\mu\varrho)(\partial^\mu\varrho) + \frac{1}{2}e^2(v+\varrho)^2 B_\mu B^\mu - \frac{\lambda}{4}(2v\varrho + \varrho^2)^2 \\
&= \left(-\frac{1}{4}B_{\mu\nu}B^{\mu\nu} + M_B^2 B_\mu B^\mu\right) \\
&\quad + \left(\frac{1}{2}(\partial_\mu\varrho)(\partial^\mu\varrho) - \lambda v^2\varrho^2 + \text{terms cubic and quartic in } \varrho\right) \\
&\quad + \frac{1}{2}e^2(\varrho^2 + 2v\varrho)B_\mu B^\mu
\end{aligned}$$

with $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. This is a remarkable result for various reasons. First of all, observe that the field G does not appear at all. In fact, (5.92) is nothing but a gauge transformation. Instead, the gauge transformed vector field B_μ becomes massive, since the first term in brackets is the Proca Lagrangian (5.55), and we read off the mass ($M_B = ev/\sqrt{2}$). The second term is the Lagrangian part of a scalar field with mass m_ϱ^2. The last term describes the interaction between the scalar and the vector field. This phenomenon, that a would-be Goldstone is "gauged away" and gives rise to a massive gauge boson, was discovered in 1964 more or less independently by R. Brout and F. Englert, by P. Higgs, and by G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble.[18]

If we compare this case of a local symmetry with the previous example of a global symmetry, we recall:

- In the case of a local symmetry (Higgs mode), a theory with two "massive" fields (keep in mind: $v^2 < 0$) and a massless gauge boson can be rewritten as a theory with one massive scalar field and one massive vector field.
- In the case of global symmetry (Goldstone mode), the original theory has two "massive" fields and this is equivalent to a theory with one massive and one massless scalar field.

In both cases, the balance of degrees of freedom is preserved, namely $(2+2) \leftrightarrow (1+3)$ and $2 \leftrightarrow (1+1)$.

"Higgs-Kibble Model"

With regard to the Glashow-Salam-Weinberg model, it will now be shown with the example of a non-Abelian gauge theory that it is possible to spontaneously break a local symmetry in such a way that it is partially preserved and not all gauge bosons

[18] All six physicists were jointly awarded the 2010 J. J. Sakurai Prize for Theoretical Particle Physics.

become massive. The starting point for this example is a $\mathbf{SU(2)}$ gauge theory coupled to an iso-scalar triplet ϕ with Lagrange density

$$\mathcal{L} = -\frac{1}{4}\, F^a_{\mu\nu} F^{a\mu\nu} - \frac{1}{2}(D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi). \tag{5.93}$$

We assume that the minimum of the potential is not located at $\phi_\alpha = 0$, but that the vacuum expectation value is

$$\langle \phi \rangle_0 = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The choice that only the third component is different from zero is without loss of generality since we can always rotate the scalar triplet into such a configuration. From the two previous examples, we anticipate how to introduce new fields as

$$\phi = \frac{v}{\sqrt{2}}(v + \varrho) e^{i(G^1\, T^1 + G^2\, T^2)/v} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$B^k_\mu = A^k_\mu - \frac{1}{gv} \partial_\mu G^k \quad (k = 1, 2), \qquad B^3_\mu = A^3_\mu.$$

The Lagrange density (5.93) becomes

$$\begin{aligned} \mathcal{L} = &-\frac{1}{4} B^a_{\mu\nu} B^{a\mu\nu} + \frac{1}{2} g^2 v^2 (B^1_\mu B^{1\mu} + B^2_\mu B^{2\mu}) + \frac{1}{2}(\partial_\mu \varrho)(\partial^\mu \varrho) - \lambda v^2 \varrho^2 \\ &+ \text{ terms of higher order in the fields } \ (\varrho, B^1, B^2). \end{aligned} \tag{5.94}$$

The fields G^1 and G^2 no longer appear, and the fields B^1 and B^2 become massive. The only symmetry remaining is a local $\mathbf{U(1)}$ symmetry

$$B^3_\mu \mapsto e^{i\alpha(x)} B^3_\mu \qquad \varrho \mapsto \varrho.$$

Non-Abelian Gauge Theory

The Higgs mechanism works for other gauge theories as well, and next we will derive the fact that the number of massive vector bosons is the same as the number of broken generators of the original gauge group. For this purpose, consider a field theory with scalars ϕ_α which is invariant under global transformations

$$\phi_\alpha \to (1 + i\epsilon^a T^a)_{\alpha\beta} \phi_\beta.$$

The fields may, just for the convenience in the notation, be considered to be real-valued by interpreting for instance n complex as $2n$ real fields. Then the representation matrices T^a may advantageously be rewritten as $(T^a)_{\alpha\beta} = i(\bar{T}^a)_{\alpha\beta}$, where the $\bar{T}^a$ are real and antisymmetric. Promoting the global symmetry to a local one by introducing gauge fields and the covariant derivative $D_\mu \phi = (\partial_\mu + gA^a_\mu \bar{T}^a)\phi$. The kinetic energy term for the scalar field thus becomes

$$\begin{aligned}\frac{1}{2}(D_\mu \phi_\alpha)(D^\mu \phi_\alpha) =& \frac{1}{2}(\partial_\mu \phi_\alpha)^2 + gA^a_\mu (\partial_\mu \phi_\alpha)(\bar{T}^a)_{\alpha\beta}\phi_\beta \\ &+ \frac{1}{2} g^2 A^a_\mu A^{b\mu} (\bar{T}^a \phi)_\alpha (\bar{T}^b \phi)_\beta. \end{aligned} \tag{5.95}$$

Now assume that the potential in the scalar theory is such that the ϕ_α acquire vacuum expectation values $\langle \phi_\alpha \rangle = (\phi_0)_\alpha$. Think of expanding the scalar fields in the kinetic term around this value. Then the last term in (5.95) acquires a contribution

$$\frac{1}{2} g^2 \, A^a_\mu A^{b\mu} (\bar{T}^a \phi_0)_\alpha (\bar{T}^b \phi_0)_\beta := \frac{1}{2} \, m^2_{ab} A^a_\mu A^{b\mu},$$

where

$$m^2_{ab} = g^2 (\bar{T}^a \phi_0)_\alpha (\bar{T}^b \phi_0)_\beta$$

can be interpreted as a mass matrix (observe that it is semi-definite). Thus, in general, the gauge bosons become massive, except those for which the transformations leave the vacuum invariant, i.e. $\bar{T}^c \phi_0 = 0$. The second term in the expression (5.95) inherits a term

$$gA^a_\mu (\partial_\mu \phi_\alpha)(\bar{T}^a \phi_0)_\alpha.$$

This vanishes if all transformations are symmetries of the vacuum. Otherwise it constitutes an interaction term of the gauge bosons with the surviving Goldstone bosons.

5.5 Discrete Symmetries

5.5.1 General Preliminary Remarks and Definition of Terms

(A)

In this section, three types of discrete symmetries are considered: Space inversions P, time reversal T, and charge conjugation C. According to the Wigner theorem, it will turn out that P and C are associated with linear unitary operators, whereas T is associated with an anti-linear and anti-unitary operator.

Up to the 1950's it was assumed (without really questioning) that "physics" is symmetric under each of these three discrete symmetries. With the discovery of violation of parity (T. D. Lee and C. N. Yang, 1956; C-S. Wu *et al.*, 1956) and of the combined CP-transformation (J. Cronin and V. Fitch, 1964) in certain weak interaction processes, it was necessary to concede that these symmetries are not exact. However, under rather mild and general assumptions one can prove the CPT-theorem, i.e. the assertion that the product of the three discrete operations is an exact symmetry. This is the reason why the CP violation is interpreted as a violation of T.

(B)

Each of the symmetries in $S \in \{P, T, C\}$ constitutes a group isomorphic to $\mathbf{Z_2}$, with the composition $S \circ S = I$. Due to the Wigner theorem (see Subsect. 4.2.1), either a linear and unitary or an anti-linear and anti-unitary operator U_S with $U_S U_S = \alpha I$ is associated to each transformation, where α is a phase factor. In the following, the operators are labeled with the same symbol as the transformations, but written in bold face font: $\mathbf{S} = U_S$. Thus

$$\mathbf{SS} = \alpha\mathbf{I}.$$

On the other hand, $\mathbf{SS}^{-1} = \mathbf{I}$ and thus

$$\alpha\mathbf{I} = \mathbf{S}^{-1}\alpha\mathbf{S}.$$

Choosing for $\mathbf{S}$ either $\mathbf{P}$ or $\mathbf{C}$, this relation is trivially fulfilled for any phase, since in this case the unitary $\mathbf{S}$ commutes with α. However, for $\mathbf{S} = \mathbf{T}$, because of $\mathbf{T}^{-1}\alpha = \alpha^*\mathbf{T}^{-1}$ (compare (4.1)), we have

$$\mathbf{TT} = \pm\mathbf{I}.$$

The fact that there are two distinct choices in sign plays a role in the distinction of bosons and fermions: bosons belong to the sector with $\mathbf{TT} = +\mathbf{I}$, fermions to the sector with $\mathbf{TT} = -\mathbf{I}$.

(C)

In the following, for all three symmetries we will consider expressions of the form

$$\mathbf{S}\Phi(x)\mathbf{S}^{-1} = \mathbb{S}_\Phi \Phi(\mathcal{S}x). \tag{5.96}$$

Here Φ stands for some field, $\mathbb{S}_\Phi$ is an operation on the field (and, as we will see, depends on the nature of the field), and $\mathcal{S}x$ denotes the action of a symmetry operator S on spacetime coordinates, i.e. $\mathcal{P}x = (t, -\vec{x})$, $\mathcal{T}x = (-t, \vec{x})$, $\mathcal{C}x = (t, \vec{x})$.[19] From this

$$\Phi_S(x') = \mathbb{S}_\Phi \Phi(x). \tag{5.97}$$

[19] I apologize for this abundance of letters S used with different fonts. It may be pedantic, but we must conceptually distinguish the symmetry transformations from their unitary representations and their realizations as e.g. matrix multiplication on the fields. The expression (5.96) is comparable to (5.21) and (5.22).

It goes without saying that a Lagrangian which is invariant under any of these transformations obeys

$$\mathbf{S}\mathcal{L}(x)\mathbf{S}^{-1} = \mathcal{L}(\mathcal{S}x).$$

5.5.2 Space Inversion P

The space inversion $\mathcal{P}$ is a transformation by which the coordinates of space-time change according to $\mathcal{P}\colon (t, \vec{x}) \to (t, -\vec{x})$. This transformation is an element of the Lorentz group, given by the matrix

$$(\mathcal{P}^{\mu}{}_{\nu}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{5.98}$$

Prior to the discovery of parity non-conservation it was an unquestioned assumption that a unitary (or anti-unitary operator) $\mathbf{P} \equiv U(\mathcal{P}, 0)$ exists, such that

$$\mathbf{P}U(\Lambda, a)\mathbf{P}^{-1} = U(\mathcal{P}\Lambda\mathcal{P}^{-1}, \mathcal{P}a) \tag{5.99}$$

for any proper orthochronous Lorentz transformation $\Lambda^{\mu}{}_{\nu}$ and any translation a^{μ}. Since 1956, we know that for weak interaction processes this holds only approximately.

In order to find out, whether $\mathbf{P}$ is unitary or anti-unitary, we restrict U to infinitesimal translations, that is $U(1, \epsilon) = 1 + i\epsilon_{\mu}P^{\mu} + \ldots$. For this specific transformation (5.99) can be analyzed by the help of (5.1b) to yield

$$\mathbf{P}iP^{\mu}\mathbf{P}^{-1} = i\mathcal{P}^{\mu}_{\ \nu}P^{\nu}. \tag{5.100}$$

Specializing this to $P^0 = H$ results in $\mathbf{P}iH\mathbf{P}^{-1} = i\mathcal{P}^0_{\ \nu}P^{\nu} = iH$. If $\mathbf{P}$ were anti-unitary and anti-linear it would anti-commute with the imaginary unit i, which would yield $\mathbf{P}H\mathbf{P}^{-1} = -H$. Then for each state $|\psi\rangle$ with positive energy E (that is $H|\psi\rangle = E|\psi\rangle$), there would exist a state $\mathbf{P}^{-1}|\psi\rangle$ with negative energy, because of $H\mathbf{P}^{-1}|\psi\rangle = -\mathbf{P}^{-1}H|\psi\rangle = -E\mathbf{P}^{-1}|\psi\rangle$. Since this does not make sense physically, $\mathbf{P}$ must be unitary and linear. This established, the relation (5.100) tells us that $\mathbf{P}P^k\mathbf{P}^{-1} = \mathcal{P}^k_{\ \nu}P^{\nu} = -P^k$. And this is consistent with Newtonian mechanics, in which the momentum $\vec{p}$ is a polar vector. It is also compatible with the canonical commutator of quantum mechanics: $[\mathbf{P}x^k, \mathbf{P}p^j] = [-x^k, -p^j] = i\delta^{kj}$.

Evaluating (5.99) for infinitesimal Lorentz transformations by taking recourse to (5.1b) yields

$$\mathbf{P}\,M^{\mu\nu}\,\mathbf{P}^{-1} = \mathcal{P}^{\mu}{}_{\varrho}\mathcal{P}^{\nu}{}_{\sigma}\,M^{\varrho\sigma} \quad \text{or} \quad \mathbf{P}\vec{J}\mathbf{P}^{-1} = +\vec{J}, \quad \mathbf{P}\vec{K}\mathbf{P}^{-1} = -\vec{K}. \tag{5.101}$$

These relations are reasonable, since in Newtonian mechanics the angular momentum $\vec{J}$ transforms as an axial vector due to $\vec{J} = \vec{r} \times \vec{p}$, and since a Galilei boost $\vec{K} = \vec{v}t$ changes its orientation after space inversion.

Next one may consider the action of the operator $\mathbf{P}$ on single-particle states, and then extend it to the Fock space spanned by operating with further creation operators on the ground states in each case of fields (scalar, spinor, vector,...) in order to find their transformation behavior. This formal path is rigorously followed in Weinberg's QFT textbook.

Here instead let us argue by looking at the physics involved. We can for instance bring forward the argument that each term in a Lagrangian should behave like a scalar under the parity transformation. As we have seen, gauge invariance requests that amongst other things, we have a term $j_\mu A^\mu$ in the Lagrangian, where A^μ is a vector gauge potential and j_μ a current. Under a parity transformation the zero component of the current, that is a charge distribution, does not change its sign. But $\vec{j} \to -\vec{j}$ because the direction of the current is reversed. Therefore

$$\mathbf{P} j_\mu(x) \mathbf{P}^{-1} = \mathcal{P}_\mu^{\ \nu} j_\nu(\mathcal{P}x), \qquad \mathbf{P} A^\mu(x) \mathbf{P}^{-1} = \mathcal{P}^\mu_{\ \nu} A^\nu(\mathcal{P}x). \tag{5.102}$$

Then $\mathbf{P} j_\mu A^\mu(x) \mathbf{P}^{-1} = j_\mu A^\mu(\mathcal{P}x)$. The vector field changes as

$$A'^\mu(t, -\vec{x}) = \mathcal{P}^\mu_{\ \nu} A^\nu(t, \vec{x}),$$

which means $\mathbb{P}_A A^\mu = \mathcal{P}^\mu_{\ \nu} A^\nu$ in the notation of (5.96). Indeed from electrodynamics it is known that the vector potential transforms as a polar vector (which in turn implies that the magnetic field behaves as an axial vector). Noticing that $\partial_\mu \to (\partial_t, -\vec{\partial}_k)$, a field strength transforms as $\mathbf{P} F^{\mu\nu}(x) \mathbf{P}^{-1} = \mathcal{P}_\mu^{\ \rho} \mathcal{P}_\nu^{\ \sigma} F^{\rho\sigma}(\mathcal{P}x)$.

We next determine the behavior of spinors under parity transformations by two different lines of reasoning. In the first we start from Weyl spinors. Remember that these were introduced as the smallest non-trivial representations of the $\mathbf{SU(2)}$–algebras generated by A^j and B^j, respectively, where $\vec{A}$ and $\vec{B}$ are determined from $\vec{J}$ and $\vec{K}$ according to (5.19). From (5.101), we observe the transformation behavior

$$\mathbf{P} \vec{A} \mathbf{P}^{-1} = \vec{B}, \qquad \mathbf{P} \vec{B} \mathbf{P}^{-1} = \vec{A}, \tag{5.103}$$

which means that under space inversion, the object $\vec{A}$ is transformed into $\vec{B}$ and *vice versa*. That is, a parity transformation is equivalent to the interchange of the $(\frac{1}{2}, 0)$-with the $(0, \frac{1}{2})$-representation. This is exactly the reason why parity-conserving QED is written in terms of Dirac spinors as representation of $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$. Equations (5.103) imply for the spinor components (5.36) that

$$\mathbf{P} : \xi_\alpha(x) \to \bar{\eta}^{\dot{\alpha}}(\mathcal{P}x) \qquad \mathbf{P} : \bar{\eta}^{\dot{\alpha}}(x) \to \xi_\alpha(\mathcal{P}x).$$

In terms of Dirac spinors this transformation rule becomes

$$\mathbf{P}\psi(x)\mathbf{P}^{-1} = \gamma_0\psi(\mathcal{P}x) \tag{5.104}$$

which in the notation of (5.96) identifies the operation $\mathbb{P}_\psi = \gamma^0\psi$.

Another way to arrive at this result is by multiplying the Dirac equation by γ^0 from the left and then rearranging the expression as

$$\begin{aligned}\gamma^0(i\partial\!\!\!/ - m)\psi(x) &= \gamma^0(i\gamma^0\partial_0 + i\gamma^k\partial_k - m)\psi(x)\\ &= (i\gamma^0\partial_0 - i\gamma^k\partial_k - m)\gamma^0\psi(x) = (i\gamma^\mu\partial'_\mu - m)\gamma^0\psi(x) = 0,\end{aligned}$$

where $\partial'_\mu = \partial/\partial x'^\mu = \mathcal{P}^\nu{}_\mu\partial_\nu$. Therefore $\psi'(x') = \eta_P\gamma^0\psi(x)$ satisfies the Dirac equation in the space-inverted world. Here, η_P is a phase (which we have set to one), and we have arrived back at (5.104).

As we will see in Subsect. 5.5.4, the anti-particle related to ψ is (in the chiral basis) defined by $\psi^C(x) := i\gamma^2\psi^*$. Under a parity transformation:

$$\mathbf{P}\psi^C(x)\mathbf{P}^{-1} = \mathbf{P}i\gamma^2\gamma^{0*}\psi^*(x)\mathbf{P}^{-1} = -\gamma_0\psi^C(\mathcal{P}x),$$

showing that particle and anti-particle indeed have opposite parity.

Having found the transformation behavior of spinors with respect to space inversions, we might derive how the Dirac bilinears $\overline{\psi}\Gamma\psi$ built from $\Gamma = \{1, \gamma^5, \gamma^\mu, \gamma^5\gamma^\mu, \gamma^{\mu\nu}\}$ transform. Since

$$\overline{\psi}'(t, -\vec{x}) = \eta_P^*\overline{\psi}(t, \vec{x})\gamma^0,$$

we derive

$$\begin{aligned}\mathbf{P}\overline{\psi}\psi(x)\mathbf{P}^{-1} &= +\overline{\psi}\psi(\mathcal{P}x)\\ \mathbf{P}\overline{\psi}\gamma^5\psi(x)\mathbf{P}^{-1} &= -\overline{\psi}\gamma^5\psi(\mathcal{P}x)\\ \mathbf{P}\overline{\psi}\gamma^\mu\psi(x)\mathbf{P}^{-1} &= \mathcal{P}^\mu{}_\nu\,\overline{\psi}\gamma^\nu\psi(\mathcal{P}x)\\ \mathbf{P}\overline{\psi}\gamma^5\gamma^\mu\psi(x)\mathbf{P}^{-1} &= -\mathcal{P}^\mu{}_\nu\,\overline{\psi}\gamma^5\gamma^\nu\psi(\mathcal{P}x)\\ \mathbf{P}\overline{\psi}\gamma^{\mu\nu}\psi(x)\mathbf{P}^{-1} &= \mathcal{P}^\mu{}_\rho\mathcal{P}^\nu{}_\sigma\,\overline{\psi}\gamma^{\rho\sigma}\psi(\mathcal{P}x).\end{aligned} \tag{5.105}$$

Observe that according to these relations the current $j^\mu = \overline{\psi}\gamma^\mu\psi$ transforms consistent to what was found previously by (5.102).

The absolute internal parity, that is the phase η_P for a particle/field, can be fixed only under specific circumstances: We will see in the next chapter for instance that mesons are bound states of a quark and an anti-quark. Quarks are fermions, so the relative parity of a quark and an anti-quark is (-1). As bound states with total angular momentum J they have the parity $(-1)^{J+1}$. Mesons in the s-state are labeled as $J^P = 0^-$ and identified with the (pseudo scalar) π mesons, those in the p-state are labeled as $J^P = 1^+$, identified with the scalar ϱ mesons. These two states inhabit two different octets in the "pre-QCD" quark model.

5.5.3 Time Reversal T

Time reversal $\mathcal{T} : (t, \vec{x}) \to (-t, \vec{x})$ is an element of the Lorentz group, in the four-dimensional representation given by

$$\mathcal{T}^\mu_{\ \nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Assuming the conservation of CPT and given the experimental results of CP-violation in certain weak interaction processes, the relation

$$\mathbf{T}U(\Lambda, a)\mathbf{T}^{-1} = U(\mathcal{T}\Lambda\mathcal{T}^{-1}, \mathcal{T}a) \tag{5.106}$$

with $\mathbf{T} \equiv U(\mathcal{T}, 0)$ is valid only approximately. In any case, we now need to investigate the pendant to (5.100), or

$$\mathbf{T}iP^\mu\mathbf{T}^{-1} = i\mathcal{T}^\mu_{\ \nu}P^\nu.$$

Specializing this to $P^0 = H$ results in $\mathbf{T}iH\mathbf{T}^{-1} = -iH$. If $\mathbf{T}$ were unitary and linear, the operator would commute with the factor i, with the result that $\mathbf{T}H\mathbf{T}^{-1} = -H$, and we would arrive at a conclusion about states with negative energy similar to the argumentation in the case of $\mathbf{P}$-transformations. Thus, $\mathbf{T}$ is anti-unitary and anti-linear. Additionally one derives

$$\mathbf{T}\vec{P}\mathbf{T}^{-1} = -\vec{P}, \qquad \mathbf{T}\vec{J}\mathbf{T}^{-1} = -\vec{J}, \qquad \mathbf{T}\vec{K}\mathbf{T}^{-1} = +\vec{K}.$$

This is consistent with the transformation of momenta (being proportional to velocities), angular momenta (containing the momenta) and Galilei boosts (being proportional to a velocity and to time). The property of $\mathbf{T}$ being anti-unitary is also consistent with the observation that the Schrödinger equation is invariant with respect to time-reversal:

$$\mathbf{T}\left(i\frac{\partial\Psi(t,\vec{x})}{\partial t} - H\Psi(t,\vec{x})\right)\mathbf{T}^{-1} = i\frac{\partial\Psi(-t,\vec{x})}{\partial t} - H\Psi(-t,\vec{x}).$$

In considering the action of $\mathbf{T}$ on single-particle states, we must make allowance for a phase factor η_T—similar to the case of space inversions $\mathbf{P}$. This factor can, however, be eliminated by redefining the states, that is by reabsorbing the phases. Therefore, no intrinsic conserved quantities ("time parity") exists. This in turn is consistent with basic postulates of quantum physics since $\mathbf{T}$ as an anti-unitary operator does not have an associated Hermitean observable.

The transformation behavior of a vector field with respect to time reversal will once more be derived from considering the term $j_\mu A^\mu$ in the Lagrangian. If time is reversed, the zero component of the current is unchanged but the direction of $\vec{j}$ is changed. Therefore

$$\mathbf{T} j_\mu(x)\mathbf{T}^{-1} = -\mathcal{T}_\mu{}^\nu j_\nu(\mathcal{T}x), \quad \mathbf{T}A^\mu(x)\mathbf{T}^{-1} = -\mathcal{T}^\mu_\nu A^\nu(\mathcal{T}x). \tag{5.107}$$

As for the Dirac field we will, similarly to the case of space inversion, derive a matrix $\mathbb{T}_\psi$ such that $\psi'(x') = \mathbb{T}_\psi \psi(x)$ is a solution of the Dirac equation in the time-reversed world $x' = (-t, \vec{x})$ if $\psi(x)$ is a solution in the original system. Write the Dirac equation as $i\gamma^0\partial_0\psi(x) = -(i\gamma^k\partial_k + m)\psi(x)$ and multiply it from the left with γ^0 to obtain

$$i\partial_0\psi(x) = H\psi(x) \quad \text{with} \quad H = -\gamma^0(i\gamma^k\partial_k + m) = H'.$$

The requirement that $i\partial_0'\psi'(x') = H\psi'(x')$ leads to $-i\mathbb{T}_\psi\partial_0\psi(x) = H\mathbb{T}_\psi\psi(x)$ because of $(\partial_0' = -\partial_0)$ and–by making use of the fact that $\psi(x)$ is a solution of the original Dirac equation–to $i\mathbb{T}_\psi iH = H\mathbb{T}_\psi$. Since time reversal is mediated by an anti-unitary operator, we arrive at the conclusion that the matrix $\mathbb{T}_\psi$ commutes with the Hamiltonian. Therefore $\mathbb{T}_\psi i\gamma^0\gamma^k \stackrel{!}{=} i\gamma^0\gamma^k\mathbb{T}_\psi$ and $\mathbb{T}_\Psi\gamma^0 \stackrel{!}{=} \gamma^0\mathbb{T}_\psi$. Writing $\mathbb{T}_\psi = UK$, these two conditions amount to

$$\gamma^0 U \stackrel{!}{=} U\gamma^{*0} \quad \text{and} \quad \gamma^k U \stackrel{!}{=} -U\gamma^{*k}.$$

In both the Dirac basis (B.10) and the Weyl basis (B.11), the only imaginary gamma-matrix is γ^2 (since σ^2 is the only imaginary Pauli matrix). From this observation we guess (and verify) that

$$\mathbb{T}_\psi = \gamma^1\gamma^3 K \tag{5.108}$$

(up to a phase). The presence of the complex-conjugate operator K in $\mathbb{T}_\psi$ solves the following "mystery": Since $\partial'_\mu = \mathcal{T}_\mu{}^\nu\partial_\nu$ one may expect that in order for $\gamma^\mu\partial_\mu$ to transform like a scalar, $\gamma'^\mu = \mathcal{T}^\mu_\nu\gamma^\nu$. Instead, one finds $\mathbb{T}_\Psi\gamma^\mu\mathbb{T}_\psi^{-1} = -\mathcal{T}^\mu_\nu\gamma^\nu$, that is the opposite sign. The dilemma is solved in that $i(\gamma^\mu\partial_\mu)$ is a scalar with respect to time reversal.

The conjugate spinor transforms as

$$\psi^\dagger K\gamma^{3\dagger}\gamma^{1\dagger}\gamma^0 = \overline{\psi}K\gamma_3\gamma_1.$$

From this one derives the transformation behavior of the Dirac bilinears:

$$\begin{aligned}
\mathbf{T}\overline{\psi}\Gamma\psi(x)\mathbf{T}^{-1} &= +\overline{\psi}\Gamma\psi(\mathcal{T}x) \qquad\qquad \text{for} \quad \Gamma = (1, \gamma^5)\\
\mathbf{T}\overline{\psi}\gamma^\mu\psi(x)\mathbf{T}^{-1} &= -\mathcal{T}^\mu_\nu\ \overline{\psi}\gamma^\nu\psi(\mathcal{T}x)\\
\mathbf{T}\overline{\psi}\gamma^\mu\gamma^5\psi(x)\mathbf{T}^{-1} &= -\mathcal{T}^\mu_\nu\ \overline{\psi}\gamma^\nu\gamma^5\psi(\mathcal{T}x)\\
\mathbf{T}\overline{\psi}\gamma^{\mu\nu}\psi(x)\mathbf{T}^{-1} &= -\mathcal{T}^\mu_\rho\mathcal{T}^\nu_\sigma\ \overline{\psi}\gamma^{\rho\sigma}\psi(\mathcal{T}x).
\end{aligned} \tag{5.109}$$

Specifically $\overline{\psi}\gamma^\mu\psi$ transforms as a vector, compatible with (5.107) since it is the current j^μ of QED.

It was mentioned at the beginning of this subsection that the time-reversal operator because of its anti-unitarity obeys either $\mathbf{TT} = +I$ or $\mathbf{TT} = -I$. And we observe indeed that the former is the case for bosonic fields, since these directly transform with the matrix $\mathcal{T}^\mu_\nu$, and that for fermions $\mathbb{TT} = \gamma^1\gamma^3\gamma^1\gamma^3 = -I$.

5.5.4 Charge Conjugation C

The discrete symmetry C is not (like P- and T-transformations) related to Lorentz transformations, but rather to internal symmetries. A more correct denomination would be "particle—antiparticle" transformation. Nevertheless the name "charge" conjugation is justified in that all elementary particles–in compliance with the three basic interactions–do have three variants of charge, namely an electric charge, a weak charge ("weak isospin"), and a color charge (details will follow in the next chapter).

The explicit form of charge conjugation can be directly derived by starting with the Dirac equation of a charged particle in an electromagnetic field:

$$[i\gamma^\mu(\partial_\mu - ieA_\mu) - m]\psi = 0. \tag{5.110}$$

The antiparticle described by the Dirac spinor ψ^C has opposite charge and thus obeys

$$[i\gamma^\mu(\partial_\mu + ieA_\mu) - m]\psi^C = 0.$$

The opposite sign for the charge also arises if we complex conjugate (5.110)

$$[-i\gamma^{*\mu}(\partial_\mu + ieA_\mu) - m]\psi^* = 0.$$

Here, the complex conjugates of the gamma matrices emerge. The $(-\gamma^{*\mu})$ also obey the Clifford algebra (B.5). Therefore a similarity transformation exists such that $(-\gamma^{*\mu}) = S^{-1}\gamma^\mu S$. If you plug this into the previous expression with the complex-conjugate field (and multiply with the matrix S from the left) you arrive at

$$[i\gamma^\mu(\partial_\mu + ieA_\mu) - m]S\psi^* = 0,$$

and $S\psi^*$ can be identified with ψ^C. For diverse reasons it has become standard to write the matrix $S = C\gamma^0$ with an explicit γ^0:

$$\psi^C = C\gamma^0\psi^*.$$

The defining equation for C is $-\gamma^{*\mu} = (C\gamma^0)^{-1}\gamma^\mu(C\gamma^0)$. In Appendix B.1 it is shown that this is solved by $C = \eta\gamma^2\gamma^0$. If one chooses $\eta = i$, the matrix C obeys $C^2 = -1$ and $C^\dagger = C^T = -C$. Thus

$$\mathbb{C}_\psi = i\gamma^2\, K.$$

For the Dirac bilinears, one finds

$$\begin{aligned} \mathbf{C}\overline{\psi}\Gamma\psi\mathbf{C}^{-1} &= +\overline{\psi}\Gamma\psi \qquad \text{for} \quad \Gamma = (1, \gamma^5\gamma^\mu) \\ \mathbf{C}\overline{\psi}\Gamma\psi\mathbf{C}^{-1} &= -\overline{\psi}\Gamma\psi \qquad \text{for} \quad \Gamma = (\gamma^5, \gamma^\mu, \gamma^{\mu\nu}). \end{aligned} \tag{5.111}$$

These in turn result in the correct transformation behavior that we expect for the current, namely $j_\mu^C = -j^\mu$. And furthermore, because of the field equations $\partial_\nu F^{\nu\mu} = j^\mu$, the potential A^μ changes sign.

5.5.5 *CPT Theorem*

The combined symmetry operator **CPT** is anti-linear and anti-unitary because of the presence of **T**. And because of the appearance of **PT**, the four-vector (x) becomes $(-x)$.

The CPT theorem implies that a Lagrange density which obeys the postulate of locality, i.e. which contains only a finite number of space-time derivatives of the fields, transforms as

$$(\mathbf{CPT})\mathcal{L}(x)(\mathbf{CPT})^{-1} = \mathcal{L}(-x). \tag{5.112}$$

In collecting the previous results of the different variants of fields with respect to the three distinct discrete symmetries, we guess that for bosonic fields

$$\mathbf{CPT}\phi_{\mu_1\ldots\mu_n}(x)\mathbf{T}^{-1}\mathbf{P}^{-1}\mathbf{C}^{-1} = (-1)^n\phi_{\mu_1\ldots\mu_n}(-x). \tag{5.113}$$

This is obviously true for scalar fields. For vector fields we have the sequence

$$\begin{gathered} A^\mu(\vec{x}, t) \xrightarrow{T} -\mathcal{T}^\mu_{\ \nu} A^\nu(\vec{x}, -t) \xrightarrow{P} \\ -\mathcal{P}^\mu_{\ \rho}\mathcal{T}^\rho_{\ \nu} A^\nu(-\vec{x}, -t) \xrightarrow{C} \mathcal{P}^\mu_{\ \rho}\mathcal{T}^\rho_{\ \nu} A^\nu(-\vec{x}, -t) = -A^\nu(-\vec{x}, -t) \end{gathered}$$

which is an example of (5.113) for $n = 1$. Since a tensor $T^{\mu\nu}$ transforms with two $\mathcal{P}$ and two $\mathcal{T}$ matrices the sign in the final expression is $+1 = (-1)^2$ and so forth. For a Dirac spinor, the sequence of discrete transformations is

$$\psi(\vec{x}, t) \xrightarrow{T} \gamma^1\gamma^3\psi^*(\vec{x}, -t) \xrightarrow{P} \gamma^1\gamma^3\gamma^0\psi^*(-\vec{x}, -t) \xrightarrow{C} \gamma^1\gamma^3\gamma^0(i\gamma^2)\psi(-\vec{x}, -t)$$

leading to the CPT transformation formula

$$\mathbf{CPT}\psi(x)\mathbf{T}^{-1}\mathbf{P}^{-1}\mathbf{C}^{-1} = -\gamma^5\psi(-x).$$

Observe that space inversion is mediated by γ^0, time reversal by $\gamma^1\gamma^3$, charge conjugation by γ^2, and the CPT-transformation by γ^5.

Next consider the Dirac bilinears. Let me make the calculation explicit for $\bar{\psi}\psi$:

$$\bar{\psi}^C\psi^C = (\psi^T C)(C\bar{\psi}^T) = -\psi^T\overline{\psi}^T = +(\overline{\psi}\psi)^T = (\overline{\psi}\psi)$$

where the next-to-last change in the algebraic sign is due to the spin-statistics relation by which fermions anti-commute. You are invited to show that the complete list of CPT-transformations for the Dirac bilinears can be simply expressed as

$$\mathbf{CPT}\bar{\psi}\Gamma_l\psi(x)\mathbf{T}^{-1}\mathbf{P}^{-1}\mathbf{C}^{-1} = (-1)^l\bar{\psi}\Gamma_l\psi(x) \tag{5.114}$$

where l is the number of Lorentz indices appearing in Γ_l.

The QFT Lagrange density is a Lorentz scalar constructed from the bosonic and fermionic fields and their derivatives. In case of gauge theories, the ordinary derivatives ∂_μ are replaced by covariant derivatives $D_\mu = \partial_\mu - igA_\mu$. These transform as

$$\mathbf{CPT}D_\mu(\mathbf{CPT})^{-1} = -D_\mu.$$

The bosonic fields carry an explicit Lorentz index and so do the Dirac bilinears. In a Lorentz scalar all indices are saturated. The number of Lorentz indices relates to the transformation under **CPT** according to (5.113) and (5.114) in such a way that the signs potentially arising cancel. From this it is conceivable that a Lagrangian with a finite number of terms obeys (5.112). More details of the CPT theorem can be found in e.g. Sect. 2.6. of [536] or in [235]. Weinberg in particular carries a very careful bookkeeping of phases in his QFT textbook. As a matter of fact, each of the single C-, P-, T-transformations can have a phase, whereas CPT, which is connected to the identity, cannot have an arbitrary phase.

The CPT symmetry is basic to the picture underlying Feynman diagrams, where the anti-particles of a particle $|p\rangle$ is visualized as an object moving backwards in spacetime[20] because of $C|p\rangle = PT|p\rangle$.

CPT conservation entails in particular that the masses of particles and anti-particles be the same. This gives a limit for its validity as for example

$$\left|\frac{m_{\bar{K}_0} - m_{K_0}}{m_{K_0}}\right| < 10^{-18}.$$

[20] This was proposed as early as 1941 by E. Stückelberg.

5.6 Effective Field Theories

5.6.1 EFT: The Very Idea

The issue of renormalizability of field theories saw a “change in attitude” [538] and a “new philosophy” [68] beginning in the 1970’s. This was influenced by the further development and interpretation of the renormalization group for solid state physics by K. Wilson. Originally, renormalization was a tricky procedure to eliminate divergences in QED. This trick starts with introducing an arbitrary cut-off Λ into the theory in order to render transition amplitudes finite; this being denoted as *regularization*. At a certain point of the calculations, the cutoff is sent to infinity. By this the bare mass and the bare charge of the classical theory become finite (renormalized). If you penetrate this mystery as subtracting infinity from infinity in order to get a finite result, you won’t be surprised that renormalization left people unsure. Nevertheless, the criterion of renormalizability became a must for other theories as well, and somehow it was added to the list of “Nature’s principles”. This led to the power-counting argument advocated in Subsect. 5.4.1. Power counting played a significant part in restricting the Lagrangians for the basic fields, in that only a finite number of interaction terms were allowed, and all those with negative mass dimension of the interaction couplings were dropped. On the other hand, non-renormalizable quantum field theories were considered dubious in terms of their status as “fundamental” theories. This was for example the case for the Fermi model, the first model for weak interaction processes. Another case would be quantized general relativity, which because of the non-positive mass-dimension of the gravitational constant cannot be renormalized. In these theories, there is no way to render the parameters (i.e. the Fermi constant or the gravitational and the cosmological constant) finite. However, even if certain terms in the action are classically not allowed by power-counting arguments, quantum corrections tend to introduce them, and this is one of the indications for doubting the renormalization “principle” in its original form.

These days all quantum field theories are considered to be effective field theories (EFT) in the following sense: They are an “appropriate” description of the “important” physics in a given region of the parameter space [210]. If as a parameter one takes for instance an energy scale, it is currently agreed upon that an appropriate and important description of physics in the spirit of an EFT depends on the energy at which the interactions are studied. Thus an effective field theory is understood to be appropriate for an energy below a certain cutoff Λ_M where the cut-off is no longer arbitrary but given in terms of a typical mass in the theory. The cut-off and renormalized entities are interdependent, and the physical observables are independent of the cut-off, as controlled by the renormalization group (RG) equation; more on this below.

The very idea of EFT lets one contemplate about a chain of EFT’s, starting from a lowest mass theory and successively elongating to higher and higher mass scales. In its extreme form, one can identify each particle mass with a boundary between two effective theories. If the conditions of decoupling (see below) are given, one gets valid results from the effective low-energy theory describing interactions of those particles with masses beyond the cut-off mass, together with quantifiable corrections

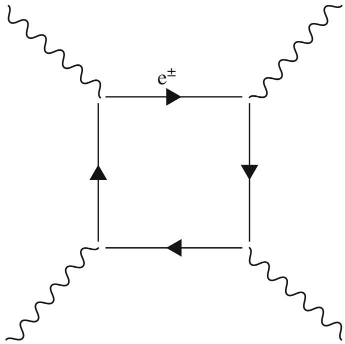


Fig. 5.2 Lowest order photon-photon scattering

from the high-energy part. There is no reason to hide our ignorance about the most appropriate theory of fundamental physics (not to mention the final theory), since that ignorance is qualified and quantified by the parameter range available to present-day experimental technology.

Although there are a lot of technical details (some of which are sketched below), the very idea of effective field theories is nearly trivial and self-evident: In atomic physics we need not care about the properties of nuclei in order to calculate spectra. In nuclear physics (at moderate energies–which tautologically is the energy range which defines the applicability of nuclear models and reactions), we need not care about the quark constituents of the nuclei. Quantum electrodynamics describes the interaction of electrons and photons below energies of the order MeV rather well.[21] Or, in the words of H. Georgi "If we had to know everything about all particles, no matter how heavy, we would get nowhere" (see Sect. 8.2 in [212]).

5.6.2 Historical Examples

The Heisenberg-Euler Theory

Before Dirac's 1931 theory of electrons and positrons interacting with an electromagnetic field became full-fledged QED at the end of the 1940's, H. Euler and W. Heisenberg investigated in 1936 low-energy photon-photon scattering. "Low" energy refers in this case to energies well below the production threshold of electrons and positrons: $E_\gamma \ll m_e$. In our modern understanding virtual electrons/positrons participate in this action, in that two photons create an electron-positron pair which afterwards decays in two photons again; visualized for instance by the Feynman diagram Fig. 5.2.

Off course the classical Maxwell action cannot reflect this process. Therefore it was augmented by Euler and Heisenberg through further terms meant to represent

[21] Although, as argued in Sect. 5.6.3, QED does not really exist on its own.

photon-photon interactions. In a "modern" notation they started from the most general photon Lagrangian respecting (1) U(1) invariance, (2) Lorentz invariance, and (3) parity invariance:

$$\mathcal{L}_{\gamma\gamma} = -\frac{1}{4}\,F^{\mu\nu}F_{\mu\nu} + \frac{a}{m_e^4}(F^{\mu\nu}F_{\mu\nu})^2 + \frac{b}{m_e^4}F^{\mu\nu}F_{\nu\rho}F^{\rho\sigma}F_{\sigma\mu} + \mathcal{O}(\frac{F^6}{m_e^8}),$$

with m_e being the electron mass, and the entities a and b dimensionless constants. Notice that here photon-photon scattering is described without any reference to electron and positron fields, although its complete description would need the full QED (and potentially even more since at energies sufficiently large, also pion/anti-pion pairs and other objects could be produced). Here you see the very idea of EFTs in that the "exact" theory contains only the first Maxwell term plus a Dirac spinor term. Although the Lagrangian $\mathcal{L}_{\gamma\gamma}$ and the QED Lagrangian differ in structure, the former yields approximate results, and these can be made as good as you are willing to make the expansion in $(1/m_e)$. Notice that the Euler-Heisenberg Lagrangian is completely defined in terms of the electromagnetic field, containing all terms allowed by the symmetries, and not at all restricted by dimensional renormalisation criteria.

In order to relate the effective theory $\mathcal{L}_{\gamma\gamma}$ with the (possibly effective) QED, one needs to determine the constants a and b; generically this is called the matching of two (effective) theories. Today, with the power of functional integrals one may integrate out the fermionic degrees of freedom from the full generating functional. This was of course not known at the time of Euler and Heisenberg. By some ingenieous techniques they were able to show that $a = -\frac{\alpha^2}{36}$ and $b = \frac{7\alpha^2}{90}$ with the fine structure constant $\alpha = 1/137$. Written in the original form, the Heisenberg-Euler Lagrangian becomes

$$\mathcal{L}_{EH} = \frac{1}{2}(\vec{E}^2 - \vec{B}^2) + \frac{2\alpha^2}{45m_e^4}\left[(\vec{E}^2 - \vec{B}^2)^2 + 7(\vec{E}\cdot\vec{B})^2\right],$$

an effective Lagrangian experimentally verified and still used today (for instance in quantum optical applications).

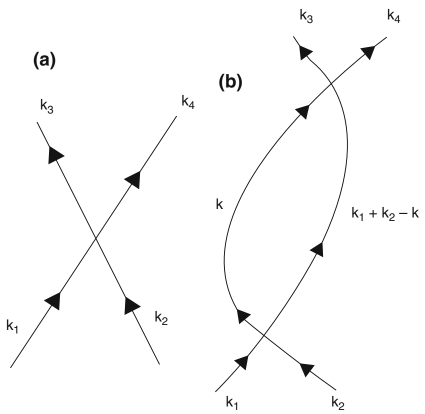
The Fermi Theory

As will be explained in Sect. 6.3. weak decay processes are today successfully described by the Glashow-Salam-Weinberg (GSW) model up to energies in the TeV-range. Back in the 1930's, only one of these processes had been observed, namely β-decay. This was thought of as a process in which a neutron in a atomic nucleus decays into a proton, an electron and a neutrino[22] $(n \to pe\nu)$. Fermi used as an *ansatz* for the interaction Hamiltonian:

$$\mathcal{H}_\beta = \frac{G_F}{2}(p\gamma^\lambda n)(e\gamma_\lambda \nu).$$

[22] At that time, one did not know that there are three varieties of neutrinos and their antiparticles.

Fig. 5.3 $\varphi\varphi$-Scattering processes: tree diagram and one-loop diagram



This was later extended–after more and more weak processes, and especially parity violation, were discovered–to the so-called (V-A)-theory. Its agreement with experimental results was quite good for energies up to several GeV. Today these "old" theories of weak interactions may be understood as effective field theories which can be derived from the GSW action. Notice that the coupling constant G_F has mass dimension (-2), and thus the model is not renormalizable.

Again we see that the effective theory takes only the "relevant" fields into account, in this case the proton, neutron, electron and neutrino fields. On the energy scale on which the theory delivers nearly exact results, the "irrelevant" W- and Z-bosons can be neglected. And the theory delivers astonishingly good results in this range, despite it is non-renormalizable and should–according to the old philosophy–be completely banned.

5.6.3 Renormalization (Group)

Regularization and Renormalization

This is not the place to explain the regularization and renormalization techniques in detail. Instead I will follow a story line by A. Zee (see Sect. III.1 in [579], a chapter with the felicitous title "Cutting Off Our Ignorance"), using the scalar theory

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - \frac{1}{2}\, m^2\varphi^2 - \frac{1}{4!}g\varphi^4.$$

According to the (Feynman) rules of the game, the loop diagram Fig. 5.3b contributes an amplitude

$$M_L = \frac{1}{2}(-ig)^2 \int \frac{d^4k}{(2\pi)^4}\, \frac{i}{k^2 - m^2 + i\epsilon}\, \frac{i}{(k_1 + k_2 - k)^2 - m^2 + i\epsilon}.$$

This integral diverges logarithmically as $\int d^4k/k^4$ if the integration is taken between the limits $0 < k < \infty$. If instead the integration is taken up to a finite value Λ, the result is finite:

$$M_L(\Lambda) = ig^2\, C\left[\ln\left(\frac{\Lambda^2}{s}\right) + \ln\left(\frac{\Lambda^2}{t}\right) + \ln\left(\frac{\Lambda^2}{u}\right)\right]$$

where (s, t, u) are the Mandelstam variables $s = (k_1 + k_2)^2$, $t = (k_1 - k_3)^2$, $u = (k_1 - k_4)^2$ and $C = 1/(32\pi^2)$. The prescription of imposing the cutoff to make a divergent expression finite is known as Pauli-Villars regularization.[23] Whereas in the early history of renormalization this was understood a mere mathematical technicality, today in the context of effective theories the cutoff describes the range of the validity of the theory. For energy-momenta bigger than Λ we enter a range of ignorance.

The expression for M_L contains two undetermined parameter, namely the coupling g and the cutoff Λ. The latter is introduced by hand, the former is something that should be measured in experiments. Now the (renormalization) trick is to understand $g = g(\Lambda)$ and to get rid of the cutoff dependence of the $\varphi\varphi$ scattering amplitude by an appropriate definition of the physical (that is measurable) coupling. In other words: If the cutoff is changed, the coupling is changed in such a way that the amplitude does not change. And in more detail: The amplitude for $\varphi\varphi$-scattering is given by

$$M = M_T + M_L(\Lambda) = -ig + ig^2 C\left[\ln\left(\frac{\Lambda^2}{s}\right) + \ln\left(\frac{\Lambda^2}{t}\right) + \ln\left(\frac{\Lambda^2}{u}\right)\right] + \mathcal{O}(g^3)$$

where the first term is the tree contribution from Fig. 5.3a. If g_0 is measured in a $\varphi\varphi$-scattering experiment with (s_0, t_0, u_0), it holds that

$$-ig_0 = -ig + ig^2 C\left[\ln\left(\frac{\Lambda^2}{s_0}\right) + \ln\left(\frac{\Lambda^2}{t_0}\right) + \ln\left(\frac{\Lambda^2}{u_0}\right)\right] + \mathcal{O}(g^3).$$

This can be solved for g and re-introduced into the amplitude. A little arithmetic yields

$$M = -ig_0 + ig_0^2 C\left[\ln\left(\frac{s_0}{s}\right) + \ln\left(\frac{t_0}{t}\right) + \ln\left(\frac{u_0}{u}\right)\right] + \mathcal{O}(g_0^3). \tag{5.115}$$

And this is the renormalization (miracle): Starting out with an expression for the amplitude with two unphysical parameters we end up with an amplitude containing a measured coupling constant.

This interplay of "bare" constants (here g) with the regularization cutoff resulting in renormalized/physical constants is not a peculiarity of the previous toy theory, but also happens for QED and more general quantum field theories.

---

[23] There are other regularization techniques as well. The one mostly used today is dimensional regularization, where one first treats the theory in $(d - \epsilon)$ dimensions and then takes the limit $\epsilon \to 0$.

Renormalizability vs. Non-Renormalizability

Consider QED as derived from the classical Lagrangian

$$\mathcal{L}_{QED} = \bar{\psi} i \not\partial \psi - e\bar{\psi} \not{A} \psi - m\bar{\psi}\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$

together with some regularization procedure with a cutoff Λ. Removing from this theory all those states for which the energy-momenta Q are larger than some $\Lambda_0 < \Lambda$ can be compensated by adding to the Lagrangian a term

$$\Delta\mathcal{L}_{QED} = -eC(\Lambda_0/\Lambda)\, \bar{\psi} \not{A} \psi - m\tilde{C}(\Lambda_0/\Lambda)\, \bar{\psi}\psi.$$

Here C and $\tilde{C}$ are dimensionless functions depending logarithmically on Λ_0/Λ. Notice that each of the terms has the same structure as terms in the Lagrangian $\mathcal{L}_{QED}$. In effect, the theory defined by the Lagrangian

$$\mathcal{L}_{QED} + \Delta L_{QED} = \bar{\psi} i \partial \psi - e_0 \bar{\psi} \not{A} \psi - m_0 \bar{\psi}\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + c\frac{m_0}{\Lambda_0^2} \bar{\psi} \gamma^{\mu\nu} F_{\mu\nu} \psi$$

leads to the same field equations as the original Lagrangian with cutoff Λ up to corrections $1/\Lambda_0^2$ if

$$e_0 = e\,[1 + C(\Lambda_0/\Lambda)] \qquad\qquad m_0 = m\left[1 + \tilde{C}(\Lambda_0/\Lambda)\right].$$

Here we meet again the circumstance that the cutoff can be compensated by changing the parameters of the theory. And we see another fact, namely that the procedure generates a term in the Lagrangian which is suppressed by a factor $1/\Lambda_0^2$. This "Pauli term" respects all symmetries of QED, but it is non-renormalizable.

The distinctive difference of renormalizable and non-renormalizable theories lies in the fact that for the former all counterterms can be removed by redefining the parameters in the theory, whereas for the latter the counterterms stay in the theory. These counterterms have the same symmetries as the original Lagrangian, and they can be organized as power series in $1/M^{|k|}$ where, as argued before, M is a cutoff mass in the theory.

Renormalization Group Flow

From the expression (5.115) for the $\varphi\varphi$-scattering amplitude, one understands that g_0 is a function of $s_0, t_0,$ and u_0. To make things easier (and since the results in any case only hold to order g_0^2), one may assume that $s_0, t_0,$ and u_0 are of the same order μ^2, and therefore

$$M = -ig(\mu) + iCg^2(\mu)\left[\ln\left(\frac{\mu^2}{s}\right) + \ln\left(\frac{\mu^2}{t}\right) + \ln\left(\frac{\mu^2}{u}\right)\right] + \mathcal{O}(g^3(\mu)).$$

Now the amplitude stays the same if one replaces μ by μ' and the relation between $g(\mu)$ and $g(\mu')$ for $\mu \sim \mu'$ is changed to

$$g(\mu') = g(\mu) + 3\,Cg^2(\mu)\ln(\frac{\mu'^2}{\mu^2})$$

or, expressed as a differential "flow" equation

$$\mu \frac{d}{d\mu} g(\mu) = 6Cg^2(\mu) + \mathcal{O}(g^3(\mu)).$$

This is the *renormalization group equation* for the $g\varphi^4$ model.[24] Generically for a theory with various coupling "constants" g_i, the flow equations are usually written in terms of the Callen-Symanzik β function as

$$\mu \frac{d}{d\mu} g^i(\mu) = \beta^i(g(\mu)) \qquad \text{or} \qquad \frac{dg^i}{dt} = \beta^i(g) \tag{5.116}$$

where the latter comes about by defining $t = \ln(\mu/\mu_0)$. Notice that the β functions depend on μ only through the $g(\mu)$. Any physical observable Σ depends on the g_i and on μ in such a way that

$$\frac{d}{d\mu}\Sigma(g^i(\mu), \mu) = 0 \qquad \text{or} \qquad \left(\mu \frac{\partial}{\partial \mu} + \beta^i \frac{\partial}{\partial g^i}\right)\Sigma = 0.$$

The differential equation (5.116) can formally be integrated to yield

$$\ln \frac{\mu}{\mu_0} = \int_{g(\mu_0)}^{g(\mu)} \frac{dg}{\beta(g)}. \tag{5.117}$$

Take for instance the scalar theory for which the (lowest-order) β function is $\beta_\varphi(g) = 6Cg^2$, so that

$$g(\mu) = \frac{g_0}{1 - 6Cg_0\ln(\mu/\mu_0)}; \qquad g_0 := g(\mu_0).$$

Trivialiy, Asymptotic Freedom, and Asymptotic Safety

From the latter expression we see that the coupling for the scalar theory diverges for a finite energy μ_∞:

$$\mu_\infty = \mu_0 \exp\left(\frac{1}{6Cg_0}\right),$$

---

[24] The group acting here is simply the additive group of transformations $\mu \to \mu + \delta\mu$. In a narrow sense, it is a semi-group since the equation can only be integrated if one moves from small to large scales, but not the other way round.

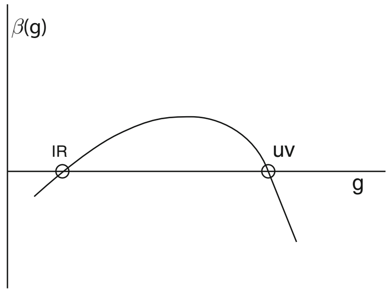


Fig. 5.4 Infrared and ultraviolet fixed points

called the Landau singularity or Landau ghost, named after L. Landau, who observed this in the case of QED for which

$$\beta_{QED}(e) = \frac{1}{12\pi^2}\, e^3.$$

Since, however, both β_φ and β_{QED} assume the given functional dependence (being proportional to g^2 and e^3, respectively) in the one-loop expansion, one might hope that higher-loop corrections remedy this unphysical behavior. As it turns out, this is not the case. Thus a Universe with only one type of scalar fields or pure QED cannot exist.

In any case, this gives rise to the question of how a generic quantum field theory can qualitatively behave for large energies. Of course, if it were possible to calculate the complete beta function one could determine this behavior from evaluating (5.117). Since however this is wishful thinking, one needs to argue qualitatively: S. Weinberg identified a general class of theories in which unphysical singularities are almost certainly absent, and this relates to the existence of "ultraviolet fixed points" g_* of the Callan-Symanzik equation. These are defined as being approached by couplings $g(\mu)$ as $\mu \to \infty$. In order for this to become true, necessarily $\beta(g_*) = 0$, and also the couplings must lie on a trajectory $g(\mu)$ which passes through the fixed point. (Things are a little bit more complicated than presented here; for a thorough treatment see [533] and Chap. 18 in [536].)

Consider the two possible behaviors of the β function near $\beta(g_*) = 0$ according to Fig. 5.4. Expand β around g_* as

$$\beta(g) = (g - g_*)\beta'(g_*) + \ldots$$

In the point labeled as UV the slope of β is negative at the fixed point. If g is smaller than g_* the negative sign cancels against the negative sign of β' so that $\partial g/\partial\mu$ is positive. Therefore, g increases with increasing μ, in other words: g is driven towards g_* for increasing energy. The same is true for $g > g_*$, since then by $\beta' < 0$ the derivative $\partial g/\partial\mu$ is negative. Thus for increasing μ the coupling g is driven downwards, back to g_*. A different behavior is observed in the case that the derivative of the β function is positive at the fixed point, in the figure sketched as

IR: Now the coupling is driven towards g_* for decreasing values of g. In the former case one is talking about an ultraviolet and in the latter case about an infrared stable fixed point. In both cases the solution of the Callan-Symanzik equation near the fixed point becomes

$$(g(\mu) - g_*) \propto \mu^{-\beta'} .$$

In this language, we interpret the fixed point $g_* = 0$ in the scalar theory and in QED as infrared stable. Conceivably, further terms in the β function with higher powers in g might lead to an ultraviolet stable fixed point and thus to an avoidance of the Landau singularity. But this seems not to be the case, and thus one must concede that both these theories are unphysical beyond a certain energy and even not mathematically consistent (see Sect. 18.3 of [536]). Sometimes these theories are dubbed "trivial" because in essence they are only consistent for zero couplings. For QED perturbation theory breaks down at very high energies. This is not harmful in itself, since QED becomes "part" of the electroweak theory in the range of the heavy vector mesons. On the other hand, and this is of much greater interest, theories are known for which the fixed point at $g = 0$ is ultraviolet stable. These are certain non-Abelian gauge theories with not too many fermions and scalars; more about this in Subsect. 6.4.4. These theories are called *asymptotic free*.

In order to understand the landscape of fixed points, the previous consideration of theories with only one coupling needs to be extended to the case with various couplings: Assume that the RG equations (5.116) have fixed point solutions $\beta^k(g_*) = 0$. The surface formed by those trajectories $g^i(\mu)$ which pass through the fixed points is called the ultraviolet critical surface. Around the g_*^i one has

$$\mu \frac{d}{d\mu}\left[g^i(\mu) - g_*^i\right] = \sum_k M_k^i \left[g^k(\mu) - g_*^k\right] \qquad M_k^i = \left[\frac{\partial \beta^i}{\partial g^k}\right]_{g=g_*} .$$

and, as shown by Weinberg, the behavior for large energies now depends on the signs of the eigenvalues of this matrix: The dimension of the ultraviolet critical surface equals the number of negative eigenvalues, and if the number is finite and (different from zero) the theory is called *asymptotically safe*. Under this point of view, a renormalizable, asymptotically free field theory is a special case of an asymptotically safe theory. In this case the fixed point is $g_*^i = 0$ and the UV critical surface is spanned, near the fixed point, by the couplings that are renormalizable in the perturbative sense. Recently there have been various hints that general relativity is asymptotically safe.

5.6.4 Chain of Effective Theories

The Structure of the Chain

Consider a transition matrix element from an initial to a finite state. Assume that these do have energies lower than Λ. Then the transition matrix can be written in terms of an effective Hamiltonian as

$$\langle F|\mathcal{H}_{eff}|I\rangle = \sum_k C_k(\Lambda)\,\langle F|\mathcal{O}_k|I\rangle|_\Lambda .$$

Here the $C_k(\Lambda)$ contain the "high"-energy contributions (energies higher than the scale Λ) and the matrix elements $\langle F|\mathcal{O}_k|I\rangle|_\Lambda$ contain the "low"-energy contributions. In a next step, one can organize the right-hand side by inverse powers of Λ. Assuming that high-energy and low-energy contributions decouple, the $C_k(\Lambda)$ do not depend on any low-energy scale, their mass dimension can come only from appropriate mass dimension of Λ. Therefore the transition matrix element becomes

$$\langle F|\mathcal{H}_{eff}|I\rangle = \sum_k \frac{1}{\Lambda^k}\sum_i c_{ki}\,\langle F|\mathcal{O}_{ki}|I\rangle|_\Lambda . \tag{5.118}$$

In this expression, the c_{ki} are dimensionless constants, and the i in $\mathcal{O}_{ki}$ enumerates all Hamiltonian terms with mass dimension $(k+4)$. The sum in (5.118) now being ordered according to the dimension of the operators, $\mathcal{O}_{ki}$ also contains contributions with mass dimension four (that is $k=0$) which define a renormalizable theory. Higher terms are supressed by factors $1/\Lambda^k$.

Instead of separating the low- and high-energy regime by a degree of freedom with a heavy mass, one can introduce a scale parameter $\mu \leqslant \Lambda$ and write instead of (5.118)

$$\langle F|\mathcal{H}_{eff}|I\rangle = \sum_k \frac{1}{\Lambda^k}\sum_i c_{ki}(\Lambda/\mu)\,\langle F|\mathcal{O}_{ki}|I\rangle|_\mu .$$

With the understanding that the coefficients c_{ki} contain the "physics" above the scale μ and the matrix elements $\langle F|\mathcal{O}_{ki}|I\rangle|_\mu$ the "physics" below μ we see that changing μ shuffles contributions from the coefficients to the matrix elements and *vice versa*. And because of the renormalizability of the dimension-four part the "Wilson" coefficients c_{ki} can depend on μ only via the logarithm of the ratio Λ/μ. Notice that the heavy scale/mass makes itself felt in this logarithmic dependence in the coupling parameters of the effective low-energy theory and in terms suppressed by inverse powers of the heavy mass.

Next we can make the connection of the previous power expansion with the renormalization group: Since μ is arbitrary it must hold that $\mu\frac{d}{d\mu}\langle F|\mathcal{H}_{eff}|I\rangle = 0$, or explicitly, for each index k

$$\sum_i \left(\mu\frac{d}{d\mu}c_{ki}(\Lambda/\mu)\right)\langle F|\mathcal{O}_{ki}|I\rangle|_\mu + \sum_i c_{ki}(\Lambda/\mu)\left(\mu\frac{d}{d\mu}\langle F|\mathcal{O}_{ki}|I\rangle|_\mu\right) = 0.$$

It can be shown that under a change in μ the operators $\mathcal{O}_{ki}$ become a linear combination of operators with the same mass dimension:

$$\mu\frac{d}{d\mu}\langle F|\mathcal{O}_{ki}|I\rangle\Big|_\mu = \sum_j \gamma_{ij}(\mu)\langle F|\mathcal{O}_{kj}|I\rangle\Big|_\mu .$$

The matrix γ is called the anomalous-dimension matrix. Inserting this into the previous expression and observing that for each k the $\mathcal{O}_{ki}$ form a complete basis, we finally arrive at

$$\sum_j \left[\delta_{ij}\mu\frac{d}{d\mu} + \gamma_{ij}(\mu)c_{ki}\right](\Lambda/\mu).$$

What makes renormalizable theories (prime examples being QED and the SM) special in this scheme? If one adds to a renormalizable theory all terms allowed by symmetry (which are infinite in number) there are necessarily interaction terms with coupling constants having mass dimension $[g_I] = -|\Delta_I|$ with Δ_I defined in (5.23). It is not unreasonable to assume that the couplings are of the order

$$g_I \sim M^{-|\Delta_I|}$$

where M is some typical (particle) mass. If one calculates processes for which $k \ll M$, the coupling g_I must on dimensional grounds be accompanied by a factor $k^{|\Delta_I|}$. As a consequence the non-renormalizable interaction part is suppressed by a factor $(k/M)^{|\Delta_I|} \ll 1$. In other words: When the mass in question is large the effective theory is approximately renormalizable.

Moving Up and Down the Chain

There are two procedures to make efficient and effective use of effective theories: Assume you have a theory which is experimentally verified up to a scale Λ_0. You may then either derive effective theories valid for $\Lambda \ll \Lambda_0$ by integrating out heavy modes or you may guess a theory which is valid beyond Λ_0 but of course delivers the same results as the older theory below Λ_0.

- Moving down the chain
  There are examples where "moving down" can be made explicit. In the electroweak part of the SM, moving down from the scale of the W- and Z-boson mass of about 100 GeV to the β-decay scale of 1 GeV we will find the Fermi theory, and lowering the scale we may go beyond the electron mass to disclose the Euler-Heisenberg model. But why, at all, having an underlying "fundamental" theory (which even is renormalizable) should we derive lower-lying effective theories. As will be discussed for the SM, this is because the calculations within the full theory are to such an extent technically complicated that one preferably resorts to a well-understood lower-lying effective theory.
  If the underlying theory is known, the effective action can systematically–but formally–found from the functional integral by integrating out the heavy fields: Assume that the full action is $S[\varphi, \phi]$. Define

$$e^{iS_{eff}[\varphi]} = \int \mathcal{D}\phi \, e^{iS[\varphi,\phi]}.$$

The effective Lagrangian,[25] implicitly defined by $S_{eff} = \int d^4x \, \mathcal{L}_{eff}[\varphi]$, is in general not local, but a functional of the fields.

- Moving up the chain
  Whereas moving down the chain is more ore less straightforward by observing the mass scales in the theory one starts with, moving up involves guessing. And we know from the historical examples that the only guidepost in guessing are symmetry consideration: The Euler-Heisenberg Lagrangian was derived by adding to the Maxwell-Lagrangian ($F_{\mu\nu}F^{\mu\nu}$) being quadratic in the field strength further term quartic in the field strength, etc. Although these further terms spoil the dimensional renormalization criteria, it seems as if the effective action gives the same results as QED (of course for energies well below the electron threshold). As for the other historic example, the Fermi theory is the most general effective theory that exhibits local $\mathbf{SU(3)} \times \mathbf{SU(2)} \times \mathbf{U(1)}$-symmetry[26].
  In the next section the standard model will be explained as a theory which survived all experiments over the last forty years with astounding precision. Nevertheless, these days the opinion prevails that this model, although renormalizable, is merely an effective theory, valid at least up to the order of TeV, but which will be superseded by another more fundamental theory valid at still higher energies. Some words about what this theory might look like will be spent in Subsect. 6.4.5.
  The other "fundamental" theory, namely general relativity as the theory of gravitation, might as well be the lowest term of an effective theory, and thus it is no surprise that theorists rack their brains as to how the other terms might look like. Again, symmetry arguments, in this case the quest for diffeomorphism invariance, are a guidepost (see Sect. 7.6.).

5.7 Concluding Remarks and Bibliographical Notes

This chapter essentially dealt with the logical consequences of special relativity as applied to field theory, and on the consequences of marrying special relativity with quantum physics. And these consequences are astoundingly restrictive. To begin with, irreducible unitary representations of the Poincaré group (respectively its covering group) lead to definite parametrizations of particle properties in the Wigner program. This program was first laid down more than seventy years ago in [555]. For a detailed exposition see [398]; and for a modernized treatment [487]. At the same time, we saw that the representations of the Poincaré group determine all linear equations which are invariant under this group. The most important wave equations for scalar, spinor, vector fields could be derived in the Bargmann-Wigner program [29]. O. Klein and W. Gordon (and the many others who discovered the KG equation), P. A. M. Dirac and J. C. Maxwell certainly would have been surprised to see their "single-particle"

---

[25] This effective action is not the same as the quantum action defined in Appendix D.2.2., which in the literature very often is also called effective action.

[26] And this without gauge bosons but at the cost of a coupling constant with mass dimension -2.

wave equations in this perspective.[27] None of them derived "their" equations from symmetry arguments.

We also followed another path towards the field equations by constructing appropriate actions. This construction started with building Lorentz scalars from the fields and their derivatives. And it turned out that the free-field part for spin-0 and spin-$^1/_2$ field are fixed from Poincaré symmetry and arguments of dimensional renormalizability. The spin-1 field Lagrangian is unique if additionally gauge invariance is imposed. Furthermore, the interaction terms are unique if one requires dimensional renormalizability. Thus, the full theory of interacting scalars, spinors, and gauge vector fields is fixed by the requirement of Poincaré and gauge invariance (with respect to symmetry group to be determined by experiment and perhaps within another theoretical frame) together with the constraint of the renormalization criterion. As a very good description of the grand arc from quantum mechanics via relativistic field theory towards particle physics I recommend [427] (although it may be hard to find in libraries today).

A central theme in fundamental physics (and not only for particle physics but also for general relativity) is the idea of a gauge theory. The term "gauge invariance" (in German: "Eichinvarianz") was introduced into theoretical physics by Weyl in his 1919 endeavour to generalize the coordinate invariance of general relativity; and it was related to scale invariance. Although Weyl's attempt failed, the "gauge" terminology survived, albeit with a different meaning. The basic idea[28] is to promote a global invariance to a local invariance by introducing connections ("gauge fields"). Instead of talking about $\mathbf{U}(\mathbf{1})$ gauge invariance of electrodynamics, a more appropriate terminology would be $\mathbf{U}(\mathbf{1})$ phase invariance.[29] The non-Abelian counterpart, today known as Yang-Mills theories, and invented more than fifty years ago had a varied fate in particle physics, and we will see some of the reasons in the next chapter. However, they were fully established after Gerard 't Hooft proved in 1971 that among all spin-one fields, only Yang-Mills theories are renormalizable, and after it was shown that only a non-Abelian theory can be asymptotically free. A well-done collection of articles reflecting the basic structure of Yang-Mills theories and their properties is [502].

For a modern understanding of QFT–aside from the by now already classic [536]–see the reviews [135, 193, 559]. Quantum field theory as for instance used in comparing experimental results with the SM predictions is based on perturbative methods derived within the Feynman path-integral approach to quantum physics. These are therefore treated in Appendix D, as they most clearly exhibit which consequences the

[27] Even Dirac, mentioned as the pioneer in higher-spin relativistic field equation, tried to derive these with algebraic acrobatics; there are no notions of symmetry.

[28] Interestingly, but rounding our understanding of symmetries, a gauge theory can also be obtained by enforcing constants of motion–originating from global symmetries in a theory–to constraints with Lagrangian multipliers [303].

[29] C. N. Yang: "If we were to rename them today, it is obvious that we should call the gauge invariance phase invariance, and the gauge fields should be called phase fields." in "Geometry in physics", in "To fulfill a vision–Jerusalem Einstein Centennial Symposium on Gauge Theories and Unification of Physical Forces," (ed.) Y. Ne'eman, Addison-Wesley, New York, 1981.

Noether theorems do have for QFT. In the same appendix, I also sketch the BRST symmetries, originally discovered in the Faddeev-Popov path-integral method for quantizing gauge field theories. The other essential notion of QFT's, namely renormalization, was only outlined in this chapter. I highly recommend the article [328] in which by the example of a model quantum field theory Feynman diagrams, perturbation theory, dimensional regularization, and the renormalization processes are worked out. And for the far-reaching re-interpretation of renormalizable and non-renormalizable field theories as effective theories see [539].

Introductory text books on relativistic field theory are [307] who also gives glimpses at concepts still speculative (grand unified theories, Kaluza-Klein theories, supersymmetry); [431] which offers a balanced approach of formality and applicability (in particle physics); [456] from which you can get the basic ideas without extraordinary mathematical rigor; [293] which provides a technically excellent account of Feynman path-integral quantization in field theories; and [412] in which you find both the essentials of quantum field theory and the standard model. In general, there is an overlap of books stressing more the field-theortical aspects ot the particle physics aspects; thus it is a matter of taste. A remarkable introduction into the field is [579], in that it also covers topics from condensed matter physics, given that the most recent advances (like for instance spontaneous symmetry breaking, renormalization group theory) came both from particle and solid state physics.

Wigner's program reveals that–given Poincaré symmetry in $4D$ and postulates from quantum physics–nature can accommodate particles with integer and half-integer spins. In the next sections we will see that nature does so to a certain degree: The standard model hosts spin-1 fields (vector-bosons) as carriers of electromagnetic, weak and strong interactions. Three generations of quarks and leptons are represented as spin-$^1/_2$ fields.There are also spin-0 fields, namely the Higgs bosons. In Sect. 7.6. we will meet the graviton, which according to the linearized version of general relativity is a spin-2 object. And according to supergravity (see Sect. 8.3) nature might also hold in store a spin-$^3/_2$ object, namely the gravitino. However it seems that nature dislikes fields with spin higher than two. But maybe this is part of our ignorance: only free-field actions are known both for massless and massive fields.

Additional Points

There are many more fascinating results in merging special relativity with quantum physics, or–to stay in the context of this book–combining spacetime symmetry with the unitary representations. Limited space allows me only to sketch these:

- The program of Bargmann and Wigner to find the unitary representations of the Poincaré group and the relativistic field equations associated with the classifying parameters (mass, spin, ...) was applied to dimensions other than four and to other groups as well.
  An extensive and detailed account dealing with the Poincaré group (or rather with $\mathbf{ISO}(\mathbf{D}-\mathbf{1},\mathbf{1})^{\uparrow}$) for dimensions $D \geqslant 3$ is [37]. Generically the universal covering

group of $\mathbf{ISO(D-1,1)}^{\uparrow}$ is $\mathbf{Spin(D-1,1)}$,[30] and the mapping of orbits to the little groups is essentially like that given in Table 5.1, simply replacing $\mathbf{SO(3)}$ by $\mathbf{SO(D-1)}$, $\mathbf{ISO(2)}$ by $\mathbf{ISO(D-2)}$, and so forth.

Quite a few authors dealt with the unitary irreducible representations of the de Sitter group and the wave equations in de Sitter spacetimes [249]

As for the conformal group it is known since long [544] that dilatation cannot be a symmetry of nature because it would imply that all masses vanish, or that the mass spectrum is continuous. This comes about since $\left[S, P^2\right] = P^2$ and thus P^2 is not a Casimir operator.

- The Wigner classification is not without twist. It does not work for particles carrying an electric charge. The reason being that charged particles carry along clouds of low energy photons and therefore cannot be described by eigenvectors of the mass operator or by vectors in some Lorentz invariant superselection sector of the physical Hilbert space. Since they are "almost" discrete eigenstates of the mass operator they are called *infraparticles*.

  On the other hand, the classification gives rise to objects with some unusual (and not observed) properties like the infinite spin objects or the tachyons. Excluding them from the classification just by postulating that they do not exist seems unsatisfactory. Thus, it comes of no surprise that now and then these "black-sheep-particles" are reconsidered.

  In the literature you find also speculations about particles not directly related to the Wigner classification: mirror particles which should serve to re-establish parity invariance. The idea is to associate to each of the observed left-handed fermions a right-handed partner which might be heavier due to spontaneous symmetry breaking.
- In [531], S. Weinberg derived electrodynamics and general relativity from the Lorentz invariance of the S-matrix and the property of the creation and annihilation operators to transform according to irreducible representations of the Poincaré group: "Maxwell's theory and Einstein's theory are essentially the unique Lorentz-invariant theories of massless particles with spin $j = 1$ and $j = 2$."
- The Weinberg-Witten theorem(s) [541] forbids the existence of massless particles in a very wide class of QFT's. The theorem states that (1) A theory containing a Poincaré covariant-conserved current J^{μ} forbids massless particles of spin $j > 1/2$ with a non-vanishing eigenvalue of the conserved charge $\int J^0 d^3x$. (2)A theory containing a Poincaré covariant-conserved tensor $T^{\mu\nu}$ forbids massless particles of spin $j > 1$ for which $\int T^{0\nu} d^3x$ is the conserved energy-momentum four-vector. For a careful discussion of the proof of the theorem and some of its consequences, see [345].
- I only briefly mentioned the (existentially) important connection between spin and statistics, or in a narrower sense the fact that integer spin fields represent bosons, whereas half-integer spin field represent fermions. This was visible in the context of replacing Fourier coefficients in the classical field solutions by creation and annihilation operators. In order to have a positive definite energy, consistency

[30] $\mathbf{Spin(3,1)}$ is isomorphic to $\mathbf{SL(2,\mathbb{C})}$.

required that spinor fields should satisfy anti-commutation relations. The original form by W. Pauli [407] makes use only of the invariance of a field theory with respect to $\mathbf{SL}(\mathbf{2}, \mathbb{C})$. This may have led Pauli to end his paper by "In conclusion we wish to state, that according to our opinion the connection between spin and statistics is one of the most important applications of the special relativity theory". From Pauli's proof results that integral-spin fields cannot satisfy anti-commutation relations, and half-integral-spin fields cannot satisfy the commutation relations. This does not exclude the possibility that fields exist which satisfy other commutation relations and show statistics that differ from Bose and Fermi statistics. Therefore in the literature you now and then find reflections on notions beyond the bosonic and fermionic classification of fields. These are generically named as "parastatistics" [239]. A special case are are so-called anyons which should exist in two-dimensional systems.

- I did not mention at all non-perturbative aspects of gauge theories, which you find under the names of solitons, instantons, monopoles, …Some of these are related to topology, as for example instantons are linked to the θ-term in the gauge field Lagrangian; more about this is found in [453]. Monopoles are related to a symmetry called duality. This is neither a variational symmetry nor a discrete symmetry. The simplest example is the exchange of the E- and the B-field in Maxwell's equations. Obviously the free equations are invariant under this exchange. They would also stay invariant if there exists a current from a magnetic source and–as pointed out by Dirac–this would lead to quantization of electric charges:

$$g \cdot e = 2\pi\, n.$$

This relation also shows that in regions in which the coupling constant e is small, the coupling g becomes large, and *vice versa*. Monopoles are solutions in non-Abelian Yang-Mills theories as was shown by A. M. Polyakov and by G. 't Hooft. And duality has become nearly an engineering discipline in string theory. In the same tenor as the innocent-looking Dirac charge quantization, it turned out that the weak coupling limit of one string version is dual change to the strong coupling of another one.

- The exchange of fields with spin 0 and spin 2 leads to attraction, and that of fields with spin 1 to repulsion. This can be derived from the generating function $W(J)$ in each of these theories which typically is of the form

$$W(J) = -\frac{1}{2}\int \frac{d^4k}{(2\pi)^4} J^*_{..}(k) \frac{\hat{D}_{....}(k)}{k^2 - m^2 + i\epsilon} J_{..}(k).$$

The dots represent Lorentz indices, and one finds for scalars, vectors and gravitons

$$D = 1 \qquad D_{\mu\nu} = -G_{\mu\nu} = -g_{\mu\nu} + k_\mu k_\nu/m^2$$
$$D_{\mu\nu\rho\sigma} = (G_{\mu\rho}G_{\nu\sigma} + G_{\mu\sigma}G_{\nu\rho}) - \frac{2}{3}G_{\mu\nu}G_{\rho\sigma};$$

see e.g. Sect. I.5 in [579] for a readable and understandable derivation. Consider the scalar case first: Introduce a current $J(x) = J_a(x) + J_b(x)$ where $J_{a/b}(x) = \delta^{(3)}(\vec{x} - \vec{x}_{a/b})$ represent two "infinitely sharp spikes" as sources. Inserting these into the generating function yields

$$W_0(J) = (\int dx^0) \int \frac{d^3k}{(2\pi)^3} \frac{e^{i\vec{k}\cdot(\vec{x}_a - \vec{x}_b)}}{k^2 + m^2}.$$

The second term can be identified as the interaction energy of the two narrow lumps, and its evaluation yields

$$E_0(r) = -\frac{1}{4\pi r} e^{-mr}.$$

This demonstrates that the two lumps attract. In the case of vector fields, the current has one extra Lorentz index. Current conservation amounts in Fourier space to $k_\mu J^\mu = 0$. Therefore the second term in $G_{\mu\nu}$ gives no contribution. The only thing that changes as compared to the scalar part is $1 \to -g_{00} = -1$. Thus $E_1(r) = -E_0(r)$, and now we see why two like charges repel. In the case of photons (with $m = 0$) this is Coulomb's law. But the expression also shows that the interaction between the vector fields becomes short-range with increasing mass. This was indeed in the 1930's the reason for H. Yukawa to reflect on a scalar field mediating the short range nuclear force. In the spin-2 case, by a similar line of reasoning, current conservation $k_\mu J^{\mu\nu} = 0$ amounts to replacing the $G_{\mu\nu}$ by $g_{\mu\nu}$. Considering the (00)-component we easily find that the 1 for the scalar case is to be replaced by $g_{00}^2(2 - (\frac{2}{3}))$ which is positive (it better be, because we would not like to be thrown off the earth)!

- The photon is massless. A mass term in the Maxwell Lagrangian would spoil gauge invariance. Nevertheless, massive electrodynamics (that is, the Proca theory) can be made gauge invariant by introducing a Stückelberg field. (Since the original articles by the Swiss physicist E. C. G. Stückelberg from 1938 are written in German–and a bit hard to read anyhow–I refer to the excellent review [454] in which also modern use of "Stueckelization" is described.) Start from the Proca Lagrangian (5.55)

$$\mathcal{L}_P = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} M^2 \, A^\mu A_\mu. \tag{5.119}$$

Stückelberg's trick is to introduce a scalar field B by the replacement

$$A_\mu \to A_\mu - \frac{1}{M} \partial_\mu B \tag{5.120}$$

to arrive at

$$\begin{aligned}\mathcal{L}_{St} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}M^2\left(A^\mu - \frac{1}{M}\partial_\mu B\right)^2 \\ &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\partial_\mu B\partial^\mu B + \frac{1}{2}M^2\ A^\mu A_\mu + MB\partial_\mu A^\mu\end{aligned}$$

(up to a boundary term after an integration by parts). This Lagrangian is gauge invariant with respect to

$$\delta A_\mu = \partial_\mu\theta, \qquad\qquad \delta B = \frac{1}{M}\theta.$$

The field equations are

$$[\mathcal{L}_{St}]^\nu = \partial_\mu F^{\mu\nu} - M^2\ A^\nu + M\partial^\nu B = 0, \quad [\mathcal{L}_{St}]_B = M^2\partial_\mu A^\mu - M\partial_\mu\partial^\mu B = 0$$

revealing immediately the Noether identity $\partial_\nu\,[\mathcal{L}_{St}]^\nu + [\mathcal{L}_{St}]_B \equiv 0$ and showing that the field equations for the scalar field B are just the derivatives of those for the vector field A_μ. Chosing the gauge $B = 0$ we regain the Proca Lagrangian. Both the Stückelberg and the Proca Lagrangian describe the three degrees of freedom of a massive spin-1 particle (in four dimensions).

In comparing the replacement (5.120) with (5.92), you anticipate that the Stückelberg trick can mimic the Higgs mechanism in the **U**(**1**) hypercharge sector of the SM, and indeed this has been worked out; details in [454]. Non-Abelian gauge theories cease to be renormalizable for massive vector fields.

Chapter 6
Particle Physics

Lynda Williams: "Quark Sing-a-long"

(Refrain) Up, Down, Charm, Strange, Top and Bottom! The World is made up of Quarks and Leptons! Up, Down, Charm, Strange,Top and Bottom! Yum! Yum!

Quarks come in six flavors. They live in families of two. Up Down, Charm Strange, Top and Bottom! They come in anti-flavors too!

Each family makes a generation between which is a mass gap. The up quark is the lightest and the top quark is the most fat!

The second and third generations do not live for very long. That's why everything in the Universe is made up of Ups and Downs!

Quarks carry a color charge. They come in red, green and blue. You'll never see a quark all by itself cuz they stick together with a strong force glue.

In particle physics a "Standard Model" has been established for roughly the past forty years. This model is from its very structure surprisingly simple (being based on internal symmetries) and has been experimentally justified to an astounding precision. Everything needed in terms of symmetries and field theory was derived in previous chapters. The "only" thing remaining to be done to arrive at the Standard Model is to fill this framework using the results from half-century high-energy accelerator experiments, giving clues about which internal symmetry group acts in elementary particle physics and which fields actually participate.

6.1 Particles and Interactions

6.1.1 Standard Model Constituents

Building the constituents of matter is as easy as assembling your furniture offered by a worldwide, well-known Swedish company: You must carefully consider the

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_6,

	First Generation	Second Generation	Third Generation	
Quarks	up down	charm strange	top bottom	matter particles
Leptons	electron electron-neutrino	muon muon-neutrino	tau tau-neutrino	
Gauge Bosons	Weak Interaction: (weakons) W^+, W^-, Z^0			force particles
	Electromagnetic Interaction: photon			
	Strong Interaction: 8 gluons			
Higgs particle				

Fig. 6.1 Standard model construction kit

construction kit with its list of parts, the connecting pieces, and the assembly instructions. The list of parts and the connecting pieces, together with a strange beast called the "Higgs" are given in Fig. 6.1.

- List of parts: "matter particles"
  (i) quarks and anti-quarks in three families and in three *colors/anti-colors*. The six quark types {up, down, strange, charm, bottom, top} are called *flavor*.
  (ii) leptons and anti-leptons in three families.
- Connecting pieces or "force particles", the gauge bosons
  (i) eight gluons as mediators of the strong interaction,
  (ii) three weakons $(W^{\pm}, Z^0)$ as mediators of the weak interaction,
  (iii) the photon γ as mediator of the electromagnetic interaction.
- The "Higgs" boson
  which–as will be shown in the sequel–in a well-defined sense results from symmetry breaking and is responsible for giving mass to the matter particles.

All particles–and as the majority of particle physicists believe – even the Higgs boson have been detected experimentally[1]. Quarks and gluons are identified only indirectly[2]. The dynamical theory of quarks and gluons (QCD: QuantumChromo-Dynamics) shows strong hints that these can not exist as free particles (catchword: *confinement*).
The properties of the elementary particles are (details in Fig. 6.2 and Fig. 6.3)

[1] On July 4th, 2012, CERN announced at first the discovery of a scalar boson with mass around 126 GeV identified by its decays and having properties of the Higgs. Further data collection and data evaluation for another year led to the opinion that indeed it was the long sought after Higgs boson. This prompted the Nobel prize committee to award the physics prize for the year 2013 to F. Englert and P.W. Kibble, two of the theoreticians that worked back in the 1960's on what became known as the Higgs mechanism.

[2] the last to be established were the top-quark in 1995 and the τ–neutrino in the year 2000

Matter Particles		Mass in MeV/c^2	electric charge	weak charge	strong charge
QUARKS					
1st. Generation	up	1,5 - 3,3	+2/3	1/2, + 1/2	r-g-b
	down	3,5 - 6,0	-1/3	1/2, - 1/2	r-b-g
2nd. Generation	charm	1160 - 1340	+2/3	1/2, + 1/2	r-b-g
	strange	70 - 130	-1/3	1/2, - 1/2	r-b-g
3rd. Generation	top	169100 - 173300	2/3	1/2, + 1/2	r-b-g
	bottom	4130 - 4370	-1/3	1/2, - 1/2	r-b-g
LEPTONS					
1st. Generation	electron-neutrino	< 2,2 eV/ c^2	0	1/2, + 1/2	-
	electron	0,511	1	1/2, - 1/2	-
2nd. Generation	muon-neutrino	< 0,17	0	1/2, + 1/2	-
	muon	105,7	1	1/2, - 1/2	-
3rd. Generation	tau-neutrino	< 15,5	0	1/2, + 1/2	-
	tau	1777	1	1/2, - 1/2	-

Fig. 6.2 Quarks and leptons

Force Particles	coupling at q=91 GeV/c	mass in GeV/c^2	electric charge	weak charge	strong charge
STRONG INTERACTION					
$g_1, g_2, g_3, g_4, g_5, g_6,$ g_7 g_8	g_s = 1,22	0	0	0	$r\bar{g}$, $r\bar{b}$, $b\bar{g}$, $b\bar{r}$, $g\bar{b}$, $g\bar{r}$ $r\bar{r} - b\bar{b}$ $r\bar{r} + b\bar{b} - 2g\bar{g}$
WEAK INTERACTION					
W- Boson W^+	g_w = 0,63	80,4	+1	1, +1	-
Z-Boson Z^0	g_z = 0,72	91,2	0	(1, 0)	-
W- Boson W^-	g_w = 0,63	80,4	-1	1, -1	-
ELECTROMAGNETIC INTERACTION					
Photon	g_{EM} = 0,31	0	0	(0, 0)	-

Fig. 6.3 Gauge bosons

- Mass
  The masses cover a broad range between 0 (as ascribed to the photon) and ≈ 170 GeV for the top quark, a mass which is comparable to the mass of a gold atom. Here the quark masses are to be understood as "current" masses, to be distinguished from "constituent" masses; more about this below.
- Spin
  (i) quarks and leptons are fermions with spin $^1\!/_2$
  (ii) gauge bosons have spin 1, the Higgs boson has spin 0.
- Charges
  (i) electrical charge Q: fractional (in units of $1/3\,e_0$) for quarks, integral for all others.
  (ii) weak charge/weak isospin $\vec{I} = (I^1, I^2, I^3)$ respectively $I := |\vec{I}|, I_3$, the values cited in the tables. For the Z^0 and the photon this charge is not defined in a strict

sense since they are a mixture of the bosons W^0 with $(I, I_3) = (1, 0)$ and B^0 with $(I, I_3) = (0, 0)$ (more details in Subsect. 6.3.3).
(iii) strong charge, characterized by colors {red, blue, green} for quarks and gluons. In Fig. 6.3. the notation $r\bar{g}$ indicates red–anti-green, etc. Leptons do not carry a strong charge and are therefore not involved in strong interaction processes.
All charges are additive (as scalars, vectors, ..) and charge conservation applies for all interactions.

- Couplings
  The gauge bosons have associated (energy dependent) coupling "constants" typical for the strength of the interaction. The electromagnetic interaction is characterized by the fine structure "constant".

$$\alpha = \frac{e^2}{\hbar c(4\pi\varepsilon_0})\simeq\frac{1}{137},$$

from which $g_{EM} := e = \sqrt{4\pi\alpha} = 0.31$, and similarly $g_W = 0.63, g_Z = 0.72$, as well $g_S = 1.22$.

6.1.2 Quarks as Building Blocks of Hadrons

Aside from the term lepton (derived from the Greek "light weight"), in the 1950's one used the terms mesons ("medium weight") and baryons ("heavy weight") for classifying nuclear and subnuclear particles. However, today mesons and baryons are no longer considered elementary particles. Instead mesons are understood as bound states of a quark and an anti-quark, and baryons are bound states of three quarks in a colorless/white (w) combination. Examples are

$$\text{Proton} : (uud)_w \qquad \pi^+ : (u\bar{d})_w \qquad K^+ : (u\bar{s})_w \qquad J/\psi : (c\bar{c})_w.$$

For historical reasons, mesons and baryons are *hadrons*, originally denoting strongly interacting particles. Today we know that protons and neutrons in the nuclei of atoms are not bound directly by strong forces, but by van-der-Waals like effects of the gluon exchange between the quarks.

You may wonder how three "tiny" up and down quarks with masses of the order MeV can make up a proton with a mass of roughly 1 GeV. In a comment to the quark masses in Fig. 6.2. I remarked that these are meant as the "current" masses. These are the ones which appear in the Lagrangian of the strong interaction model. They are the quantities which are probed when a hadron experiences weak or electromagnetic processes. And they are to be distinguished from the "constituent" masses, where the name refers to quarks being the constituents of hadrons, the proton, say. Three quarks constituting the proton we get a constituent mass in the order of $M_S \sim GeV$. On this mass scale the "current masses" of the u- and d-quark are nearly the same, they even may be set to zero. The difference between the current and the constituent masses comes from the gluons that energetically bind the constituent quarks together, or, to use another picture, the constituent quark mass refers to the current quark mass plus the energy of the gluons surrounding the constituent quark. Of course this is only a picture based on our imagination of screening of electrical charges. Within the Standard Model, the difference of current and constituent masses is explained by chiral symmetry breaking, a mechanism explained in 6.4.3.; see for instance [17].

In the 1950's it was recognized that the many particles which had been discovered show specific patterns in their scattering and/or decay behavior. This led to the heuristic definition of numbers which served to explain the existence or non-existence of certain scattering and decay processes. Among these were for example the *baryon number* B and the *lepton number* L: A lepton carries the lepton number $L = +1$, but zero baryon number. An anti-lepton has $L = -1$. Quarks have baryon number $B = +1/3$ (this choice, as will soon be seen, gives baryons the baryon number one and mesons baryon number zero). The net baryon and the net lepton number appear to be conserved (as is of course the electrical charge) in all elementary particle scattering and decay processes. This should ring a bell to those acquainted with Noether's theorem, since conserved quantities originate from conserved currents, currents that may arise from global symmetries. And indeed, we will recover the baryon and the lepton number in an altogether different light. Baryon number conservation implies the stability of the Universe: The lightest baryon is the proton, and thus every conceivable decay process involves leptons. Therefore, the rule of baryon number conservation prevents for instance all the protons in the universe from gradually changing into positrons (plus other leptonic stuff)! Very helpful in the early history of particle physics was also a number called *strangeness* S which is conserved in all strong and electromagnetic processes, but follows a $|\Delta S| = 1$ rule for weak interaction processes. In hindsight the attribute strangeness is given to the strange quark s, for historical reasons as $S(s) = -1$, $S(\bar{s}) = +1$.

The first attempts to bring order into the zoo of the ever-increasing number of particles and resonances were based on the idea of W. Heisenberg (1932) that the proton and the neutron constitute a douplet of a new symmetry, namely the $\mathbf{SU(2)}$ isospin group. However more spin $^1\!/_2$ baryons were found in the 1950's: three Σ's and the Λ^0 (collectively called by the name hyperons), and two Ξ's with strangeness $S = -2$. So the question arose, of whether there exists a symmetry group into which these eight baryons could fit. Investigations at that time covered among others[3]

- **SO(7)** has an 8-dimensional representation, which can be directly associated with the eight baryons known at that time (J. Tiomno).
- **SO(4)** allows for a direct sum of representations as $\mathbf{4} \oplus \mathbf{4}$. The three Σ's and the Λ were associated to one of the 4-dimensional representations, the two nucleons (p and n) and the Ξ's to the other (J. Schwinger).
- The exceptional group $\mathbf{G_2}$ has a representation splitting as $\mathbf{7} \oplus \mathbf{1}$ (R. Behrends and A. Sirlin).

[3] I dispense with giving the full references of the authors in the following list since these approaches are now of historical interest only. But I mention them here to show how keen particle physicists have been in symmetry argumentation.

- **SU(3)** has a representation splitting as $\mathbf{3} \oplus \mathbf{6}$: The idea was to arrange (p, n, Λ) into a triplet, and the remaining ones into an incomplete sextet (S. Sakata).

All these approaches aimed at putting the baryons directly into the representation of some group. In 1961, M.Gell-Mann[4] and independently G. Zweig again proposed an **SU(3)** symmetry group. But in their model three constituents–the quarks (up, down, strange)–make up the fundamental **3** representation of $\mathbf{SU}(3)$:

$$\mathbf{3} = \begin{pmatrix} u \\ d \\ s \end{pmatrix}. \tag{6.1}$$

This is the defining representation of $\mathbf{SU}(\mathbf{3})$ in terms of the Gell-Mann λ-matrices (see Appendix A.3.4) or their Hermitean conjugates. The objects which transform under these representations are three-component $\mathbf{SU}(\mathbf{3})$-vectors ψ, i.e. $\psi' = e^{i(\alpha^a \lambda_a)}\psi$. In that case, compliant with

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \tag{6.2a}$$

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \tag{6.2b}$$

mesons can be arranged into octets, and baryons into octets and decuplets. This will be illustrated for the meson octet: If one performs the reduction of $\mathbf{3} \otimes \overline{\mathbf{3}}$ with the $\mathbf{SU}(3)$-vectors ψ one gets according to (A.32) the octet matrix

$$M^{(8)} = \begin{pmatrix} \left(2u\bar{u} - d\bar{d} - s\bar{s}\right)/3 & u\bar{d} & u\bar{s} \\ d\bar{u} & (2d\bar{g} - u\bar{u} - s\bar{s})/3 & d\bar{s} \\ s\bar{u} & s\bar{d} & (2\,s\bar{s} - u\bar{u} - d\bar{d})/3 \end{pmatrix}. \tag{6.3}$$

The non-diagonal elements of this matrix can directly be identified with known particles of definite isospin and strangeness. These are the pseudoscalar mesons

$$\begin{aligned} u\bar{d} &= \pi^+ & \qquad d\bar{u} &= \pi^- \\ u\bar{s} &= K^+ & s\bar{u} &= K^- \\ d\bar{s} &= K^0 & s\bar{u} &= \bar{K}^0. \end{aligned}$$

Since $M^{(8)}$ is traceless, only two of its diagonal components are independent. These correspond to two further mesons. One of them is the third partner in the π-meson isotriplett, the π^0 with $I_3 = 0$:

$$\pi^0 = \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$$

[4] 1969 Nobel prize in appreciation "for his contributions and discoveries concerning the classification of elementary particles and their interactions".

which is its own antiparticle. The other meson state, chosen to be orthogonal to π^0 (and compliant with $\operatorname{tr} M^{(8)} = 0$) is

$$\eta = \frac{1}{\sqrt{6}}(2s\bar{s} - u\bar{u} - d\bar{d}).$$

The meson singlet built as $(u\bar{u} + d\bar{d} + s\bar{s})$ is identified with the η' pseudoscalar meson

The pattern according to (6.2a, 6.2b) was indeed observed (see below). However, at that time in this scheme, called "the eightfold way" by Gell-Mann, a baryon with electrical charge (-1) and strangeness (-3) was missing, the so-called Ω^-. Its discovery in 1964 paved the way for the establishment of the quark model. The symmetry becomes particularly apparent if particles are arranged in a $(I_3, \bar{Y})$ plane, where I_3 is the third component of the isospin and the hypercharge $\bar{Y}$ is defined simply as the sum of the baryon number and the strangeness $\bar{Y} := B + S$. According to the Gell-Mann/Zweig quark model the baryon number could more appropriately be called the total quark number since $B = 3\{N(q) - N(\bar{q})\}$, where $N(q)$ denotes the number of quarks in the baryon or the meson.

The entities I_3 and $\bar{Y}$ are commuting elements of the algebra of the $\mathbf{SU(3)}$-generators (see Appendix A.3.4):

$$I_3 = F_3, \qquad \bar{Y} = \frac{2}{\sqrt{3}} F_8.$$

They are related to the electrical charge by the Gell-Mann-Nishijima relation

$$Q = I_3 + \frac{\bar{Y}}{2}. \tag{6.4}$$

For quarks, with $B = 1/3$, this implies charges in units of $1/3$ of the charge of an electron, e_0. In the (I_3, Y)-plane the (d, u, s)-quarks and $(\bar{d}, \bar{u}, \bar{s})$-quarks are represented as triplets; see Fig. 6.4a. The octet splits into a $\mathbf{SU(2)}$ triplet and two $\mathbf{SU(2)}$ doublets. The meson and the baryon octets are shown in Fig. 6.4b, respectively. Finally the assignment of particles and resonances to the positions in the decuplet is shown in Fig. 6.4c.

Although these diagrams are quite impressive and suggest that $\mathbf{SU(3)}$ is a *de facto* symmetry group, this symmetry cannot be exact since the particles in the multiplets have different charges and masses. The disparity in electrical charge is not relevant if the symmetry is assumed to hold with respect to strong interactions. But, not only the masses of the quarks

$$m_u \approx 2.5\ MeV, \qquad m_d \approx 4.5\ MeV, \qquad m_s \approx 100\ MeV$$

are different, but as seen from Fig. 6.5, the masses of mesons in the octet and singlet range from 135 to 1020 MeV, and there is also a large mass variation in the baryon multiplets. The symmetry would be exact only in case of equal masses. In any case, be aware on a previous remark on "current" and "constituent" masses. Although the "current" mass of a strange quark is large compared to the masses of the u- and d-quark, it is still differs by a factor of 10 from the characteristic mass M_S of strong interactions. Given this, many relations were derived in the sixties from the "near" exactness of the three quark masses, like the Gell-Mann-Okubo mass formula for the members of a multiplet. The older books on symmetries in particle physics (a typical representative is [166]) deal extensively with these relations; see also [211]. As a matter of fact, Gell-Mann was not only able to predict the quantum numbers of the yet unknown Ω^- but also its mass of approximately 1.68 GeV.

This story did not end with the (u, d, s)-quarks. In the period that followed the success of the "eightfold way", S. Glashow, J. Iliopoulos, and L. Maiani postulated in 1970 the existence of a fourth quark c (termed the "charm" quark) in order to remedy problems with the description of weak interactions in a clever manner (see Sect. 6.3.2). Indeed a narrow resonance discovered by two experimental teams in 1974 could only be interpreted as a bound state of a c-quark and a $\bar{c}$-quark[5]. With this new quark, the symmetry group had to be extended from $\mathbf{SU(3)}$ to four flavors in $\mathbf{SU(4)}$. Already in 1973 M. Kobayashi and T. Maskawa postulated the existence of a further family of quarks in order to formulate a consistent model for the CP violation of weak interactions (see Sect. 6.3.2). And indeed the b ("bottom" or "beauty")[6] and the t ("top" or "truth") quark were eventually discovered in 1977 and 1995, respectively. With these new flavors, the symmetry group became $\mathbf{SU(6)}$. This symmetry is even more approximate than the symmetry of the former "eightfold way", since

$$m_c \approx 1.25\, GeV, \quad m_b \approx 4.2\, GeV, \quad m_t \approx 171 GeV.$$

Although the mass of the c-quark is already in the order of the hadronic scale, astonishingly enough the allocation of hadrons with charm components into the multiplets

$$\mathbf{4} \otimes \overline{\mathbf{4}} = \mathbf{15} \oplus \mathbf{1}$$
$$\mathbf{4} \otimes \mathbf{4} \otimes \mathbf{4} = \mathbf{20} \oplus \mathbf{20} \oplus \mathbf{20} \oplus \overline{\mathbf{4}}$$

works quite well, as seen in Fig. 6.6 and Fig. 6.7. Other conclusions derived from the approximate symmetry also conform with experiments.

To each additional flavor degree of freedom, we can formally associate a further quantum number, namely C(harmness), $\tilde{B}$(ottomness), T(opness), and the previously

[5] This example of a "charmonium" is the only particle carrying two names: It is the J/ψ, the letters standing for the discovery of the US East coast team with S. Ting and the West coast team with B. Richter.

[6] Of course you realize that the names of the quarks are not at all representatives of their attributes. I suppose that the naming of new phenomena and new entities in physics is a very interesting cultural and sociological theme to investigate.

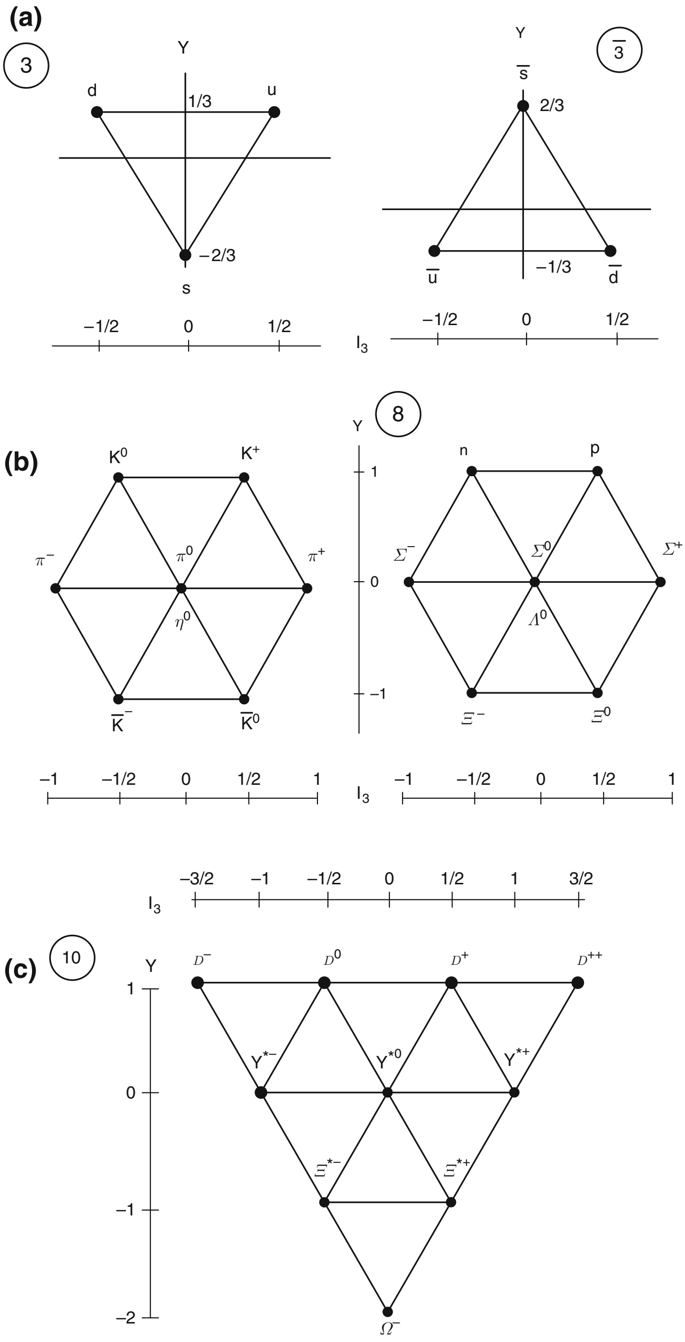


Fig. 6.4 SU(3) and particle physics

(a)

$q_i\bar{q}_j$	$J = 0$	$J = 1$
$\|u\bar{d}\rangle$	$\pi^+(140)$	$\rho^+\,(770)$
$2^{-1/2}\|d\bar{d} - u\bar{u}\rangle$	$\pi^0(135)$	$\rho^0\,(770)$
$\|u\bar{d}\rangle$	$\pi^-(140)$	$\rho^-\,(770)$
$2^{-1/2}\|d\bar{d} + u\bar{u}\rangle$	$\eta(549)$	$\omega(783)$
$\|u\bar{s}\rangle$	$K^+(494)$	$K^{*+}(892)$
$\|d\bar{s}\rangle$	$K^0(498)$	$K^{*0}(892)$
$\|\bar{u}s\rangle$	$K^-(494)$	$K^{*-}(892)$
$\|\bar{d}s\rangle$	$\bar{K}^0(498)$	$\bar{K}^{*0}(892)$
$\|s\bar{s}\rangle$	$\eta'(958)$	$\phi(1020)$

(b)

$q_iq_jq_k$	$J = 1/2$	$J = 3/2$
$\|uuu\rangle$		$\Delta^{++}(1230)$
$\|uud\rangle$	$p(938)$	$\Delta^+(1231)$
$\|udd\rangle$	$n(940)$	$\Delta^0(1232)$
$\|ddd\rangle$		$\Delta^-(1234)$
$\|uss\rangle$	$\Sigma^+(1189)$	$\Sigma^+(1383)$
$2^{-1/2}\|(ud + du)s\rangle$	$\Sigma^0(1192)$	$\Sigma^0(1384)$
$\|dds\rangle$	$\Sigma^-(1197)$	$\Sigma^-(1387)$
$2^{-1/2}\|(ud - du)s\rangle$	$\Lambda(1116)$	
$\|uss\rangle$	$\Xi^0(1315)$	$\Xi^0(1532)$
$\|dss\rangle$	$\Xi^-(1321)$	$\Xi^-(1535)$
$\|sss\rangle$		$\Omega^-(1672)$

Fig. 6.5 Mesons and baryons from u/d/s-Quarks

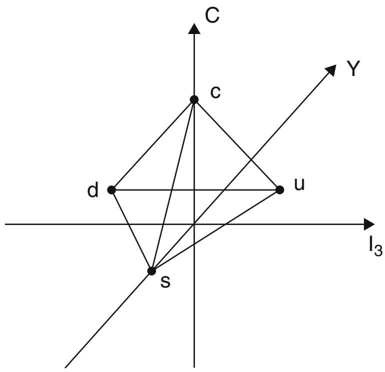


Fig. 6.6 SU(4) multiplet

defined hypercharge then contains further terms:

$$Y = B + S + C + \tilde{B} + T.$$

From today's perspective, it also would make sense to introduce D(ownness) and U(pness) instead of the historically-motivated assignment of isospin and its third component. Each of the numbers counts the number of flavors and anti-flavors, e.g. $U = N(u)\text{-}N(\bar{u})$, $D = N(d)\text{-}N(\bar{d})$, $S = -\{N(s)\text{-}N(\bar{s})\}$ (the minus sign coming

(a)

$q_i\bar{q}_j$	$J=0$	$J=1$
$c\bar{c}$	$\eta_c(2980)$	$J/\psi(3097)$
$c\bar{d}$	$D^+(1869)$	$(D^*)^+(2010)$
$c\bar{u}$	$D^0(1865)$	$(D^*)^0(2007)$
$\bar{c}u$	$\bar{D}^0(1865)$	$(\bar{D}^*)^0(2007)$
$\bar{c}d$	$D^-(1869)$	$(D^*)^-(2010)$
$c\bar{s}$	$D_s^+(1970)$	$D_s^{*+}(2109)$
$\bar{c}s$	$D_s^-(1970)$	$D_s^{*-}(2109)$

Mesons with charm

(b)

$q_i\bar{q}_j$	$J=0$	$J=1$
$b\bar{b}$	η_b	$\Upsilon(9460)$
$u\bar{b}$	$B^+(5278)$	$(B^*)^+(5324)$
$d\bar{b}$	$B^0(5278)$	$(B^*)^0(5324)$
$\bar{u}b$	$\bar{B}^0(5278)$	$(\bar{B}^*)^0(5324)$
$\bar{d}b$	$B^-(5278)$	$(B^*)^-(5324)$

Mesons with bottom

Fig. 6.7 Mesons from c-quarks (a) and from b-quarks (b)

from the assignment of strangeness -1 to the strange quark). It is an easy algebraic exercise to express both the baryon number and the charge as

$$B = \frac{1}{3}(D+U-S+C-\tilde{B}+T) \qquad Q = \frac{2}{3}(U+C+T)-\frac{1}{3}(U-S-\tilde{B}). \quad (6.5)$$

The F-ness numbers are conserved in all strong interaction processes, since the flavor identities are untouched. Only in weak interaction processes is the flavor allowed to change, as for instance in beta decay where a d-quark transforms into a u-quark, and by this a neutron decays into a proton, an electron and an electron-antineutrino, see Fig. 6.9b.

In 1971 M. Gell-Mann and H. Fritzsch "re-animated" the symmetry group $\mathbf{SU(3)}$, but gave it another twist. They introduced the concept of color charge and proposed that instead of the flavor degrees of freedom, the colors of quarks are representatives of an exact $\mathbf{SU(3)}$ symmetry. Instead of (6.1) the correct assignment is now

$$\mathbf{3} = \begin{pmatrix} r \\ g \\ b \end{pmatrix}.$$

Since no colored hadrons are known to exist, hadrons must be composed of colored quarks in specific combinations. In labeling by q_c a quark with flavor q $(q = u, d, s, c, b, t)$ and color c:

$$\text{meson} \equiv \; q_c\,\bar{q}'_c \qquad\qquad \text{baryon} \equiv \; \epsilon^{cc'c''} q_c\, q_{c'}\, q_{c''}. \quad (6.6)$$

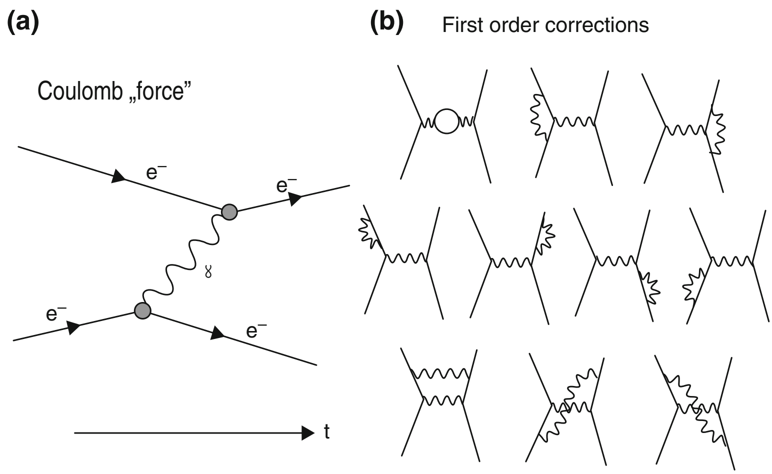


Fig. 6.8 Coulomb interaction

These combinations are possible because $\mathbf{SU(3)}$ contains a singlet in its product representations (compare (6.2a) and (6.2b)). The inhibition of the octet and decuplet representations is not yet completely understood. It could either be inherently caused within QCD (for instance because only the singlet states are stable) or else it is related to a large mass gap between the energetically favored singlet states and the other multiplet states.

Other combinations such as diquarks ($q_c\, q'_c$) (i.e. mesons with baryon number $^2/_3$) have not been observed. As we will see, they are excluded because the theory of strong interaction is a gauge theory of the color group. Other beasts like the tetra-quark (shorthand: $(q\bar{q})(q\bar{q})$) or the pentaquark (shorthand: $(qqq)(q\bar{q})$) are not forbidden by QCD and are the object of experimental searches and theoretical consideration. It may even be the case that in an effective field theory approximation of QCD the mesons and baryons are not simply the combinations (6.6) but superpositions with these bound states.

Playing with the color group representation we observe from $(\mathbf{8}\otimes\mathbf{8})_S = (\mathbf{1}\oplus\ldots)$ and $(\mathbf{8}\otimes\mathbf{8}\otimes\mathbf{8})_S = (\mathbf{1}\oplus\ldots)$ that there should exist colorless combinations of gluons, called glueballs. So far they have not been observed or identified with certainty.

6.1.3 Interaction Processes

QED (Quantum Electro Dynamics) was the first successful quantum field theory. Therein the hitherto established perception that two charged objects exert a Coulomb force on one another was substituted by a picture that their mutual interaction is mediated by the exchange of photons. The Feynman graph Fig. 6.8a does not only illustrate this process, but comprises recipes for calculating its amplitudes and cross section.

The same holds for correction terms of first and higher order as in Fig. 6.8b. This subsection serves to convey the logical arguments relating the S-matrix, transition amplitudes, Green's functions and Feynman diagrams.

In Subsect. 5.3 I briefly explained how free fields are quantized. In the case of interaction one needs to resort to perturbative techniques. This is most appropriately formulated in the Dirac picture of quantum physics: One splits the total Hamiltonian of the interacting system (take QED with interacting electron, positron and photon fields) into a free and an interaction part as $\mathcal{H} = \mathcal{H}_F + \mathcal{H}_I$. States are described as

$$|\psi(t)\rangle_I = e^{i\mathcal{H}_F t}|\psi(t)\rangle$$

which evolve according to (4.7) as

$$i\frac{d}{dt}|\psi(t)\rangle_I = \mathcal{H}_I(t)|\psi(t)\rangle_I.$$

This has the formal solution

$$|\psi(t)\rangle_I = |\psi(-\infty)\rangle_I - i\int_{-\infty}^{t} d\tau\, \mathcal{H}_I(\tau)|\psi(\tau)\rangle_I.$$

Experiments in particle physics deal with decay and scattering processes. In a scattering experiment some particles move initially freely, then they interact during a limited time span, and afterwards newly created particles escape the interaction region as free particles. This setting is embodied by the concept of a scattering operator (also scattering matrix) S. It is defined as the operator that connects the initial state and the final state via

$$|\psi(+\infty)\rangle_I = S\,|\psi(-\infty)\rangle_I$$

and it is a limiting case of the Green's function:

$$S = G(+\infty, -\infty), \qquad\qquad i\partial_t G(t, t_0) = \mathcal{H}_I(t)G(t, t_0).$$

By solving the differential equation for the Green's function, one arrives at

$$S = 1 - i\int_{-\infty}^{+\infty} dt\, \mathcal{H}_I(t) + \frac{(-i)^2}{2!}\int_{-\infty}^{+\infty} dt \int_{-\infty}^{+\infty} dt'\; P(\mathcal{H}_I(t)\mathcal{H}_I(t')) + \cdots$$

where P denotes the time-ordering operation

$$P(\mathcal{H}_I(t)\mathcal{H}_I(t')) = \begin{cases} \mathcal{H}_I(t)\mathcal{H}_I(t') & \text{for} \quad t < t', \\ \mathcal{H}_I(t')\mathcal{H}_I(t) & \text{for} \quad t > t'. \end{cases}$$

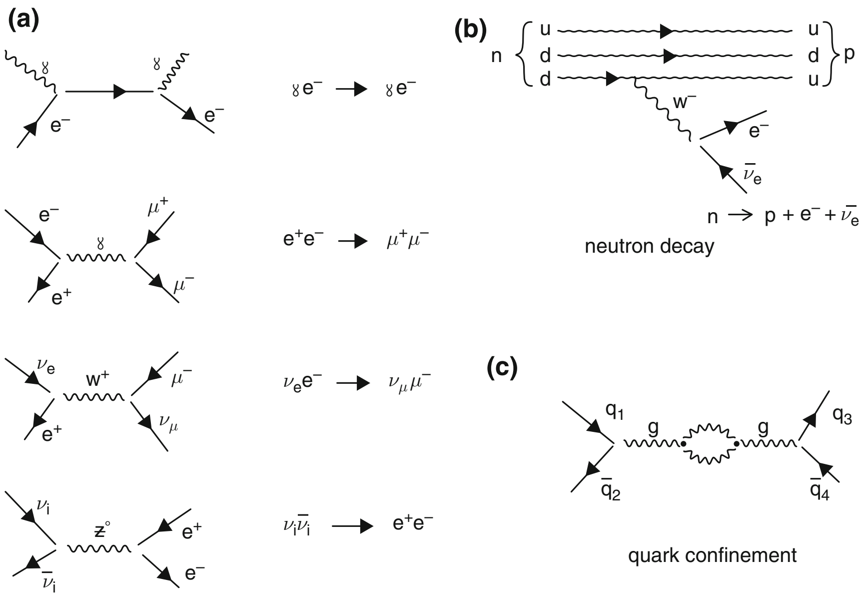


Fig. 6.9 Examples of scattering processes

The Feynman graphs are a graphical representation of the infinite sum expression of the S-matrix, the first non-trivial term representing the lowest order interaction process. The graphs comprise computational prescriptions, and eventually make it possible to calculate transition amplitudes and cross sections for scattering processes.

This picture by which the primary entities are not forces, but instead that exchange particles are mediators could successfully be applied also to the strong and weak interactions within the Standard Model. The two lower Feynman graphs in Fig. 6.9a depict processes with exchange of "weakons". Also, β decay obtains a concise representation, in that a down quark within the neutron decays into an up quark and a W^- which in turn decays into an electron and an electron anti-neutrino, illustrated in Fig. 6.9b. A peculiarity arises for the gluon exchange in strong interaction, in that gluons can couple to themselves (Fig. 6.9c), a further hint to the confinement of quarks.

One of the testing grounds of the Standard Model and its possible extensions is the calculation of the anomalous magnetic moment $a = (g\text{–}2)/2$ of the muon. Experiments at the Brookhaven National Laboratory reach a precision of 0.5 parts per million, and theoretical predictions become possible with the same precision. Fig. 6.10 is taken from [53]. This should convey a feeling for the level of precision when discussing the extremely good agreement between SM predictions and experiment.

6.1.4 Lagrangian of the Standard Model

QCD and QFT

The Standard Model has two facets, one of them describing strong interactions (QCD) and the other one describing electroweak processes. Details follow in the next two sections. At this point–mainly for orientation–we consider only the following:

- QCD (QuantumChromoDynamic) is a Yang-Mills theory with gauge group $\mathbf{SU_C(3)}$, where the index C refers to the color charges. The fermion sector is made up of quarks.
- QFD (QuantumFamily/FlavorDynamics) is the Glashow-Salam-Weinberg model of the pseudo-unified electromagnetic and weak interactions. It is a Yang-Mills theory with a Higgs potential; the gauge group is $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ with charges I (weak isospin) and Y (weak hypercharge), spontaneously broken to $\mathbf{U_{EM}}(1)$. The fermion sector is made up of leptons and quarks.

Gauge Bosons, Quarks and Leptons, Higgs

The Standard Model Lagrangian consists of three terms:

$$\mathcal{L} = \mathcal{L}_{\text{gauge bosons}} + \mathcal{L}_{\text{fermions}} + \mathcal{L}_{\text{Higgs}}.$$

Since $\mathbf{SU(3)} \times \mathbf{SU(2)} \times \mathbf{U(1)}$ is a $(8+3+1=12)$ parameter group with twelve gauge bosons, we find eight *gluon*s, three *weakon*s ($W^{\pm}$, Z^0), and the "photon" γ in $\mathcal{L}_{\text{gauge bosons}}$. Later, we will see that originally–prior to spontaneous breakdown via the Higgs mechanism–there are four electro-weakons. The fermion part $\mathcal{L}_{\text{fermions}}$ comprises both the quarks and the leptons.

Fields and Covariant Derivatives

A more detailed structure of the Lagrangian in terms of gauge fields (A), fermions (ψ), and Higgs scalars (ϕ) is

$$\mathcal{L} = \mathcal{L}_{\text{YM}}(A) + \mathcal{L}_{\text{Dirac}}(A, \psi, \overline{\psi}) + \mathcal{L}_{\text{scalar}}(A, \phi, \psi, \overline{\psi}). \tag{6.7}$$

Here

$$\mathcal{L}_{YM}(A) = -\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a \tag{6.8a}$$

$$\mathcal{L}_{Dirac}(A, \psi, \overline{\psi}) = i\overline{\psi}\not{D}\psi \tag{6.8b}$$

$$\mathcal{L}_{scalar}(A, \phi, \psi, \overline{\psi}) = (D\phi)^2 + V(\phi) + Y(\overline{\psi}, \psi, \phi) \tag{6.8c}$$

with covariant derivatives

$$D_\mu(A, g) = \partial_\mu + igA^a_\mu T^a$$

		contribution to a_μ $(\times 10^{11})$
QED	coefficient of $\left(\frac{\alpha}{\pi}\right)^n$	
$n=1$	0.5	116140972.9(0.4)
$n=2$	0.765857376(27)	413217.6
$n=3$	24.05050898(44)	30142.4
$n=4$	126.07(41)	366.9(1.2)
$n=5$	930(170)	6.3(1.1)
QED total		116584705.7(1.7)
hadronic		6739(67)
weak		
one-loop		195
two-loop		−43(4)
gravitation [Be75]		4×10^{-30}
total result		$a_\mu^{\text{th}}(\times 10^{11}) = 116591597(67)$

Fig. 6.10 Contributions to the anomalous magnetic moment of the muon

and the field strength tensor

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f^{abc} A^b_\mu A^c_\nu .$$

As explained in Sect. 5.3.4 and as detailed in the subsequent two sections, the explicit form of the covariant derivative $D_\mu(A, g)$ not only depends on the gauge group, but also on the transformation properties of the fields ψ and φ. Since the gauge group of the Standard Model is the product of three groups, there are three independent coupling constants $(g_C, g_I, g_Y) \equiv (g_S, g_W, g_B)$.

6.2 Strong Interactions

6.2.1 Lagrangian of Quantum Chromo Dynamics

The strong interaction is described in the Standard Model by an $\mathbf{SU(3)}$ Yang-Mills theory[7] with eight gauge bosons A^a_μ (the gluons) and a number of quark fields, which are Dirac spinors carrying a flavor q and a color index c, denoted by q_c.

---

[7] Conceiving the strong interaction as a YM-theory is going back to 1972 with work of M. Gell-Mann, H. Fritzsch, and H. Leutwyler, see e.g. [199].

Being of Yang-Mills type and given the gauge group, the Lagrange density of QCD is determined to be

$$\begin{aligned}\mathcal{L}_{QCD} &= -\frac{1}{4}F^2 + \bar{q}(i\not{D} - m_q)q \\ &= -\frac{1}{4}\sum_a F^a_{\mu\nu}F^{\mu\nu}_a + i\sum_{q,c,c'}\bar{q}_c\gamma^\mu (D_\mu)_{cc'}q_{c'} - \sum_{q,c} m_q\bar{q}_c q_c.\end{aligned} \tag{6.9}$$

The a-summation ranges over the eight gauge field indices, the c-summation(s) over the three color indices, and the q-summation over the N_f flavors. The $F^a_{\mu\nu}$ and the $(D_\mu)_{cc'}$ are the Yang-Mills fields and the covariant derivatives with respect to the local $\mathbf{SU_C(3)}$ symmetry:

$$\begin{aligned}F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_S\epsilon^{abc}A^b_\mu A^c_\nu \\ (D_\mu)_{cc'} &= \delta_{cc'}\partial_\mu + ig_S\frac{\hat{\lambda}^a_{cc'}}{2}A^a_\mu,\end{aligned}$$

where $\hat{\lambda}^a_{cc'}$ are matrices in the adjoint representation of $\mathbf{SU(3)}$. The m_q are the quark masses, g_S is the coupling constant of the strong interaction, that is the quark-gluon coupling. Aside from these constants and the number N_F of flavors, the Lagrangian (6.9) is completely fixed by the requirements of color gauge invariance and renormalizability.

6.2.2 Symmetries of QCD

(1) To start with, the Lagrangian (6.9) is invariant under Lorentz transformations including space reflections P and time reversal T.
(2) By construction as a gauge theory, the QCD Lagrangian is invariant under the color gauge group $\mathbf{SU_C(3)}$

$$q \to e^{i\theta^a(x)\lambda_a}q, \qquad q = \begin{pmatrix} q_r \\ q_b \\ q_g \end{pmatrix},$$

where $\lambda^a_{cc'}$ are the (3×3) matrices in the defining representation of $\mathbf{SU(3)}$. This is the local invariance of QCD with respect to color transformations.
(3) Further global symmetries act in the flavor space. Each flavor field might transform according to

$$q \to e^{i\varphi_q}q.$$

This gives rise to N_F currents $J^\mu_q = i\bar{q}\gamma^\mu q$. The associated charges are the flavor numbers F-ness. This is conservation of flavor type.

(4) If $m_u = m_d$ there is a global $\mathbf{SU(2)}$ symmetry

$$\begin{pmatrix} u \\ d \end{pmatrix} \mapsto e^{i\sigma^k \phi_k} \begin{pmatrix} u \\ d \end{pmatrix}$$

where σ^k are the Pauli matrices. Since on the hadronic scale, the masses of the up and down quarks may be considered equal, we indeed recover from QCD the successful "old" reflections on $\mathbf{SU_f(2)}$ (flavor) symmetry. This generalizes to further quark flavors, but it is clear that the more massive the quark becomes, the more broken is the associated flavor symmetry. This was visible at the span of masses in the multiplets of $\mathbf{SU_f(3)}$ and $\mathbf{SU_f(4)}$
(5) In the limit of vanishing quark masses, QCD is invariant under

$$q \to e^{i\theta} q \qquad q \to e^{i\gamma_5 \theta} q \tag{6.10}$$

$$q \to e^{i\tau^a \theta_a} q \qquad q \to e^{i\gamma_5 \tau^a \theta_a} q \tag{6.11}$$

where τ_a are from the defining representation of $\mathbf{SU(N_f)}$. This symmetry can be better organized by introducing chiral fields $q_L = \frac{1}{2}(1-\gamma^5)q$, $q_L = \frac{1}{2}(1-\gamma^5)q$. Neglecting the mass term, the quark part of the QCD Lagrangian becomes

$$i\bar{q}_L \not{D} q_L + i\bar{q}_R \not{D} q_R.$$

This term is invariant under two separate unitary transformations $q_L \to U_L q_L$ and $q_R \to U_R q_R$. Thus the theory exhibits a $\mathbf{U_L(N_f)} \times \mathbf{U_R(N_f)}$ symmetry. Now $\mathbf{U(N)}$ is isomorphic to $\mathbf{SU(N)} \times \mathbf{U(1)}$, and therefore the transformations (6.10) are identified with two $\mathbf{U(1)}'s$ and (6.11) are identified with two $\mathbf{SU(N_f)}'s$. These symmetries, holding for vanishing current quark masses, cannot be realized exactly. But they may hold approximately for the lightest quarks $\{u, d, s\}$. Therefore one may expect the conserved vector currents and axial vector currents

$$\mathbf{SU_V(3)} \times \mathbf{U_V(1)}: \qquad j_a^\mu = \sum_{u,d,s} \bar{q}\gamma^\mu \frac{\tilde{\lambda}^a}{2} q \qquad j^\mu = \sum_{u,d,s} \bar{q}\gamma^\mu q$$

$$\mathbf{SU_A(3)} \times \mathbf{U_A(1)}: \qquad j_a^{5\mu} = \sum_{u,d,s} \bar{q}\gamma^\mu \gamma^5 \frac{\tilde{\lambda}^a}{2} q \qquad j^{5\mu} = \sum_{u,d,s} \bar{q}\gamma^\mu \gamma^5 q$$

where the $\tilde{\lambda}^a$ are eight $\mathbf{SU(3)}-$ matrices in the u, d, s flavor space. The V-symmetries and their associated charges result in flavor invariance and baryon number conservation. But the $\mathbf{SU_L(3)} \times \mathbf{SU_R(3)}$ chiral symmetry does not hold even approximately in nature. By the breaking of flavor chiral symmetry, one understands the mechanism through which in the end only the $\mathbf{SU_V(3)}$ survives. It is believed and supported by experimental results that this mechanism is due to spontaneous symmetry breaking. Here I refer to textbooks in particle physics, where you learn that "the pion is the Goldstone mode of chiral symmetry breaking", and that this mechanism is a reason for many experimentally verified relations among hadronic processes involving the u, d, s-quarks.

6.2.3 Theoretical Consistency and Experimental Support

The “Advantages of the Color Octet Gluon Picture” were elucidated for the first time in [199]. Let me mention the following:

- Spin-statistic problem
  According to the spin-statistic theorem, spin-$^1\!/_2$ fields must be totally antisymmetric under exchange of their quantum numbers. For quite some time the problem persisted that certain baryons seem to violate this theorem. Consider for instance the $\Delta^{++} = (uuu)$. Its wave function is a product of three pieces:

$$\Psi_{\Delta^{++}} = \Psi_{SU_f(3)}\Psi_{\text{orbit}}\Psi_{\text{spin}}.$$

  The flavor part is symmetric, since the Δ^{++} is composed from three identical quarks. The spin part is symmetric, since with $J = \frac{3}{2}$ all spins point in the same direction. The orbital angular momentum part behaves as a factor $(-1)^L$, and thus is likewise symmetric because of $L = 0$. Only with a further antisymmetric part in the wave function would the spin-statistics problem be solved. And indeed, a color part $\Psi_{SU_c(3)} = \epsilon^{cc'c''}u_c u_{c'} u_{c''}$ serves this purpose and saves the spin-statistics theorem.
- Number of colors and of generations
  Already the previous argument is a strong point for three colors. But there is also an experimental confirmation:
  The total cross section of electron-positron pair annihilation into hadrons

$$e^+ + e^- \to \gamma \to q + \bar{q} \to \text{hadrons}$$

  depends on the number and the charges of the quarks. One specifically derives for the ratio of cross sections

$$R = \frac{\sigma(e^+ + e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} = N_c \sum_q Q_q^2.$$

  This is a function of energy. Below the threshold energy for the creation of a quark pair, this and heavier quarks cannot contribute to R. Thus the threshold behavior of R and its absolute value provide an estimate for $N_c \approx 3$.
  Another estimate derives from the pion decay rate $\pi \to \gamma + \gamma$. The Feynman graph for this process contains a loop built of quarks and is thus sensitive to the sum of colors. The result is $N_c = 3.01 \pm 0.05$.
  The thermodynamics of the early universe is influenced essentially by the number of species which are able to contribute to thermal equilibrium at a given temperature

(or equivalently: energy). The astrophysical data thus give bounds[8] to the number of colors and generations, and give support to the findings that only three types of colors and three families can exist.
In electron-positron colliders, one can produce the weak gauge boson Z^0 at a high rate and determine its mass, width and decay modes to a high degree of precision. As it turns out, the decay sensitively depends on the number of neutrino types, and the measurements are only compatible with three generations. Since, as will be explained later, in the weak sector of the Standard Model all leptons and quarks come in families, the Z^0-results indicate that no further fermions remain to be discovered.

- Non-occurrence of exotic states
The original quark model was not able to explain why nature does not exhibit particles with non-integer baryon number, e.g. bound states of two quarks. These would belong to representations of $\mathbf{3} \otimes \mathbf{3} = \overline{\mathbf{3}} \oplus \mathbf{6}$. But, according to QCD, only states which are invariant under the color group can be formed.
- Asymptotic freedom
In the perturbative analysis of QCD it was found that the strength of strong interaction goes to zero for smaller and smaller distances, or, higher and higher energies; more about this in Subsect. 6.4.5 below. Thus the constituent quarks behave like free particles in the limit of high energies (D. Gross and F. Wilczek, D. Politzer). As a consequence, the strong interaction processes can be treated by perturbation techniques for high energies.
- Confinement
On the other hand the coupling constant g_S increases with increasing distances. This has been supported by lattice calculations (discretized QCD). However, so far there is no strong proof for this feature[9]. Thus for low energies the quarks are confined and only appear as color-neutral hadronic bound states or "jets". Perturbative techniques are not applicable in this range. "Confinement" is, so to speak, a phase in QCD and illustrates that although gluons are massless the range of strong interaction is not infinite.

6.3 Weak and Electromagnetic Interaction

We saw in the previous section that strong interactions are surprisingly simple and are elegantly described by the QCD Lagrangian (6.9). It is completely analog to QED if the Abelian $\mathbf{U}(\mathbf{1})$ gauge group, the massless photon and the Dirac field for the electron-positron pair are replaced by a non-Abelian $\mathbf{SU}(\mathbf{3})$ gauge group with eight massless gluons and the Dirac fields for three generations of colored-anticolored quarks. Both theories, QED and QCD, are by construction locally invariant with respect to their gauge groups.

---

[8] These were known even prior to bounds from experiments in particle physics.

[9] If you succeed in delivering a proof, this would earn you a 1 Million Dollar Clay Prize.

The strong interactions and QCD is now understood as an arena in which only quarks act as players. In electroweak interactions electrons e^- and (electron)-neutrinos ν_e together with their antiparticles and their relatives from the other generations, the μ, the τ and their associated neutrinos ν_μ, ν_τ (again with their antiparticles) enter the scene. Three types of electroweak processes are observed: (i) processes involving only leptons such as the decay of a muon: $\mu^- \to e^- \bar{\nu}_e \nu_\mu$ (ii) processes involving only hadrons, such as the decay $\Lambda \to p\pi^-$ (iii) semi-leptonic processes like β decay $n \to pe^- \bar{\nu}_e$.

The historically very first weak interaction process to be observed was β decay, discovered in the early 20th century. First explanations in the prevailing models of atomic and nuclear physics interpreted this as a decay by which an atom changed its atomic number by ejecting an electron: $(Z, A) \to (Z + 1, A) + e^-$. Irritatingly this picture violates the conservation of energy-momentum. This led W. Pauli to postulate the existence of a light neutral particle[10] which he called "neutron". It was renamed neutrino by E. Fermi after the discovery of the "real" neutron. A first model (1934) for β decay is due to Fermi. With more and more experimental data and especially after the discovery of parity violation, weak interactions were described by the so called $(V\text{-}A)$ model, from which culminated the work on current algebra in the fifties and sixties.

Since weak interactions are a finite-range phenomena one would expect on contemplation of the Yukawa potential

$$V_Y(r) = -g_W^2 \frac{e^{Mr}}{r}$$

that they are mediated by fields with mass M. We have seen in the Sect. 5.3.4 that a mass for gauge particles is not compatible with local gauge symmetry–and this seems to obstruct a formulation of weak interactions within the frame of a Yang-Mills theory. As a matter of fact, historically the progress in forming a theory of weak interactions was retarded for a decade. Only in conjunction with the concept of spontaneous symmetry breaking it was possible to stay in (or rather "to return to") the framework of Yang-Mills theory. In the late sixties S. Weinberg and A. Salam–based upon previous ideas by S. Glashow–applied the Higgs mechanism to a $\mathbf{SU(2)} \times \mathbf{U(1)}$ gauge theory. This resulted in today's model of the electroweak interaction[11].

---

[10] Note that an apparently non-conservation of hitherto established conserved quantities related to basic symmetries of nature initiates the postulation of previously unobserved particles.

[11] Glashow, Salam and Weinberg received the Nobel price in 1979, although the massive vector bosons predicted by their model were found only in 1983 at CERN.

6.3.1 Fermi-Type Model of Weak Interactions

Formulated in modern language, Fermi postulated for the β decay a sum of interaction terms built from four fermion operators[12] as

$$\mathcal{L}_F = -\frac{G_F}{\sqrt{2}}\bar{p}(x)\gamma^\mu\psi_n\bar{\psi}_e\gamma_\mu\psi_n + h.c,$$

where G_F is a coupling constant determined by experiment to be $G_F/(\hbar c)^3 = 1.166 \times 10^{-5} GeV^{-2}$. In 1936 G. Gamow and E. Teller enlarged Fermi's *ansatz* and opted for a more general form of the four-point interaction with

$$\mathcal{L}_{GT} = -\frac{G_F}{\sqrt{2}}\sum_{AB}\Omega_{AB}\left(\bar{\psi}_p\Gamma^A\psi_n\right)\left(\bar{\psi}_e\Gamma^B\psi_\nu\right) + h.c,$$

where the summation ranges over those $\Gamma^A = \{1, \gamma^\mu, \gamma^5, \gamma^\mu\gamma^5, \gamma^{\mu\nu}\}$ for which the products lead to parity-conserving Lorentz invariant bilinear expressions of the Dirac spinors. After the discovery of parity violation in 1956 and with further experimental data, R. Marshak and C. G. Sudarshan, and then R. Feynman and M. Gell-Mann, as well as J. J. Sakurai proposed the $(V\text{-}A)$ form

$$\mathcal{L}^\beta_{\text{V-A}} = -\frac{G_F}{\sqrt{2}}\left(\bar{\psi}_p\gamma^\mu(1-g\gamma^5)\psi_n\right)\left(\bar{\psi}_e\gamma_\mu(1-\gamma^5)\psi_\nu\right) + h.c.$$

It is called V(ector)-A(xialvector) because in the lepton sector the combination $(\gamma^\mu - \gamma^\mu\gamma^5)$ appears. In the proton-neutron term, a constant $g \simeq 1.26$ occurs, which, as we know today, arises because the hadrons are composed of quarks. As a matter of fact also the weak interaction of quarks obey the $(V\text{-}A)$ form (with $g = 1$), if the Glashow-Salam-Weinberg model is approximated by the Fermi model. It is the $(1-\gamma^5)$ term which gives rise to parity violation.

Even though the Fermi model makes good predictions for weak processes, it is theoretically not acceptable. The model is not renormalizable since the coupling constant G_F has dimension $(\text{mass})^{-2}$; and indeed the criterion (5.23) yields for the Fermi interaction term $\Delta = 4\text{-}0\text{-}4(\frac{3}{2}) = -2$.

The Lagrangian term $\mathcal{L}_{\text{V-A}}$ was found for all weak interaction processes to have the structure

$$\mathcal{L}_{\text{V-A}} = -\frac{G_F}{\sqrt{2}}J_1^\mu J_{2\mu}$$

where the currents are of the form $J_k^\mu = \bar{\psi}_1(O_{12})_k^\mu\psi_2$. Therefore the investigation of weak currents and their algebra became prominent in the 1960's.

[12] If you read Fermi's original article, you can appreciate how progress in physics is connected with the development of adequate mathematical notions.

6.3.2 Current Algebra

Generically one can decompose the current-current Lagrangian as

$$\mathcal{L}_{jj} = -\frac{G_F}{\sqrt{2}}\Big(j_\mu^{(+)}(x)j^{(-)\mu}(x) + j_\mu^{(n)}(x)j^{(n)\mu}(x)\Big).$$

with the charged currents $j_\mu^{(\pm)}(x)$ (together with $j_\mu^{(-)} = (j_\mu^{(+)})^\dagger$) and the neutral current $j^{(n)\mu}(x)$. Both are composed of a leptonic and a hadronic part:

$$j_\mu^{(\pm)}(x) = l_\mu^{(\pm)}(x) + h_\mu^{(\pm)}(x) \tag{6.12a}$$

$$j_\mu^{(n)}(x) = l_\mu^{(n)}(x) + h_\mu^{(n)}(x). \tag{6.12b}$$

(A) Leptonic Currents

Three lepton families are known to exist. To each of them is associated a separately-conserved lepton number. It is remarkable that from all experiments so far, all lepton currents do have the same structure which is

$$l_\mu^{(\text{em})}(x) = -\,\overline{e}(x)\gamma_\mu e(x) + (e \to \mu, \tau) \tag{6.13a}$$

$$l_\mu^{+}(x) = \overline{\nu}_e(x)\gamma_\mu^L e(x) + (e \to \mu, \tau) \tag{6.13b}$$

$$l_\mu^{-}(x) = \overline{e}(x)\gamma_\mu^L \nu_e(x) + (e \to \mu, \tau) \tag{6.13c}$$

where

$$\gamma_\mu^L := \gamma_\mu \frac{1-\gamma_5}{2} \qquad\qquad \gamma_\mu^R = \gamma_\mu \frac{1+\gamma_5}{2}.$$

The neutral leptonic current was found by experiments to have the structure

$$l_\mu^{(n)}(x) = l_\mu^{(0)}(x) - sin^2\theta_W \;\; l_\mu^{(\text{em})}(x) \tag{6.14}$$

that is a linear combination of the electromagnetic current $l_\mu^{(em)}$ and the current

$$l_\mu^{(0)}(x) = \frac{1}{2}\Big[\overline{\nu}_e(x)\gamma_\mu^L \nu_e(x) - \overline{e}(x)\gamma_\mu^L e(x)\Big] + (e \to \mu, \tau). \tag{6.15}$$

The mixing ratio is characterized by the *Weinberg angle* θ_W. We see from (6.13a) that the electromagnetic current is a pure vector(V) current, whereas the currents $l_\mu^{(\pm)}$ and $l_\mu^{(0)}$ exhibit a $(V\text{-}A)$ or left-handed structure. This reflects the experimental fact that for charged currents the parity violation is maximal, whereas for neutral currents the degree of parity violation depends on the charge.

The charges associated to the currents $l_\mu^{(\pm)}$ and $l_\mu^{(0)}$, defined by

$$I_l^\pm(t) := \int d^3x \ l_0^{(\pm)}(t,\vec{x}), \qquad I_l^0(t) := \int d^3x \ l_0^{(0)}(t,\vec{x}) \tag{6.16}$$

have commutators

$$[I_l^+, I_l^-] = 2I_l^0, \qquad [I_l^0, I_l^\pm] = \pm I_l^\pm.$$

This is the Lie algebra of the $\mathbf{SU(2)}$ group, in this context called the weak isospin group $\mathbf{SU_I(2)}$; in the literature sometimes also $\mathbf{SU_L(2)}$ or $\mathbf{SU_W(2)}$.

Note that in all the currents (6.13, 6.15), the neutrino fields always make their appearance in the combination $\gamma_\mu^L \nu$ which is the same as $\gamma_\mu \nu_L$. There are no right-handed neutrinos in the currents. At the same time this is compatible with the terms $\bar{\nu}\gamma_\mu^L = \bar{\nu}_R\gamma^\mu$ containing no left-handed anti-neutrinos. Weak interactions are chiral: they favor left-handed neutrinos and right-handed anti-neutrinos.

Given the structure of the currents and the chirality, it makes sense to regard the left-handed leptons as isospin doublets

$$L_l := \left\{ \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \right\}, \tag{6.17}$$

and the right-handed electrons as isospin singlets

$$R_l := \{(e^-)_R, (\mu^-)_R, (\tau^-)_R\}. \tag{6.18}$$

With this assignment the neutrinos have $I^0 = +\frac{1}{2}$, the left-handed electron family has $I^0 = -\frac{1}{2}$, and the right-handed electron family $I^0 = 0$. The weak hypercharge Y, defined by

$$\frac{Y}{2} = Q - I^0, \tag{6.19}$$

commutes with all components of the weak isospin. Since this is a $\mathbf{U(1)}$ symmetry the entire symmetry amounts to $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$, and the assignment of hypercharge to the isospin multiplets is

$$Y(L_l) = -1, \qquad Y(R_l) = -2. \tag{6.20}$$

(B) Hadronic Currents

Astonishingly enough–and this can not be explained by or within the Standard Model–the structure of hadronic currents is very similar to that of the leptonic currents. The hadronic currents are built as bilinears of the Dirac spinors for their constituents, namely the quarks:

$$h_\mu^{(em)}(x) = \sum_{q,c} \bar{q}_c(x)\ \gamma_\mu\ Q\, q_c(x). \tag{6.21}$$

The charge matrix is diagonal, that is $Q = \mathrm{diag}\{\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{2}{3}, \ldots\}$. Experiments reveal that the $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ symmetry also holds for hadronic weak processes, so that the generators (6.16) of the weak isospin can simply be extended to

$$\begin{aligned} I^\pm(t) &:= \int d^3x\ \ (l_0^{(\pm)}(t,\vec{x}) + h_0^{(\pm)}(t,\vec{x})) \\ I^0(t) &:= \int d^3x\ \ (l_0^{(0)}(t,\vec{x}) + h_0^{(0)}(t,\vec{x})). \end{aligned}$$

The hadronic charged current has the form

$$h_\mu^{(+)}(x) = \big(\bar{u}(x), \bar{c}(x), \bar{t}(x)\big)\gamma_\mu^L U_{CKM} \begin{pmatrix} d(x) \\ s(x) \\ b(x) \end{pmatrix} \qquad h_\mu^{(-)} = h_\mu^{(+)\dagger} \tag{6.22}$$

with the so-called Cabbibo-Kobayashi-Maskawa matrix U_{CKM}[13]. Because of separate lepton-number conservation in case of leptons the corresponding matrix is the unit matrix, as seen from (6.13).

The analogy with the leptonic case suggests that one can arrange the left-handed quarks into doublets

$$L_q := \left\{ \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L \right\} \qquad Y(L_q) = \ 1/3 \tag{6.23}$$

and the right-handed ones into singlets,

$$R_q := \{u_R, c_R, t_R\} \qquad Y(R_q) = 4/3 \tag{6.24a}$$

$$R_Q := \{d'_R, s'_R, b'_R\} \qquad Y(R_Q) = -\,2/3. \tag{6.24b}$$

Here

$$\begin{pmatrix} d'(x) \\ s'(x) \\ b'(x) \end{pmatrix} = U_{CKM} \begin{pmatrix} d(x) \\ s(x) \\ b(x) \end{pmatrix}$$

are states rotated within the flavor space, namely eigenstates with respect to isospin (the original "unprimed" quarks are mass eigenstates). Furthermore

$$2\, h_\mu^{(0)}(x) = \big(\bar{u}, \bar{c}, \bar{t}\big)\, \gamma_\mu^L D \begin{pmatrix} u \\ c \\ t \end{pmatrix} - \big(\bar{d}, \bar{s}, \bar{b}\big)\, \gamma_\mu^L D^\dagger \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{6.25}$$

[13] Generically it is a $(\frac{N_f}{2} \times \frac{N_f}{2})$−matrix. However, as stated above, the discovery of more than six flavors would spoil both results in particle physics and in cosmology.

Table 6.1 Weak quantum numbers

quark	I	I_3	Q	Y	lepton	I	$\mathbf{I}_3$	Q	Y
u_L	$1/2$	$1/2$	$2/3$	$1/3$	ν_e	$1/2$	$1/2$	0	-1
d_L	$1/2$	$-1/2$	$-1/3$	$1/3$	e_L^-	$1/2$	-1	-1	-1
u_R	0	0	$2/3$	$4/3$					
d_R	0	0	$-1/3$	$-2/3$	e_R^-	0	0	-1	-2

where D is the matrix

$$D = U_{CKM} U_{CKM}^{\dagger}.$$

From the fact that no flavor changing neutral currents have been observed, this matrix must be identical to the unit matrix. This is related to the so called GIM mechanism named after S. Glashow, J. Iliopoulos, L. Maiani [226]. These authors postulated in 1970 the existence of a charm quark as a partner of the strange quark in order to be able to express the current in the form (6.25)–read of course as a two-component equation since the t/b-pair came later. With D as the identity matrix, the hadronic current $h_\mu^{(0)}$ matches in structure its leptonic counterpart $l_\mu^{(0)}$, given by (6.15).

As a 3×3 unitary matrix U_{CKM} depends on nine real parameter. It can be parametrized in terms of three angles and six phases. For the three generations of quarks and anti-quarks $(2 \times 3 \text{ -} 1)$ relative phases can be absorbed into redefined quark fields. The three angles are called mixing angles, by some also Cabbibo angles, since one of them was already introduced in 1963 by N. Cabibbo: In today's quark picture this corresponds to $N_f = 4$ and a unitary matrix

$$\begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix}.$$

The angle θ_C is determined from the weak processes with $|\Delta S| = 1, |\Delta Q| = 0$ as $\theta_C \approx 13.04°$. The remaining phase in the 3×3-matrix U_{CKM} is directly related to the CP-violation of weak interaction processes. Since in the context of this model, CP-violation could not be explained in case of only two families of quarks, M. Kobayashi and T. Maskawa demanded for (at least one) further generation of quarks[14]. So when in 1977 the bottom quark was indeed discovered, no particle physicist doubted the existence of its partner, the top quark, found nearly two decades later[15].

As in the case of leptons, the neutral hadronic current $h_\mu^{(n)}(x)$ is a mixture of an electromagnetic current $h_\mu^{(em)}(x)$ and of $h_\mu^{(0)}(x)$. And again the mixing angle is the Weinberg angle, in an expression quite similar to (6.14).

In order not to loose track and for later purposes Table 6.1 shows the assignment of quantum numbers–for the first generation of quarks only.

---

[14] They received the Nobel prize in 2008 "for the discovery of the origin of the broken symmetry which predicts the existence of at least three families of quarks in nature".

[15] Here we observe anew the belief and trust of physicists in symmetries.

6.3.3 Glashow-Salam-Weinberg Model

The model named for S. Glashow, A. Salam and S. Weinberg[16] is a Yang-Mills theory of the group $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ coupled to massless leptons and quarks, in which the gauge symmetry is spontaneously broken by the Higgs mechanism. The GSW model is minimal in the sense that only as many scalar fields are introduced as are needed for the breaking

$$\mathbf{SU_I(2)} \times \mathbf{U_Y(1)} \;\curvearrowright\; \mathbf{U_Q(1)}.$$

By this breaking, three of the four gauge bosons and (by a further appropriate term in the Lagrangian) the fermions acquire a mass; one gauge boson remains massless and can be identified with the photon.
The GSW-Lagrange density is a sum of three terms

$$\mathcal{L}_{GSW} = \mathcal{L}_F + \mathcal{L}_G + \mathcal{L}_S, \tag{6.26}$$

where the indices F, G, S stand for the fermions, the gauge fields, and the scalars responsible for spontaneous symmetry breaking.

Fermion Sector

The starting point is a theory of massless leptons and quarks which is globally invariant with respect to the symmetry group $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$. Each of the fermions contributes to the Lagrangian with a Dirac term

$$i\overline{\psi}\gamma^\mu\partial_\mu\psi = i\overline{\psi}_L\gamma^\mu\partial_\mu\psi_L + i\overline{\psi}_R\gamma^\mu\partial_\mu\psi_R.$$

In order to achieve the global $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ invariance, we furthermore need to introduce the multiplets (6.17, 6.18) for the leptons and (6.23, 6.24) for the quarks. With the additional obvious notation

$$L_f = \{L_l, L_q\} \qquad R_f = \{R_l, R_q, R_Q\}$$

the fermion Lagrangian before gauging (indicated by the $*$) is the sum of Dirac terms

$$\mathcal{L}_F^* = i\overline{L}_f\gamma^\mu\partial_\mu L_f + i\overline{R}_f\gamma^\mu\partial_\mu R_f. \tag{6.27}$$

Local Symmetry and Gauging

According to the Yang-Mills recipe, the global symmetry of (6.27) can be turned into a local symmetry by introducing gauge fields and covariant derivatives.

[16] Nobel prize conferred in 1979: "For their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, including, inter alia, the prediction of the weak neutral current".

Since the group $\mathbf{SU(2)} \times \mathbf{U(1)}$ is $(3+1)$-dimensional, the gauging procedure needs four gauge fields. Denote the gauge fields addressing $\mathbf{SU_I(2)}$ and the three components of the weak isospin I^a as $W^a_\mu(x)$ $(a = 1, 2, 3)$, and the gauge field to $\mathbf{U_Y(1)}$ and the weak hypercharge Y by $B_\mu(x)$.

The covariant derivatives are

$$D_\mu = \partial_\mu + i g_I \hat{I}^a W^a_\mu + i g_Y \hat{Y} B_\mu, \tag{6.28}$$

where the $\hat{I}^a$ are representation matrices of $\mathbf{SU_I(2)}$ generators and $\hat{Y}$ is a representation of the generator of $\mathbf{U_Y(1)}$:

$$\hat{I}^a = \begin{pmatrix} \frac{1}{2}\sigma^a & 0 \\ 0 & 0 \end{pmatrix} \qquad \hat{Y} = \begin{pmatrix} y_L 1 & 0 \\ 0 & y_R \end{pmatrix}.$$

Because of the product structure of the gauge group the covariant derivative (6.28) has two different coupling constants $g_I = g_W$ and $g_Y = g_B$. Given the assignment of isospin and hypercharge for the various multiplets (see (6.20, 6.23, 6.24)) the covariant derivatives of the lepton and quark fields are explicitly

$$\begin{aligned}
D_\mu L_l &= \left[\partial_\mu + \frac{1}{2}(i g_I \vec{\sigma} \cdot \vec{W}_\mu) - i g_Y B_\mu\right] L_l \\
D_\mu R_l &= \left[\partial_\mu - 2 i g_Y B_\mu\right] R_l \\
D_\mu L_q &= \left[\partial_\mu + \frac{1}{2} i g_I \vec{\sigma} \cdot \vec{W}_\mu - \frac{1}{3} i g_Y B_\mu\right] L_q \\
D_\mu R_q &= \left[\partial_\mu + \frac{4}{3} i g_Y B_\mu\right] R_q \\
D_\mu R_Q &= \left[\partial_\mu - \frac{2}{3} i g_Y B_\mu\right] R_Q .
\end{aligned}$$

The Lagrange density contribution $\mathcal{L}_F$ is thus

$$\mathcal{L}_F = i\overline{L}_l \gamma^\mu D_\mu L_l + i\overline{R}_l \gamma^\mu D_\mu R_l + i\overline{L}_q \gamma^\mu D_\mu L_q + i\overline{R}_q \gamma^\mu D_\mu R_q + i\overline{R}_Q \gamma^\mu D_\mu R_Q . \tag{6.29}$$

The dynamics of the gauge field is determined by the field strengths

$$\begin{aligned}
W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g_I\, \epsilon^{abc} W^b_\mu W^c_\nu \\
B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu
\end{aligned}$$

and their contribution to the Lagrange density (6.26) in the form

$$\mathcal{L}_G = -\frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \tag{6.30}$$

Spontaneous Symmetry Breaking

For low energies the electroweak theory must be broken to the electromagnetic $\mathbf{U_Q(1)}$ theory. Therefore, a Higgs mechanism must operate in such a way that after symmetry breaking one massless gauge boson (the photon) remains, and three others acquire a mass. To this end, introduce four fields arranged as an isospin doublet with two complex scalar fields $\Phi = (\Phi^q, \Phi^{q'})$ with appropriate charges q, q'. These charges are to be chosen in such a way that after spontaneous symmetry breaking, the electromagnetic $\mathbf{U_Q(1)}$ symmetry survives. Therefore the lower component of Φ has to be neutral: $q' = 0$. Next, because of $Y = 2(Q\text{-}I^0)$, and since the lower component in the doublet has $I^0 = -\frac{1}{2}$, the hypercharge of the multiplet is $Y = 1$. This in turn requires the charge of the upper component to be $q = 1$. With this assignment of weak isospin and weak hypercharge the covariant derivative of the scalar field[17]

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{6.31}$$

becomes according to (6.28)

$$D_\mu \Phi = \left[\partial_\mu + (i/2) g_I \vec{\sigma} \cdot \vec{W}_\mu) + i g_Y B_\mu\right] \Phi.$$

The scalars interact with all other fields. Thus the Lagrangian has two parts, namely

$$\mathcal{L}_S = D_\mu \Phi^\dagger D^\mu \Phi - \frac{\lambda}{2}[\Phi^\dagger \Phi - v^2]^2 + \mathcal{L}_Y(\Phi, L_f, R_f).$$

The term $\mathcal{L}_Y$ describes the interaction of the scalar field with the quark and lepton fields. It is named the Yukawa interaction term, because it was originally introduced by H. Yukawa to describe the strong nuclear force between nucleons, mediated by pions. The most general Yukawa term which is allowed by the local $\mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ symmetry is for the leptons and the quarks given by

$$\mathcal{L}_Y^{\text{lept}} = -\sum_l g_l \bar{L}_l \Phi R_l + h.c, \tag{6.32a}$$

$$\mathcal{L}_Y^{\text{quark}} = -\sum_q g_q \, [\bar{L}_q \Phi R_Q + \bar{L}_Q \tilde{\Phi}^\dagger R_q] + h.c. \tag{6.32b}$$

Here $L_Q = U^\dagger_{CKM} L_q$ and $\tilde{\Phi} = i\sigma^2 \Phi$ (the charge conjugate to Φ).

Starting from the classical vacuum

$$\langle \Phi \rangle_0 = \frac{v}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

---

[17] I try to be consistent in this book in distinguishing the symmetry-breaking scalar field from the Higgs boson which survives the symmetry.

introduce four new fields $G^a(x)$ and $H(x)$ related to $\Phi(x)$ by

$$\Phi(x) = \frac{1}{\sqrt{2}} e^{i\vec{\sigma}\cdot\vec{G}/v} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}.$$

In expressing the Lagrangian (6.26) in terms of these modified fields, it is found after a suitable gauge transformation that the fields Φ^+ and G^a disappears from the theory, such that only one neutral scalar boson–the Higgs boson $H(x)$–remains.

Immediate Consequences of the GSW Model

The Lagrangian arising after spontaneous symmetry breaking can be split into three parts: a kinetic term, an interaction term, and a further term from which the masses can be read off. The kinetic term has the "usual" quadratic field strength part for the gauge field, the Dirac part for the fermions and a part $\partial_\mu \partial^\mu H$ for the Higgs.

Masses

- The Higgs boson $H(x)$ has the mass $M_H = v\sqrt{2\lambda}$.
- The vector fields that arise after the spontaneous symmetry breaking are named by the same symbols as before the symmetry breaking. Furthermore, defining the linear combinations

$$W_\mu^{(\pm)} := \frac{1}{\sqrt{2}}[W_\mu^1 \pm i W_\mu^2].$$

the mass term becomes

$$v^2 g_I^2 W_\mu^{(+)} W^{(-)\mu} + \frac{1}{2} v^2 [g_I W_\mu^3 + g_Y B_\mu]^2. \tag{6.33}$$

The charged vector bosons $W^\pm$ acquire the mass

$$M_{W^+} = M_{W^-} = \frac{1}{\sqrt{2}} g_I v = \frac{M_H}{2\sqrt{\lambda}}.$$

The masses for the neutral vector bosons W^0 and B are not directly calculable from (6.33). In order to find the proper mass eigenstates we must diagonalize the mass matrix

$$(M) = \frac{v^2}{2} \begin{pmatrix} g_I^2 & g_I g_Y \\ g_I g_Y & g_Y^2 \end{pmatrix}.$$

Since the determinant of (M) vanishes, one of the eigenvalues must be zero. This is the anticipated result of one massless gauge boson remaining after symmetry breaking. The sum of eigenvalues is $\frac{v^2}{2}(g_I^2 + g_Y^2)^2$. Denoting the mass eigenstates by A and Z, we obtain the masses of these neutral vector bosons as

$$M_A = 0 \qquad M_Z = \frac{v}{\sqrt{2}}\sqrt{g_I^2 + g_Y^2}.$$

In diagonalizing the matrix (M), introduce linear combinations of the neutral vector bosons W_μ^3 and B_μ as

$$Z_\mu := \quad \cos\theta_W \; W_\mu^3 + \sin\theta_W \; B_\mu$$

$$A_\mu := \; -\sin\theta_W \; W_\mu^3 + \cos\theta_W \; B_\mu.$$

The matrix becomes diagonal by fixing the angle θ_W. (The justification for identifying the angle with the Weinberg angle is given below.)

$$\tan\theta_W = \frac{g_Y}{g_I} \qquad\qquad \text{or} \qquad\qquad \sin^2\theta_W = \frac{g_Y^2}{g_I^2 + g_Y^2}.$$

Indeed the second term in (6.33) becomes

$$\frac{1}{2}v^2(g_I W_\mu^3 + g_Y B_\mu)^2 \;\Rightarrow\; \frac{1}{2}v^2(g_I\cos\theta_W + g_Y\sin\theta_W)Z^2 = \frac{1}{2}v^2\frac{g_I^2}{\cos^2\theta_W}Z^2.$$

There is no mass term for the field A_μ, and the Z mass can be expressed by the Weinberg angle as

$$M_Z = \frac{M_W}{|\cos\theta_W|}.$$

- The leptons acquire a mass through the Yukawa terms (6.32). If the scalar field is taken at its ground state value, we obtain

$$\mathcal{L}_Y = -\frac{v}{\sqrt{2}}[g_e\bar{e}_L e_R + g_d\,\bar{d}_L d_R + g_u\bar{u}_L U_R] + h.c.] + (e \to \mu,\tau).$$

Thus the fermion fields couple proportionally to their masses to the Higgs field (with a universal factor $v/\sqrt{2}$ of proportionality).

Interactions

The identification of the A_μ with the electromagnetic field becomes justified by investigating the Lagrangian part containing the interaction terms which couple the fermions to the vector fields:

$$g_I\sin\theta_W\, A^\mu j_\mu^{(em)} + \frac{g_I}{\sqrt{2}}\left(W^{(+)\mu}j_\mu^{(+)} + W^{(-)\mu}j_\mu^{(-)}\right) + \frac{g_I}{\cos\theta_W}Z^\mu j_\mu^{(0)}$$

with the currents $j_\mu^{(\alpha)} = l_\mu^{(\alpha)} + h_\mu^{(\alpha)}$ being the sum of the leptonic and hadronic currents (6.13, 6.14, 6.21, 6.22, 6.25). In order to recover the electromagnetic interaction of leptons from the first term we need to identify

$$e = g_I \sin\theta_W = g_Y \cos\theta_W \qquad \text{or} \qquad e^2 = \frac{g_I^2 g_Y^2}{g_I^2 + g_Y^2}.$$

6.3.4 Theoretical Consistency and Experimental Support

For low energies (that is energies small compared to the mass of the weakons) the GSW model must approximate the Fermi and the $(V\text{-}A)$ model which are formulated with local four-point interactions. The propagator of the exchange of a $W^\pm$ boson between two fermions

$$\frac{g_{\mu\nu} - k_\mu k_\nu / M_W^2}{M_W^2 - q^2} \xrightarrow{q^2 \ll M_W^2} \frac{g_{\mu\nu}}{M_W^2}$$

indeed reduces for small momentum transfers q^2 to a contact interaction. Matching this to the effective Hamiltonian

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} J^\mu J_\mu$$

with the currents given by (6.12) yields the relation

$$\frac{G_F}{\sqrt{2}} = \frac{g_I^2}{8M_W^2},$$

and allows us to determine the value of v from

$$\frac{g_I^2}{8M_W^2} = \frac{1}{4v^2} \cot^2\theta_W$$

as $v \approx 250\,\text{GeV}$. With the experimentally known Weinberg angle $(\sin^2\theta_W \approx 0.23)$ one obtains an estimate for the mass of the weakons

$$M_W \approx 79\,\text{GeV} \qquad M_Z \approx 90\,\text{GeV},$$

which is surprisingly close to the values found in experiment: $M_W = 80\,\text{GeV}$, $M_Z = 91\,\text{GeV}$. The mass of the Higgs scalar is not determined by the GSW model, since it depends on the experimentally inaccessible parameter λ. However there are independent experimental and theoretical "educated guesses" that the Higgs mass should be in the range 120-150 GeV (if nature follows the GSW model with only one Higgs field); more details in [371]. Indeed, in July 2012, CERN announced the discovery of a scalar boson with mass 126 GeV. Further data collections and analyses give strong hints that this boson has indeed properties of the minimal Higgs sector in the standard model.

6.4 Paralipomena on the Standard Model

In Sect. 6.2.2 we reflected on the symmetries of QCD. Which of these symmetries survive in the full SM? Of course P and C symmetry are no longer valid. Depending on your taste and attitude you can attribute this either to the manner in which the electroweak theory was built according to the structure of the leptonic and hadronic currents found from experiments, or you can assign this to the chiral characteristic of nature. In QCD, each flavor is separately conserved. This is no longer true for electroweak processes since by processes involving the weakons, flavors can be changed. Nevertheless the baryon number (being essentially the sum of the flavornesses according to (6.5)) remains conserved because the SM Lagrangian is invariant with respect to

$$L_q \to e^{i\alpha} L_q, \qquad R_q \to e^{i\alpha} R_q, \qquad R_Q \to e^{-i\alpha} R_Q.$$

And furthermore, the three separate lepton numbers are conserved due to

$$L_f \to e^{i\beta} L_f, \qquad R_f \to e^{-i\beta} R_f.$$

These symmetries need not be imposed, but are a consequence of the structure of the SM Lagrangian. In turn, this Lagrangian receives its structure only from the requirement of gauge invariance and renormalizability.

6.4.1 Limits of the Standard Model

In spite of its astounding success–myriads of data taken in particle accelerators are within experimental and theoretical limits in perfect agreement with the model, and even the Higgs particle seems to have been detected after searching for half a century–yet the Standard Model does not please the aesthetic aspirations of those believing in simple and appealing roots of fundamental physics. The Standard Model is in fact to good to be true.

One of the reasons is the number of parameters which must be plugged into the model from the outside. There are 19 of them, which we can list for instance as

- 3 coupling constants for the gauge group $\mathbf{SU_C(3)} \times \mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$. Although two of them become related by the Weinberg angle, the parameter $\sin^2\theta_W$ is also not determined within the model.
- 9 mass parameters for six quarks and the three charged leptons. Off course you could just as well count nine Yukawa couplings instead.
- 4 parameters in the CKM matrix: three quark mixing angles and one CP-violating phase.
- 2 Higgs parameters: self coupling (λ) and vacuum expectation value (v).

- 1 θ-parameter from a further possible Lagrangian term for the kinetic part of the gauge bosons; see Subsect. 6.4.4.

This gets even worse, since we know today that neutrinos are not massless. This requires a modification and augmentation of the Standard Model. For instance, we must conceive that the neutrino partner in the left-handed isospin doublets are mixed states, as is the case for the quarks. In other words, a sort of a CKM matrix is needed also in the lepton sector. With three further mass parameters for the neutrinos and at least four parameters for this matrix, we would arrive at 26 parameters for the Standard Model. More about massive neutrinos below.

Another reason of reluctance accepting the Standard Model as ultimate fundamental physics is the fact that its gauge group is the product of three groups. This led early to attempts to find a larger symmetry group which spontaneously breaks down to $\mathbf{SU_C(3)} \times \mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$. Although so far none of the proposals could be established, more details of these "grand unification" attempts will be given in Sect. 8.1.

The Standard Model cannot explain in itself why there are three flavors, three colors, three families, three interactions etc. And it cannot explain why charge is quantized and why quarks carry their $1/3$ elementary charge (this at the very end would yield an explanation of why the proton charge has the exact magnitude–but opposite sign–compared to the electronic charge).

Although, technically speaking, the Higgs mechanism works and constitutes one part of the electroweak sector, many feel unsure. As a matter of fact the Higgs-field interactions seem entirely arbitrary. And why is there only one Higgs dublet? Can the Higgs mechanism be mimicked by other more pleasing procedures? Yes it can: I mentioned for instance Stückelberg fields in the concluding section of Chap. 5. and will sketch technicolor models in Sect. 8.4.1.

According to the Standard Model our world is dichotomic in two senses: (1) Although the lepton and quark interactions share many structural similarities, especially their charged weak currents, they do not exhibit any relations amongst each other. As we will see in Sect. 8.1, this discrepancy is resolved in grand unified theories in which the leptons and quarks are assembled within common multiplets. (2) Furthermore, the matter fields (fermions: leptons, quarks) and force fields (vector bosons: gluons, weakons, photon)–plus sort of a bastards (Higgs scalars) are treated completely differently in the model. A way to overcome this dichotomy is to be seen in supersymmetry, a topic dealt with in Sect. 8.3[18].

Another major problem is the fact that the Standard Model for the three particle interactions is conceptually incompatible with general relativity, the theory of the fourth interaction, namely gravity. Indeed, the main challenge of theoretical physics today is to find a theoretical frame in which all interactions are treated on the same footing. Superstrings are such a frame, but indeed present zillions of solutions. At least they constitute a proof of existence of a genuine unification; further details will be given in 8.4.2.

[18] Once again the assumption of further symmetries will improve our understanding of physics.

Although as a synonym for particle physics one often speaks of "high energy physics" it should be noted, that the Standard Model has been tested experimentally only at energies smaller than a few hundred GeV. It is very well conceivable that the model is just an effective field theory–valid at least up to that energy scale. The characteristic energy of the Standard Model is given by the mass of the W-boson $M_W \sim 100\,\text{GeV}$. The next scale coming into view is the Planck energy $M_{Planck} \sim 10^{19}$ GeV or maybe scales of the order $M_{GUT} \sim 10^{15}$ GeV–if nature in fact realizes grand unification. In any case, it is hard to believe that there is no new physics in the range between the present highest energies reachable by accelerators and the extremely high mass scales.

We should also bear in mind that according to the other standard model, namely the concordance model for cosmology, the SM for particles explains only roughly five percent of the matter in the universe. The rest, dark energy and dark matter, remains unexplained.

6.4.2 Massive Neutrinos

In the Standard Model described so far, neutrinos were assumed to be massless. This is built into the model by treating them as Weyl fermions. There is no inherent logic why neutrinos should be massless, compared for instance to the photon for which a mass would destroy the $\mathbf{U(1)}$ gauge invariance. And indeed, due to the observed phenomenon of neutrino oscillation neutrinos are now believed to have nonzero masses. Neutrino oscillation stands for the effect that the three flavor neutrinos ν_e, ν_μ, ν_τ (and their antiparticles) oscillate between the three flavor sates while they propagate through space. This can happen if the neutrino flavor eigenstates are not the same as the mass eigenstates. The idea is roughly as follows: Call $|\nu_\alpha\rangle$ the flavor eigenstates, and $|\nu_k\rangle$ the mass eigenstates. These are related by a 3×3 unitary matrix U

$$|\nu_\alpha\rangle = U_{\alpha k}|\nu_k\rangle.$$

called–after the inventors of this mechanism–the Pontecorvo-Maki-Nakagawa-Sakata matrix (also termed the lepton-mixing matrix). It is the analogue to the CKM matrix describing mixing of quark flavor and quark mass eigenstates. And similarly to the CKM matrix, the matrix $U_{\alpha i}$ can be expressed in terms of three angles and some phases. The number of phases depends on whether the neutrinos are Dirac or Majorana fields. In the former case the counting of phases is the same as for the CKM matrix: one phase in case of three generations. In case of Majorana neutrinos only (3-1) relative phases can be absorbed in the neutrino states, and thus three phases are physically relevant. Each of the mass eigenstates propagates in time according to

$$|\nu_k(t)\rangle = e^{-iE_k t}|\nu_k(0)\rangle.$$

Due to their small masses[19] the energy for each species can be approximated by

$$E_k = \sqrt{p_k^2 + m_k^2} \simeq E + \frac{m_k^2}{2E}.$$

Introducing the length of traveling $L \simeq t$ we thus obtain

$$|\nu_k(L)\rangle = e^{-i\frac{m_k^2\,L}{2E}}|\nu_k(0)\rangle$$

(where the common phase $\exp\{-iEL\}$ has been dropped). This expression shows how the speed of propagation of a mass eigenstate depends on the mass. Since the flavor eigenstates are linear combinations of mass eigenstates, the differences in speed lead to interference between the flavor components. This makes it possible that a neutrino created with a given flavor changes its flavor during its propagation. The probability to observe a neutrino $|\nu_\beta\rangle$ that originally had a flavor α after it traveled a distance D is given by

$$P(\alpha \to \beta) = \langle \nu_\beta | \nu_\alpha(D)\rangle^2 = \left| \sum_k U_{\alpha k}^* U_{\beta k} e^{-i\frac{m_k^2 D}{2E}} \right|^2.$$

Due to the small masses, the distance D can be assumed to be the same for all neutrino types. Since the explicit calculation of this probability is a little laborious for the 3×3 case, the essentials will be made clear for the example with two neutrino generations. Let

$$|\nu_e\rangle = \cos\theta_{lm}|\nu_1\rangle + \sin\theta_{lm}|\nu_2\rangle, \qquad |\nu_\mu\rangle = -\sin\theta_{lm}|\nu_1\rangle + \cos\theta_{lm}|\nu_2\rangle.$$

Here it was assumed that we are dealing with Dirac neutrinos. For Majorana neutrinos, two additional phases appear in the linear combinations of the mass eigenstates. This assumption is not really relevant, since as will be seen later, the phases drop out of the calculation of transition probabilities. After a time t has elapsed the original pure $|\nu_e\rangle$ state is

$$\begin{aligned} |\nu(t)\rangle &= \cos\theta_{lm} e^{-iE_1 t}|\nu_1\rangle + \sin\theta_{lm} e^{-iE_2 t}|\nu_2\rangle \\ &= \left[\cos^2\theta_{lm} e^{-iE_1 t} + \sin^2\theta_{lm} e^{-iE_2 t}\right]|\nu_e\rangle \\ &\quad + \left[\sin\theta_{lm}\cos\theta_{lm}(-e^{-iE_1 t} + e^{-iE_2 t})\right]|\nu_\mu\rangle \end{aligned}$$

from which one calculates the probability of observing a $|\nu_e\rangle$ to be

$$P(\nu_e \to \nu_e, t) = \langle \nu_e | \nu(t)\rangle^2 = 1 - \sin^2(2\theta_{lm})\sin^2\left[\frac{(E_1 - E_2)t}{2}\right],$$

---

[19] There are only upper bounds, the most stringent one coming from the results on the microwave background radiation due to which the sum of the masses of the three neutrinos is smaller than roughly 1 eV.

and the probability of observing a $|\nu_\mu\rangle$ is $P(\nu_e \to \nu_\mu, t) = 1 - P(\nu_e \to \nu_e, t)$. Now

$$E_1 - E_2 = \frac{E_1^2 - E_2^2}{E_1 + E_2} \simeq \frac{m_1^2 - m_2^2}{2E} = \frac{\Delta m^2}{2E}$$

from which

$$P(\nu_e \to \nu_e, t) = 1 - \sin^2(2\theta_{lm}) \sin^2\left(\frac{\pi D}{L}\right) \qquad \text{with} \qquad L := \frac{4\pi E}{\Delta m^2}.$$

This probability is different from 1 if the lepton mixing angle θ_{lm} is nonzero, and if $\Delta m^2 \neq 0$. As became obvious from the previous calculation, neutrino oscillation is a genuinely quantum-mechanical effect. By this mechanism one can solve the "solar neutrino problem", that is the discrepancy between the number of electron neutrinos detected from the Sun's core and the theoretical predictions.

With the evidence of neutrinos being massive, the question of whether they are Dirac or Majorana fermions arises. If they are Dirac fermions, they obey the field equations (5.40)

$$i\gamma^\mu \partial_\mu \nu_L = m_D \nu_R \qquad\qquad i\gamma^\mu \partial_\mu \nu_R = m_D \nu_L.$$

But in the Standard Model no right-handed neutrino or left-handed anti-neutrino exist and these equations become inapplicable. However, as exemplified by (B.35) one can construct from a left-handed spinor ψ a Majorana spinor $\chi_L := \psi_L + (\psi^C)_R$. This allows for a Lorentz-invariant Majorana mass term (5.44), which in case of only left-handed neutrinos reads

$$\mathcal{L}_{\nu m} = -m_M \nu_L^T C \nu_L.$$

This looks like a promising route to incorporate massive left-handed neutrinos into the Standard Model. Observe, however, that this mass term changes the lepton number by two units, and this is hardly reconcilable with the Standard Model. Furthermore, one would need a specific Higgs mechanism (or some other procedure) to produce a Majorana mass term. These considerations seem to be the first serious deviation from the predictions of the Standard Model.

The $\Delta L = 2$ lepton-number violating property of Majorana neutrinos explains why currently a lot of attention is being devoted to double beta decay. While single beta decay $(Z, A) \to (Z + 1, A) + e^- + \bar{\nu}_e$ proceeds according to Fig. 6.9b, the double beta process is

$$(Z, A) \to (Z + 2, A) + 2e^- + \bar{\nu}_e + \bar{\nu}_e.$$

Here the lepton number is conserved. But for the case of Majorana neutrinos, it is conceivable that the neutrino annihilates with itself. This is called neutrino-less double beta decay, with a violation of lepton number ($\Delta L = 2$). If this process were to be observed, one would have an unambiguous signal of a Majorana mass.

There are many proposals in place that extend the Standard Model by massive neutrinos through querying parity violation and the ensuing left-right asymmetry. Of course, experiments tell us that parity is violated–at least up to energies presently achievable. It might be the case that parity conservation is restored at higher energies–or–the other way round, parity is a broken symmetry of another theory encompassing the Standard Model. If right-handed neutrinos are present, their weak interaction must be suppressed compared to the interaction of the left-handed ones: the smaller the neutrino mass the larger the suppression must be. Let me mention only one appealing left-right symmetric alternative [369] which is built analogously to the Glashow-Salam-Weinberg model, but being based on $\mathbf{SU(2)_L} \times \mathbf{SU(2)_R} \times \mathbf{U(1)_{B-L}}$ as symmetry group. Here, $\mathbf{U(1)}$ refers to the difference of baryon and lepton number.

Although we still have no theoretical understanding of why the quarks and leptons have a particular mass, the case of massive neutrinos is even harder to understand: Why are their masses so much smaller than the ones of the other basic constituents? A widely adopted explanation is provided by the *see-saw mechanism*: Assume that massive right-handed neutrinos exist, but that they do not interact electroweakly. You find them under the name "sterile" neutrinos in the literature. They interact only with gravitation (and the Higgs field) because of their mass. In the presence of both left- and right-handed massive neutrinos, it is conceivable that they receive a mass contribution from the Dirac mass term

$$\mathcal{L}_{Dmass} = -m_D(\bar{\nu}_R \nu_L + \bar{\nu}_L \nu_R)$$

and from a Majorana mass term

$$\mathcal{L}_{Mmass} = -\frac{1}{2} m_L(\bar{\nu}_R^C \nu_L + \bar{\nu}_L \nu_R^C) - \frac{1}{2} m_R(\bar{\nu}_L^C \nu_R + \bar{\nu}_R \nu_L^C).$$

These can (up to a sign) be brought into the matrix form

$$(\bar{\nu}_L, \bar{\nu}_L^C) \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} \nu_R^C \\ \nu_R \end{pmatrix} + h.c.$$

Now assume that there is a mass scale M in the model such that

$$m_R = M \gg m_D \gg m_L = \mu.$$

Then the mass matrix can be diagonalized with eigenvalues $m_l \sim (\mu – m_d^2/M)$, $m_h \sim M$, giving rise to one very heavy and one very light neutrino as eigenstates. In some scenarios the scale for the superheavy neutrinos is $M \sim 10^{14}$ GeV, in others it becomes of the order of the grand unified theory scale ($\sim 10^{15-16}$) GeV. The superheavy neutrinos have of course not been produced in experiments, but due to the thermodynamic history of our universe, they could be present and they qualify as candidates for the dark mass in the concordance model of cosmology.

Neutrinos were and still are something special and always full of surprises. The most recent surprise is their small but nonvanishing mass. The implication of massive neutrinos for physics and cosmology are in greater depth treated by [314, 370].

6.4.3 Anomalies

Most of the symmetry considerations in this book refer to variational symmetries, that is symmetries of the classical action. As was found in the late 1960's by S. Adler, and by J.S. Bell and R. Jackiw, quantum radiative corrections can spoil the classical symmetries. Their case (chiral anomaly, treated below) gave rise to the adjective "anomalous" symmetry, but it is not the only possible case. Today one understands that the reason why an exact symmetry of the classical action ceases to be a symmetry in the quantum theory relates formally to the functional measure in the path-integral.

Albeit anomalies are treated here in the chapter on particle physics, this should not lead to the impression that they are peculiar to the Standard Model. They also possibly appear in grand unified theories, but were shown to be absent in superstrings (this discovery lead to euphoria in the mid-1980's), and they are under investigation in theories of gravity. There are harmless and harmful (not to say mortal) anomalous symmetries, depending on whether they afflict the renormalization of a theory or not. In fact, anomalies are detrimental to gauge theories. Since the following presentation of anomalies is (regrettably) sketchy a reference to books on field theories is appropriate, and I specifically recommend Chap. 19 of [412], where the essentials are at first worked out in the more simple case of two-dimensional QED. A comprehensive account of anomalies in quantum field theories is [39].

Chiral Anomalies

Take a step back from the Standard Model and consider the Lagrangian $\mathcal{L} = \overline{\psi} i \gamma^\mu (\partial_\mu - ieA_\mu)\psi$. This simple model exhibits an axial symmetry

$$\psi(x) \to e^{i\lambda\gamma^5}\psi(x). \tag{6.34}$$

According to Noether's theorem the corresponding axial current

$$J_A^\mu = \overline{\psi}\gamma^\mu\gamma^5\psi \tag{6.35}$$

is classically conserved. As it turns out, higher-order contributions in the quantum field theoretic calculations (technically speaking: those coming from fermion loops[20]) reveal that the divergence of the current becomes

$$\partial_\mu J_A^\mu = \frac{e^2}{(4\pi)^2}\epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}.$$

[20] It would be out of the scope of this book on symmetries to derive this here. The derivation requires a lot more "meat" from quantum field theory.

The Dirac Lagrangian $\mathcal{L} = \overline{\psi} i(\gamma^\mu \partial_\mu - m)\psi$ is not invariant under (6.34) because of the mass term. By use of the Dirac equation $i\gamma^\mu \psi = m\psi$ the divergence of the axial current can be written as $\partial_\mu J^\mu_A = 2\, m\overline{\psi} i\gamma^5 \psi$. In the full theory, given by

$$\mathcal{L} = \overline{\psi}[i\gamma^\mu(\partial_\mu - ieA_\mu) - m]\psi$$

quantum fluctuations lead to the divergence relation

$$\partial_\mu J^\mu_A = 2\, m\overline{\psi} i\gamma^5 \psi + \frac{e^2}{(4\pi)^2}\epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}.$$

Thus even in the limit of vanishing mass, the axial current is not conserved. The generalization of the previous finding to non-Abelian gauge theories is

$$\partial_\mu J^\mu_A(m \to 0) = \frac{g^2}{(4\pi)^2}\epsilon^{\mu\nu\rho\sigma} \mathrm{Tr}(F_{\mu\nu} F_{\rho\sigma}). \tag{6.36}$$

Here, $F_{\mu\nu} = F^a_{\mu\nu} T^a$, where T^a are representations of the gauge algebra generators, and the trace is to be taken over all fermions in the theory. A theory is free of anomalies if the totally symmetric quantity

$$D^{abc} := \mathrm{Tr}[\{T^a, T^b\} T^c] = 0 \tag{6.37}$$

vanishes identically. For some groups this is true independently of how the fermions are represented in the gauge theory, because all representations identically fulfill (6.37). To these "safe" groups [213] belong for instance **SU(2)** and **SO(10)**. Theories with unitary symmetry groups **SU(N)** ($N \geqslant 3$) or with Abelian factors **U(1)** are not *a priori* safe against anomalies. However, anomalies may be avoided by choosing fermionic representations such that (6.37) is fulfilled. This has to be decided on a case-by-case basis.

In general, we may split the fermionic fields into left- and right-handed components. Let the left/right-handed fermions couple to the gauge field with a representation $T^a_{L/R}$. Then the anomaly cancellation condition (6.37) can be written

$$\mathrm{Tr}[\{T^a_L, T^b_L\} T^c_L] - \mathrm{Tr}[\{T^a_R, T^b_R\} T^c_R] = 0.$$

If the gauge field couples to the left- and right-handed fields with an identical presentation (so called vector-like theories) the anomalies in both sectors compensate. This is the case for QED and QCD.

The GSW part of the Standard Model, which is left-right asymmetric, is endangered towards anomalies. Its gauge group is not safe, and we need to check (6.37) for all combinations of indices from **SU(2)** and **U(1)** in the arrangement of fermions as recorded in Table 6.1: (1) let a, b, c refer to the $\mathbf{SU_I(2)}$ generators, then $D^{abc} \propto \delta^{ab}\mathrm{Tr}(I^c)$. Looking at (6.1) we observe that $\mathrm{Tr}(I^c)$ vanishes separately for the quark and the lepton sector; (2) let a, b refer to $\mathbf{SU_I(2)}$ and let c refer to $\mathbf{U_Y(1)}$, then $D^{abc} \propto \delta^{ab}\mathrm{Tr}(Y)$. For the quark sector, $\mathrm{Tr}_q(Y) = 3 \times (4/3)$ and for the lepton sector, $\mathrm{Tr}_l(Y) = -4$. Thus the two contributions cancel each other; (3) let a refer to $\mathbf{SU_I(2)}$ and b, c refer to $\mathbf{U_Y(1)}$, then $D^{abc} \propto \mathrm{Tr}(I^a Y^2) = 0$ both for the quark and the lepton sectors; (4) let a, b, c refer to $\mathbf{U_Y(1)}$, then $D^{abc} \propto \mathrm{Tr}(Y^3)$, and again the quark contribution cancels against the lepton contribution[21].

Baryon and Lepton Number Anomaly

Denote all quark fields by $\mathcal{Q} = \{L_q, R_q, R_Q\}$ and all lepton fields by the set $\mathcal{L} = \{L_l, R_l\}$. Then because of the pairwise appearance of the quark and the lepton fields in the fermionic part $\mathcal{L}_F$ (6.29) and in the Yukawa part $\mathcal{L}_Y$ (6.32) the classical SM action is invariant under the global transformations

$$\mathcal{Q}' = e^{i\epsilon_B}\mathcal{Q} \qquad\qquad \mathcal{L}' = e^{i\epsilon_L}\mathcal{L}.$$

These lead to the classically conserved vector currents

$$j^\mu_{Bar} = \bar{\mathcal{Q}}\gamma^\mu\mathcal{Q} \qquad\qquad j^\mu_{Lep} = \bar{\mathcal{L}}\gamma^\mu\mathcal{L}.$$

The conserved charges corresponding to these Noether currents are the baryon and the lepton number. Again, as in the case of the $\mathbf{U_A(1)}$ symmetry, baryon and lepton number conservation is spoiled on the quantum level. One finds an expression structurally similar to (6.36):

$$\partial_\mu j^\mu_{Bar} = \frac{1}{64\pi^2}\epsilon^{\mu\nu\rho\sigma}\Big(g_I^2\, W^a_{\mu\nu}W^a_{\rho\sigma} + 2g_Y^2\, B_{\mu\nu}B_{\rho\sigma}\sum_{quarks}(y_L^2 - y_R^2)\Big)$$
$$\partial_\mu j^\mu_{Lep} = \frac{1}{64\pi^2}\epsilon^{\mu\nu\rho\sigma}\Big(g_I^2\, W^a_{\mu\nu}W^a_{\rho\sigma} + 2g_Y^2\, B_{\mu\nu}B_{\rho\sigma}\sum_{leptons}(y_L^2 - y_R^2)\Big).$$

Therefore the baryon and lepton numbers are not conserved. Nevertheless (B-L) is conserved since

$$\sum_{quarks}(y_L^2 - y_R^2) = \sum_{leptons}(y_L^2 - y_R^2)$$

in the SM; this is even true for each family separately.

6.4.4 Strong CP Problem

CP violating processes have been known since 1964. They were first observed in decays of K mesons, and then in 2001 for the B mesons. These can, as mentioned before, be described by the electroweak part of the Standard Model in terms of the phase in the CKM matrix.

---

[21] This "magical" cancellation could give rise to the philosophical statement that a world either without leptons or without quarks is inconsistent.

There is, however, a so-called strong CP problem: Symmetry arguments do not exclude from the QCD Lagrangian a kinetic term proportional to

$$\epsilon^{\mu\nu\varrho\sigma} F_{\mu\nu} F_{\varrho\sigma}. \tag{6.38}$$

It was mentioned that this term can be written as a total derivative. Indeed with the so-called Chern-Simons current G^μ

$$\epsilon^{\mu\nu\varrho\sigma} F_{\mu\nu} F_{\varrho\sigma} = 2\partial_\mu \qquad\qquad G^\mu = 2\epsilon^{\mu\nu\varrho\sigma}[A^a_\nu \partial_\varrho A^a_\sigma + \frac{1}{3} f^{abc} A^a_\nu A^b_\varrho A^c_\sigma].$$

Thus you might wonder why one should be worried about this term at all, since as a total derivative it can be dropped from the action and has no influence on the classical field equations. However, in a quantum theory a boundary term cannot immediately be discarded. This is comprehensible from the path integral formulation where one needs to integrate over all histories and not only the classical ones. And as we saw and will see in various other places, boundary terms themselves contain “physics”, signaling for instance a non-trivial topology. Indeed, as known from the existence of instantons, Yang-Mills theories do have a non-trivial vacuum structure which can be characterized by a topological quantum number n, and the true vacuum is a superposition

$$|\theta\rangle = \sum_n e^{-in\theta} |n\rangle.$$

The topological number is given by

$$n = -\frac{g^2}{16\pi^2} \mathrm{Tr}\,(F_{\mu\nu} \widetilde{F}^{\mu\nu})$$

and this is proportional to the expression (6.38). (Notice also that this term has the same structure as the anomaly term (6.36)). Indeed, the θ-vacuum can be mimicked by adding to the Yang-Mills Lagrangian a term $\mathcal{L}_\theta = -n\theta$. Since from the previous reasoning the term (6.38) is factually present in QCD, it engenders unwanted CP violations, and this is the “strong CP problem”. In order to avoid this problem one could impose $\theta = 0$ on the theory. But this is a rather brute-force method and would provoke the next WHY question, and that in a theory claimed to be fundamental! Furthermore we need to ask, whether this restriction is not too narrow.

Consider the QCD Lagrangian complemented by the θ-term:

$$\mathcal{L} = \sum_q \bar{q}(i\not{D} - M)q - \frac{1}{2}\mathrm{Tr}\,(F_{\mu\nu}F^{\mu\nu}) + \frac{\theta g^2}{16\pi^2}\mathrm{Tr}\,(F_{\mu\nu}\widetilde{F}^{\mu\nu}). \tag{6.39}$$

If one of the quarks (q_k, say) had zero mass, then the Lagrangian would be invariant under a global $\mathbf{U(1)}$ transformation $q_k \to \exp(-i\beta\gamma^5)q_k$ of this quark. We know from the previous considerations of the chiral anomaly that this would lead to a change in the Lagrangian

$$\delta\mathcal{L} = -\beta\,\partial_\mu j_5^\mu = 2\beta\frac{g^2}{16\pi^2}\mathrm{Tr}\,(F_{\mu\nu}\widetilde{F}^{\mu\nu}).$$

Therefore the parameter θ in the Lagrangian would shift to $\theta + 2\beta$. Or, to reverse the argument, the parameter could be completely eliminated. Now however from experimental findings none of the quarks is massless; so this does not work, but it shows that the θ-parameter may very well be devoid of any physical meaning. One can argue that in the case of non-vanishing masses a physically meaningful parameter $\bar{\theta} = \theta - \arg(\det M)$ remains in the theory. Therefore the strong CP-problem amounts to explain why $|\bar{\theta}| \lesssim 10^{-9}$, an upper bound coming from the upper bound of the electric dipole moment of the neutron.

The literature on the strong CP problem is rather rich. The still most popular approach is due to a proposal by R. Peccei and H. Quinn from 1977. Again, their arguments are founded on symmetry considerations: We know that in the Standard Model the quark masses arise from spontaneous symmetry breaking and are not encoded in the full theory as in (6.39), but due to a Yukawa term $\bar{q}q\Phi$ which couples the quark fields to the Higgs doublet. These terms are not invariant under the chiral rotations of the quarks, unless...... there is a further symmetry $\mathbf{U_{PQ}(1)}$ acting also on the Higgs in such a way that the quark transformations are compensated. A closer inspection of the quark Yukawa Lagrangian (6.32) reveals that after appropriate phases in the fields are adjusted so that the first term is $\mathbf{U_{PQ}(1)}$ invariant, the second term fails to be invariant as well. Peccei and Quinn solved this dilemma by introducing into the theory a second Higgs doublet which takes the part of $\tilde{\Phi}$. The extra symmetry (called PQ-symmetry) can thus be used to remove the θ-term completely. The PQ-symmetry (and the extra Higgs doublet) of course influence the spontaneous breaking of the gauge symmetry and the PQ symmetry. The Higgs mechanism results in the massive W- and Z-boson, together with a new particle (the *axion* as it was baptized by S. Weinberg and F. Wilzcek), which in principle is a massless Goldstone boson. The complete treatment reveals a tiny mass of the axion. Since this is already ruled out by experiment, the simple Peccei-Quinn mechanism does not work. Therefore, variants have been investigated. The search for axions still continues. Not only particle physicists but also cosmologists are interested in these beasts because they are dark matter candidates. By the way, isn't it intriguing to see how the very idea of symmetries in Yang-Mills Lagrangians leads to possible consequences in the thermodynamic history of our universe?

6.4.5 Standard Model and Effective Field Theories

The Standard Model for elementary particle physics is the best theory known today for the strong, weak, and electromagnetic interactions, extensively tested and verified up to the TeV range. The SM can be interpreted as a member or link in a chain of effective theories. As explained in Subsect. 5.6.4., one may either move down or up the chain. In moving down (in the following called "The top-down *modus operandi*") one has various ways to integrate out heavy modes. In moving up, one understands the Standard Model as the lowest-order term of a series of interactions, with the only requirement that all further terms respect the symmetries of the Standard Model ("Guessing the effective theory beyond the Standard Model"). We also saw that the development of the effective field-theory philosophy is tied to our comprehension of the renormalization group. Therefore a paragraph on the β functions of the SM is included in this section.

The Top-Down Modus Operandi

I have forgone expounding pre-SM theories/models for strong and weak interactions. These were not wrong, but from the vantage point of effective field theories today they can be understood as having a limited range of applicability. One of the very first models of strong forces was the 1935 theory by H. Yukawa in which the interaction is pictured by the exchange of mesons between nucleons (indeed the pion was predicted before its discovery twelve years afterwards). In the early sixties, one found an approximate $\mathbf{SU(2)} \times \mathbf{SU(2)}$ symmetry of strong interactions which triggered work on "soft pions". Today this is interpreted as a limit of QCD where–due to the presence of two light and four further heavy quarks–the pions are interpreted as the massless Goldstone bosons of a spontaneously broken chiral symmetry (or, nearly massless pseudo-Goldstone bosons). Going top-down from a known and validated theory to effective theories has the advantage that otherwise impossible or hard calculations become possible or easier. For the SM a full arsenal of effective Lagrangians has been investigated, see e.g. [130].

- Chiral perturbation theories
  Although we have at our disposal an elegantly simple and renormalizable theory of the strong interactions, QCD is not directly amenable for phenomenological calculations, that is to compare its low energy predictions with experimental findings. Because of asymptotic freedom perturbative techniques can be used only for very high energies. Otherwise one is obstructed in that QCD is formulated in terms of quarks and gluons, whereas experiments are performed with hadrons, the bound states of the fundamental fields. The connection between the "wrong" degrees of freedom (quarks and gluons) and the experimentally accessible "particles" became possible by exploiting the idea of effective field theories. We saw in Sect. 6.2.2. that if formulated with chiral states, the QCD Lagrangian has an exact $\mathbf{SU_L(3)} \times \mathbf{SU_R(3)}$ symmetry if the light (u, d, s) quark masses are set to zero. This amounts to the existence of eight Goldstone bosons making up an octet of pseudoscalars formed by the π, K and η mesons. These are not massless (because chiral invariance is broken by the small quark masses) but they are lighter than the other hadronic bound states in the theory. Applying the theory for energy-momenta small compared to ~ 1 GeV, it was found in late sixties and in the seventies that

experimentally verifiable results can be obtained from effective chiral Lagrangians. These were expressions built solely from requesting chiral invariance for the tree level (with two derivatives of the effective fields) and the one-loop level (with four-derivative interactions) contributions. This effective field theory approach to QCD has high predictive power and is described by one of the players in the game in Chap. 19 of [536]. In a quite recent article, S. Weinberg formulates an effective field theory of constituent quarks, gluons and pions [540]. In combining this with ideas of "large N expansions", where N is the number of colors, he shows that this effective field theory is renormalizable to leading order in $1/N$ with only a finite number of terms in the Lagrangian.

- Heavy quarks
  Another useful approach is heavy-quark effective technique dealing with describing the physics of hadrons with one or two heavy quarks. Here one expands in powers of the heavy quark masses. Again theory and experiment can be compared; see [360]. A related field is "nonrelativistic QCD" used particularly in the calculations for bound states of heavy quarks.
- $(V\text{-}A)$ model
  It was mentioned already that for the electroweak part of the Standard Model, both the Fermi and the $(V\text{-}A)$ model are effective theories. Take as an example the decay of a muon into an electron and neutrinos according to $\mu^- \to e^- \bar{\nu}_e \nu_\mu$. The SM yields for the lowest-order Feynman diagram in Fig. 6.9a an amplitude

$$M_T = \frac{g_I^2}{2}(\bar{e}\gamma^\rho P_L \nu_e)\frac{g_{\rho\sigma} - k_\rho k_\sigma/M_W^2}{M_W^2 - k^2}(\bar{\nu}_\mu \gamma^\sigma P_L \mu).$$

Since the muon mass m_μ is much smaller than the W mass, one may expand the W-propagator in powers of k^2/M_W^2. The lowest-order term is

$$M_T^0 = \frac{g_I^2}{2M_W^2}(\bar{e}\gamma^\rho P_L \nu_e)(\bar{\nu}_\mu \gamma^\sigma P_L \mu).$$

This amplitude would also result from a Lagrangian interaction term

$$\mathcal{L} = 2\sqrt{2} G_F (\bar{e}\gamma^\rho P_L \nu_e)(\bar{\nu}_\mu \gamma^\sigma P_L \mu).$$

Indeed this is a $(V\text{-}A)$-Lagrangian, and an effective Lagrangian to the GSW model of electroweak interactions.

Guessing the Effective Theory Beyond the Standard Model

If we assume that the SM shows deviations from experiments at a scale $\Lambda \gtrsim 10 TeV$, say, we might think of the next member in the chain of effective theories to be given by a Lagrangian

$$\mathcal{L} = \mathcal{L}_{SM} + \frac{1}{\Lambda^2}\sum g_i \mathcal{O}_i$$

where the g_i are dimensionless, so that the $\mathcal{O}_i$ are operators with dimension six,[22] obeying the SM symmetries. Indeed these operators were derived, and as it turns out, there are so many of them that further selection criteria are needed (see [359] Sect.7.1 and references therein).

Another scenario is sketched in Sect. VIII.3 of [578] dealing with an effective theory of proton decay. Again all relevant extra interaction terms in the Lagrangian are (up to coefficients) determined by the requirement that the Standard Model symmetries are respected.

Beta Functions for the Standard Model

The β functions of Abelian and non-Abelian Yang-Mills theories with (massless quarks) have been calculated in the meantime up to four-loop approximation, parametrized as

$$\beta(g) = -\beta_0 \frac{g^3}{16\pi^2} - \beta_1 \frac{g^5}{(16\pi^2)^2} - \beta_2 \frac{g^7}{(16\pi^2)^3} - \beta_3 \frac{g^9}{(16\pi^2)^4} + \mathcal{O}(g^{11}).$$

The coefficients β_i can be expressed completely in terms of parameters characterizing the Yang-Mills symmetry group, e.g. for the first two coefficients

$$\beta_0 = \frac{11}{3} C_A - \frac{4}{3} T_F n_f, \qquad \beta_1 = \frac{34}{3} C_A^2 - \frac{20}{3} C_A T_F n_f - 4 C_F T_F n_f. \tag{6.40}$$

Here C_A and C_F is the value of the quadratic Casimir index (see Appendix A.3.3) in the adjoint and in the fundamental representation, respectively:

$$C_A \delta^{ab} = f^{acd} f^{bcd} \qquad C_F \delta_{ij} = \sum_a (T^a T^a)_{ij},$$

and T_F is the Dynkin index (that is the trace normalization) of the fundamental representation of the symmetry group: $T_F \delta^{ab} = \mathrm{Tr}(T^a T^b)$. Further n_f is the number of flavors. I refrain from displaying the expressions for β_3 and β_4 which are quite lengthy but again are written in terms of powers of the symmetry group operators C_A, C_R, T_R and higher-order group invariants, see [443] where you also find references to the other β_i. These results hold for an arbitrary compact semi-simple Lie group.

Specifically for $\mathbf{SU(N)}$, $N \geqslant 1$

$$T_F = \frac{1}{2} \qquad C_A = N \qquad C_F = \frac{N^2 - 1}{2N}.$$

[22] The only possible dim-5 operator represents a Majorana mass term for the left-handed neutrinos. I thank A. Blum for pointing this out to me.

Therefore a **SU(N)** Yang-Mills theory is asymptotically free if the number of flavors is not too large, namely

$$n_f < \frac{11}{2}N$$

which for $n_f = 6$ is surely fulfilled by all $N \geqslant 2$. This result, established in 1972/1973 by G. 't Hooft, by D. Politzer and by D. Gross and F. Wilcek was a breakthrough in the recognition of QCD as the theory of strong interactions. Since conversely, the coupling becomes larger and larger with decreasing energy, one can no longer rely on perturbation theory.

For a $\mathbf{U(1)}$ theory, one gets because of $C_A = 0$ and $C_F = 1$

$$\beta(g) = \frac{1}{16\pi^2} \times \frac{2}{3} n_f \, g^3 + \frac{1}{(16\pi^2)^2} \times 2n_f \, g^5 + \mathcal{O}(g^7).$$

Thus the beta function for an $\mathbf{U(1)}$ theory has an infrared stable fixed point at $g = 0$. Indeed the coefficient of the first term in the beta function is the same as the one given for QED in 5.6.3, if n_f is interpreted as the number of Weyl spinors in the theory.

The results above are slightly modified in the presence of scalar fields. Assume the following scenario: A Yang-Mills theory based on a compact group includes all possible renormalizable interactions with n_f Weyl fermions ψ^i belonging to unitary irreducible representations s_i and further n_ϕ scalar fields ϕ^α in the representation r_α. Then one can show that in the one-loop approximation

$$\beta_0 = \frac{11}{3} C_a - \frac{2}{3} \sum_i C_{s_i} - \frac{1}{6} \sum_\alpha C_{r_\alpha}$$

(for the proof see any advanced textbooks on QFT). Here the C_a are the Dynkin indices in the adjoint representation and the C_{s_i}, C_{r_α} the Dynkin indices in the s- and r-representations.

In case of the Standard Model with n_H Higgs doublets in the electroweak sector one obtains the $b_i = \beta_0(g_i)$ as

$$b_3 = 11 - \frac{2}{3} n_f \tag{6.41a}$$

$$b_2 = \frac{22}{3} - \frac{2}{3} n_f - \frac{1}{6} n_H \tag{6.41b}$$

$$b_1 = -\frac{2}{3} n_f - \frac{1}{10} n_H \tag{6.41c}$$

referring to the "running" gauge couplings $g_i = \{g_S, g_I, g_Y\}$ with

$$\beta(g_i) = \mu \frac{dg_i}{d\mu} = -\frac{b_i}{16\pi^2} g_i^3 + \mathcal{O}(g^5). \tag{6.42}$$

These are traditionally written in terms of $\alpha_i(\mu) := g_i^2/(4\pi)$ as

$$\frac{1}{\alpha_i(\mu)} = \frac{1}{\alpha_i(\mu_0)} + \frac{b_i}{2\pi} \ln \frac{\mu}{\mu_0} + \cdots. \tag{6.43}$$

In the spirit of effective field theories, n_f in the previous formulae is to be understood as the number of quark flavors below the threshold energy. Between two successive quark masses α_i is continuous and it needs to be matched at the boundaries by adjusting μ_0.

One should also be aware that this picture of running effective couplings (depicted in Fig. 8.1a) is valid only as long the previous perturbative calculations make sense. The coupling α_{QCD}, shown to be small for high energies, increases if going to smaller energies, and may leave the range of perturbation theory. In this domain the quarks and gluons are no longer the right objects to argue with but instead bound states in the form of nucleons, mesons and their resonances are relevant. Indeed, thus far one has not fully succeeded in deriving the effective "low"-energy description from the QCD-Lagrangian.

6.5 Concluding Remarks and Bibliographical Notes

This chapter illustrated what S. Weinberg meant in his Nobel prize talk in Dec. 1979: "To a remarkable degree, our present detailed theories of elementary particle interactions can be understood deductively, as consequences of symmetry principles and of a principle of renormalizability which is invoked to deal with the infinities."[534]

The establishment of the Standard Model for particle physics is one of the great intellectual achievements of the past century. The history of particle physics reflects how theoretical ideas and experimental results from increasingly powerful accelerators intermingled to yield the final–or better today's–picture. Contributions from theoretical and experimental physicists involved in the rise of the Standard Model are collected in [277]. The sociological background of this development is treated in [415]. A very readable account is given by [69]. Personal reflections of some of the major players can be found in [133]. S. Weinberg also considers the history in his Nobel lecture [534]. The Standard Model had attained its shape already in the early seventies, and nothing conceptually new has come to the fore since then. This is also because the resources for the funding of ever-larger accelerators are not at hand. Of course this is frustrating, given that the Standard Model cannot be the final story. Although experimentally verified to better than a percent level (for the recent precision tests see the the latest PDG report [45]), it is believed to be the low-energy version of some underlying "more fundamental"[23] theory. Particle physicists are therefore these days looking somewhat desperately to the newest machine, the LHC in Geneva. Some hopes may have been fulfilled after the discovery in July 2012 of a

[23] Here I put "more fundamental" into quotation marks because the effective field-theory philosophy throws doubts on the meaning of this very conception.

particle that has properties expected of the Higgs. But other phenomena still wait for their disclosure, perhaps supersymmetry, maybe hints for an even higher symmetry or for extra dimensions, and most probably things no one expected thus far. A fairly recent compilation for new TeV-scale physics and their LHC signatures is [371].

In this chapter I left out most of the interesting details of the "old", that is the $\mathbf{SU}(\mathbf{3})$-flavor considerations. There are numerous textbooks from the sixties with the combination "particles" and "symmetries" in the title, e.g. [166]. I also left out details about which experiments lead to the insights about the structure of leptonic and hadronic currents. These details are to be found in newer textbooks on particle physics as e.g. [364], or in standard books on weak interactions [405], [212], or in overview articles [134], [45]. As for the Glashow-Salam-Weinberg model my presentation is sketchy and incomplete. Here you may consult one of its fathers (Sect. 21.3. in [536]). Although a little outdated, I recommend [98], a book that contains more than it announces, as it goes beyond established particle physics and cosmology.

Finally, I recommend for those who are interested in more details on the theoretical aspects of the Standard Model the books [53], [283], [363], and (the more demanding) [432], the latter especially for the group-theoretical aspects.

Last but not least, a remark needs to be made about the unreflected usage of the wording "particle" in this–and in many of the other–chapters in this monograph. Indeed the word particle refers to different entities in the various fields of physics: A particle is meant to be an abstract mass point in classical mechanics, assumes an erratic character by–what is colloquially called–the particle-wave duality in quantum mechanics, is defined group theoretically in a relativistic field theory in the Wigner classification. In particle physics, the concept of a particle is even more dubious, because for instance in the pre-quark area of the field, most of the "elementary particles" are just short and ultra-short living resonances, and in the Standard Model the quarks are confined and evade detection. Further, the "force" particles (photon, W- and Z-bosons, gluons) cannot be identified with the gauge potentials because these carry more ("superfluous") degrees of freedom. A Higgs boson is in the GSW-theory well described by a spin-0 field, but the many thousands of particle physicists at CERN need to identify the "Higgs" by its peculiar decays in three-story high detectors after subtracting all other known decay modes and by using a lot of filters and heavy data reduction programs. Thus the question "What is a particle?" should not only be answered by philosophers of physics [172]–who, strange enough, do it more intensive than "particle" physicists themselves. And one should also bear in mind that in curved spacetime the notion of a particle becomes observer dependent, see e.g. [112].

Chapter 7
General Relativity and Gravitation

Mit der Einsteinschen Relativitätstheorie hat das menschliche Denken über den Kosmos eine neue Stufe erklommen. Es ist, als wäre plötzlich eine Wand zusammengebrochen, die uns von der Wahrheit trennte: Nun liegen Weiten und Tiefen vor unserem Erkenntnisblick entriegelt dar, deren Möglichkeiten wir vorher nicht einmal ahnten. Der Erfassung der Vernunft, welche dem physikalischen Weltgeschehen innewohnt, sind wir einen gewaltigen Schritt nähergekommen. [549]

H. Weyl was not the only scientist getting enthusiastic about general relativity (GR)–notwithstanding that he was one of the very first admires of general relativity for its wide-reaching influence on physics and philosophy. Here is another typical one: "General relativity is the most beautiful physical theory ever invented" (Sean Carroll's first sentence in the preface of [78].) As you will realize in this chapter, GR acquires its aesthetics and internal stringency from a "very large" group of symmetry transformations, related to general coordinate transformations.

You will notice from the heading that this chapter is not only on general relativity, but also covers theories of gravitation in general[1]. Although GR is currently the by far best theory of gravitation around, there are numerous modifications and extensions. This diversity exists because the potential action functionals for gravitation are less stringently determined by fundamental principles, as compared to internal symmetries.

7.1 Introductory Remarks

As of today we know four fundamental forces, or in the comprehension of the contemporary field-theoretic approach, four interactions. Three of them, namely the electromagnetic, the weak and the strong interactions are the theme of the standard

From the first English translation: "Einstein's theory of relativity has advanced our ideas of the structure of the cosmos a step further. It is as if a wall which separated us from Truth has collapsed. Wider expanses and greater depths are now exposed to the searching eye of knowledge, regions of which we had not even a presentiment. It has brought us much nearer to grasping the plan that underlies all physical happening."

[1] Indeed there is a seminal journal titled "General Relativity and Gravitation".

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_7,

model and were dealt with in the previous chapter. The interaction not treated so far is gravity. Although the gravitational force is the oldest force known to mankind, it seems to be the least understood. On the one hand, gravitational experiments are difficult to perform because of the smallness of the gravitational constant $\kappa = \frac{8\pi G}{c^2}$ (where G is the Newtonian gravitational constant). On the other hand, its full theoretical description is hampered by κ having mass dimension $[\kappa] = -2$ which has prevented the construction of a renormalizable quantum theory of gravity.

Gravity is the weakest among the four interactions, but because of its universality (coupling to all forms of known matter), its range, and its signature (acting on all masses and energies with the same polarity), it is of course the one we experience in our daily right from the beginning. Strangely enough, today we are able to describe the gravitational interaction only within a frame seemingly completely different from that frame familiar from the other three interactions in the standard model.

- GR is *the* theory of gravitation formulated by A.Einstein in its final form in 1915. It is not only conceptually "beautiful", but survived all tests with a high degree of precision.
- "general relativity" is an unfortunate appellation (as is "special relativity", Einstein's coup from 1905). Einstein himself wrote in a letter in 1921 that the name "invariance theory" would be more appropriate[2]. This, as we will see, is indeed much more reasonable from the perspective of its symmetries[3].
- As described in Chap. 3, the theory of special relativity came into existence as a result from the dilemma that Maxwell's electrodynamics is not Galilei invariant, in contrast to Newtonian mechanics[4]. In this sense, there was no real need for a theory of gravitation beyond the one known since Newton's time and very successfully applied in astronomy. Einstein was instead motivated by epistemological motives, e.g. how to go (1) from uniform motions of inertial systems to accelerated motions; (2) from Galilei/Minkowski coordinates to arbitrary coordinates; and (3) from Newton's "action-at-a-distance" concept to a field theory of gravitation.
- The transition from classical mechanics to special relativity meant a replacement of Euclidean geometry by Minkowski geometry. We will see that the transition to general relativity implies the substitution of Minkowski geometry by Riemannian geometry. This is accompanied by different symmetry groups:
  – Mechanics: Galilei group
  – Special relativity: Poincaré group
  – General relativity: Diffeomorphism group (related to the group of general coordinate transformations).

---

[2] "*Nun zum Namen Relativitätstheorie. Ich gebe zu, dass dieser nicht glücklich ist und zu philosophischen Mißverständnissen Anlaß gegeben hat. Der Name Invarianz-Theorie würde die Forschungs-Methode der Theorie bezeichnen, leider aber nicht den materiellen Gehalt der Theorie.*" Letter to E. Zschimmer 30.9. 1921

[3] M. Planck coined the name *Relativtheorie*, which was changed to *Relativitätstheorie* by A. Bucherer in 1906.

[4] As mentioned in Sect. 3.1.2 this only holds true for point transformations.

- Einstein himself came to his final form of general relativity along interesting and rather convoluted paths[5]. This was partly due to his belief in certain principles (equivalence principle, principle of general covariance, Mach principle) which were not properly defined at the time. Only later did one come to a more clear understanding of the semantics of general relativity, and these principles are today relevant only with regard to the history and philosophy of science. Nevertheless, from the viewpoint of symmetries, a clarification of notions like "covariance" and "invariance" is still a current topic; more about this in Subsect. 7.5.3.

7.2 Equivalence Principle

7.2.1 Different Versions of the Equivalence Principle

"There are almost as many statements of a principle of equivalence as there are authors" (J.L. Anderson in [7])

(A) Identity of Inertial and Gravitational Mass

The two force laws ascribed to Newton are

$$\vec{F} = m_I \ddot{\vec{x}} \qquad \text{for arbitray forces}$$
$$\vec{F} = m_G \vec{g} \qquad \text{for gravitation,} \qquad |\vec{g}| = \frac{GM_E}{R_E}, \tag{7.1}$$

where M_E and R_E are for example the mass and the radius of the earth. These laws are taught already in schools; however (usually)without mentioning that there are in principle two different mass quantities, the inertial mass m_I and the gravitational mass m_G, which could in principle differ by to any degree. Newton's two simple laws have as a consequence that in a gravitational field

$$\ddot{\vec{x}} = \frac{m_G}{m_I}\vec{g}.$$

Newton already knew that the inertial and the gravitational mass differ by less than a percent. Further experiments gave better bounds for the ratio of m_G and m_I. Today, it has been experimentally verified that in vacuum all objects–independent of their nature–do have the same gravitational acceleration with a precision of

$$\frac{m_G}{m_I} = 1 \pm 10^{-14}.$$

[5] These are comprehensively treated in [282], [439].

Further experiments–all performed by satellites–to test this UFF (Universality of Free Fall) are operating or planned. One of these is the project STEP (Satellite Test of the Equivalence Principle). On the Stanford university STEP home page, one can read: "STEP will advance the testing of the Equivalence Principle from several parts in 10^{13} to 1 part in 10^{18}. Whether it confirms Equivalence five to six orders of magnitude more precisely than known today or discovers a violation, it will be a landmark experiment in Fundamental Physics, with consequences extending from gravitation theory to cosmology to theories of the evolution of the Universe. It will probe a large and otherwise inaccessible domain in the parameter space of new interactions. A null result would remain for many years a severe constraint on new theories. A positive result would constitute the discovery of a new force of Nature."

(B) "Weak" Equivalence Principle

In an alternative or related manner the identity of inertial and gravitational mass can be expressed as "The effect of a homogeneous static gravitational field can be transformed away." Consider a system of N mass points in a homogeneous gravitational field and with further forces $\vec{F}_{ij} = \vec{F}(|\vec{x}_i - \vec{x}_j|)$ depending only on the relative distances. The Newton equations are

$$m_i \ddot{\vec{x}}_i = m_i \vec{g} + \sum_j \vec{F}(|\vec{x}_i - \vec{x}_j|).$$

Introduce a new coordinate system, corresponding to an accelerated free-falling coordinate system:

$$\vec{x}_i' = x_i - \frac{1}{2}\vec{g}t^2; \qquad \text{i.e.} \quad \ddot{\vec{x}}_i' = \ddot{\vec{x}}_i - \vec{g} \qquad \vec{x}_i' - \vec{x}_j' = \vec{x}_i - \vec{x}_j$$

with the consequence that in these new coordinates, because of

$$m_i \ddot{\vec{x}}_i' = \sum_j \vec{F}(|\vec{x}_i' - \vec{x}_j'|),$$

the gravitational force no longer appears. This seems like a magic trick, but be aware that this expression is only an approximation holding for small masses near the surface of the earth. Indeed we were assuming that the gravitational field is homogeneous, but the gravitational field of a massive sphere is not homogeneous. From this, the tidal forces arise, leading to a stretching of objects longitudinally and a compression transversally.

(C) "Strong" Equivalence Principle

The previous considerations give hints, that acceleration and gravity are indistinguishable only locally. This leads to the formulation of the "strong" equivalence principle: At every point of spacetime within a gravitational field, we can choose a local inertial system such that in a sufficiently small neighborhood, the laws of nature assume the same form as in a non-accelerated coordinate system without gravitation.

(D) "Mathematical" Equivalence Principle

This form of the equivalence principle is the mathematical specification of the strong versions for those spacetimes $(M, g_{\mu\nu})$ which are dealt with in most gravitational theories, namely manifolds M with a metric $g_{\mu\nu}$: At each point P of a spacetime $(M, g_{\mu\nu})$ we can install a coordinate system (x^λ) in such a way that

$$g_{\mu\nu|P} = \eta_{\mu\nu} + \frac{1}{3} R_{\mu\varrho\sigma\nu|P}\, x^\varrho x^\sigma + \mathcal{O}(x^3).$$

Here $\eta_{\mu\nu}$ is the Minkowski metric and $R_{\mu\varrho\sigma\nu}$ is the Riemann curvature tensor belonging to the metric (details to be given in Subsect. 7.3.3 below). These so-called Riemann normal coordinates are unique up to Lorentz transformations. In other words, the curved spacetime metric can be approximated by a flat Minkowski metric in regions which are small compared to $|R_{\mu\varrho\sigma\nu}|^{-1/2}$. Furthermore, this reveals that a non-vanishing curvature tensor is a signal for a gravitational field.

(E) Universality Principles

Clifford Will [561] introduced under the name "Einstein Equivalence Principle" (EEP) three distinct universality principles

- UFF: Universality of Free Fall
- LLI: Local Lorentz Invariance: "The outcome of any local non-gravitational experiment is independent of the velocity of the free-falling reference frame in which it is performed."
- LPI: Local Position Invariance: "The outcome of any local non-gravitational experiment is independent of where and when in the universe it is performed."

Very often the LPI is expressed by the more concrete Universality of Clock Rates (UCR) or the Universality of Gravitational redshift (UGR). These state that the rates of standard clocks agree if taken along the same world line and that they show the standard redshift if taken along different worldlines. It is claimed that if these principles are valid, gravity can be described by a Riemannian space-time metric. For a further detailed analysis see [480], [224]

(F) Other Versions

- geodesics: Structureless free-falling test particles follow geodesics (that is their trajectories are determined by (7.4)). This is true in all generally covariant theories for which the first order and the second order formulation yield the same field equations. Observe that this holds for structureless particles that is particles without charge, spin, …and for test particles, that is particles that do not back-react with the gravitational field they are subjected to.

- minimal coupling: The minimal coupling amounts to replacing the ordinary derivative in a flat-space (scalar, spinor, vector, …) field theory by the covariant derivative. By this one obtains the field theory in curved spacetime–or–the field theory coupled to the gravitational field; see Sect 7.4.2.

7.2.2 Reference Systems and Gravitation

Consider a mass point in uniform motion. There is an inertial system **(x)** in which this motion is described by the equation

$$\frac{d^2x^\mu}{d\tau^2} = 0 \tag{7.2a}$$

$$d\tau^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \tag{7.2b}$$

where $\eta_{\mu\nu}$ is the Minkowski metric. Now these equations shall be transformed to another arbitrary coordinate system (x'). For ease of notation we define first the expressions

$$\mathcal{J}^\mu_{\ \nu} := \frac{\partial x'^\mu}{\partial x^\nu} \qquad\qquad \mathcal{K}^\mu_{\ \nu} := \frac{\partial x^\mu}{\partial x'^\nu} \tag{7.3}$$

which shows that as a matrix, $\mathcal{K}$ is the inverse to the matrix $\mathcal{J}$, or $\mathcal{K}^\nu_{\ \lambda}\mathcal{J}^\lambda_{\ \mu} = \delta^\nu_\mu$. Furthermore

$$\frac{\partial \mathcal{J}^\mu_{\ \nu}}{\partial x^\lambda} := \mathcal{J}^\mu_{\ \nu\lambda} = \mathcal{J}^\mu_{\ \lambda\nu},$$

and a similar relation holds for $\mathcal{K}^\mu_{\ \nu}$. Then

$$\frac{dx^\mu}{d\tau} = \frac{\partial x^\mu}{\partial x'^\nu}\frac{dx'^\nu}{d\tau} = \mathcal{K}^\mu_{\ \nu}\frac{dx'^\nu}{d\tau}.$$

The equations of motion (7.2a) can be rewritten as

$$\frac{d^2x^\mu}{d\tau^2} = \mathcal{K}^\mu_{\ \nu}\frac{d^2x'^\nu}{d\tau^2} + \mathcal{K}^\mu_{\ \nu\lambda}\frac{dx'^\lambda}{d\tau}\frac{dx'^\nu}{d\tau} = 0,$$

or, after multiplication with $\mathcal{J}^\varrho_\mu$:

$$\frac{d^2x'^\varrho}{d\tau^2} + \left\{{}^{\varrho}_{\lambda\nu}\right\}\frac{dx'^\lambda}{d\tau}\frac{dx'^\nu}{d\tau} = 0, \tag{7.4}$$

with the *Christoffel symbol*

$$\left\{{}^{\varrho}_{\lambda\nu}\right\} := \mathcal{J}^\varrho_{\ \mu}\mathcal{K}^\mu_{\ \nu\lambda},$$

which is evidently symmetric in its two lower indices λ and ν. In the new coordinates, the equations of motion (7.4) reveal that in general, the motion is no longer uniform but that because of

$$\frac{du'^{\varrho}}{d\tau} =: t^{\varrho} = -\left\{ {\varrho \atop \lambda\nu} \right\} u'^{\lambda} u'^{\nu}$$

inertial forces t^{ϱ} may emerge. This phenomenon is well known from classical mechanics: In going from an inertial system to a rotating coordinate system, you put forth the centrifugal and the Coriolis force. In the new coordinates, Eq. (7.2b) becomes

$$d\tau^2 = g_{\mu\nu} dx'^{\mu} dx'^{\nu}. \tag{7.5}$$

with the new metric

$$g_{\mu\nu} = \eta_{\varrho\sigma} \mathcal{K}^{\varrho}{}_{\mu} \mathcal{K}^{\sigma}{}_{\nu}.$$

After some algebraic manipulations, we find a relation between the derivatives of the metric and the Christoffel symbols in the form

$$g_{\nu\lambda,\mu} = \left\{ {\varrho \atop \nu\mu} \right\} g_{\rho\lambda} + \left\{ {\varrho \atop \lambda\mu} \right\} g_{\nu\rho} \quad \text{or} \quad g_{\varrho\lambda} \left\{ {\lambda \atop \mu\nu} \right\} = \frac{1}{2} (g_{\varrho\mu,\nu} + g_{\nu\varrho,\mu} - g_{\mu\nu,\varrho}). \tag{7.6}$$

7.2.3 Geodesics

Because of the equivalence principle (local non-discriminability of accelerated motion from the effect of a gravitational field), the equation

$$\frac{d^2 x^{\lambda}}{d\tau^2} + \left\{ {\lambda \atop \mu\nu} \right\} \frac{dx^{\mu}}{d\tau} \frac{dx^{\nu}}{d\tau} = 0 \tag{7.7}$$

is the equation of motion for a structureless test particle in a gravitational field (free falling-object). It is reminiscent to the equation defining the curve of minimal length in the geometry specified by the metric $g_{\mu\nu}$.

Auto-Parallels, Geodesics and Curves of Minimal Length

The geometries one is dealing with in formulating gravitational theories (Riemann or Riemann-Cartan spacetimes) are affine geometries: In these, parallel lines are mapped onto parallel lines and ratios of distances along parallel lines are preserved. Affine geometries are manifolds with a connection Γ. Every connection entails prominent types of curves:

- $x(\lambda)$ is called an *auto-parallel* with respect to the connection Γ if it obeys

$$\frac{d^2 x^{\nu}}{d\lambda^2} + \Gamma^{\nu}_{\mu\rho} \frac{dx^{\mu}}{d\lambda} \frac{dx^{\rho}}{d\lambda} = 0. \tag{7.8}$$

These equations are invariant under the transformations $\lambda \to a\lambda + b$ (where a and b are constants), and parameters related by this transformation are called affine parameters.

- A *geodesic* is defined to be a curve $x^\mu(\lambda)$ whose tangent vector points in the same direction as itself when parallel propagated. (The intuitive notion is to define a geodesic as a curve as straight as possible in the given geometry.) In a geometry with an affine connection $\Gamma^\rho_{\mu\nu}$ a geodesic is defined by

$$\frac{d^2x^\rho}{d\lambda^2} + \Gamma^\rho_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = \alpha\frac{dx^\rho}{d\lambda}$$

where $\alpha(\lambda)$ is an arbitrary function of the variable that parametrizes the curve. On changing the parameter as $\lambda \to p(\lambda)$ this equation is converted to:

$$\frac{d^2x^\rho}{dp^2} + \Gamma^\rho_{\mu\nu}\frac{dx^\mu}{dp}\frac{dx^\nu}{dp} = \left(\alpha - (\frac{d^2p}{d\lambda^2})(\frac{dp}{d\lambda})^{-2}\right)\frac{dx^\rho}{dp} =: \tilde{\alpha}(p)\frac{dx^\rho}{dp}$$

which is again the equation of a geodesic. Obviously one can always find a parametrization such that $\tilde{\alpha}(p) = 0$–just by solving the differential equation

$$\frac{d^2p}{d\lambda^2} - \alpha(\lambda)(\frac{dp}{d\lambda})^2 = 0. \tag{7.9}$$

Therefore in an affine geometry auto-parallels and geodesics are not distinguishable.

The *curve of minimal length* is derived from extremizing the arc length:

$$\delta\int_A^B ds = 0 \qquad \text{with} \qquad ds^2 = g_{\mu\nu}dx^\mu dx^\nu.$$

The explicit calculation is as follows: define $x'^\mu := \frac{dx^\mu}{d\lambda} = v^\mu$. The variational problem is thus

$$\delta\int_A^B ds = \delta\int_A^B d\lambda\sqrt{g_{\mu\nu}v^\mu v^\nu} = 0 \tag{7.10}$$

with the "Lagrangian" $L = \sqrt{g_{\mu\nu}v^\mu v^\nu}$. The resulting Euler-Lagrange equations are

$$\frac{1}{2L}g_{\mu\nu,\rho}v^\mu v^\nu - \frac{d}{d\lambda}\big(\frac{1}{L}g_{\mu\rho}v^\mu\big) = 0$$

or

$$\frac{1}{L}\Big(\frac{1}{2}g_{\mu\nu,\rho}v^\mu v^\nu - \frac{d}{d\lambda}g_{\mu\rho}v^\mu\Big) = (\frac{d}{d\lambda}\frac{1}{L})g_{\mu\rho}v^\mu. \tag{7.11}$$

Now

$$\frac{1}{2}g_{\mu\nu,\rho}v^\mu v^\nu - \frac{d}{d\lambda}g_{\mu\rho}v^\mu = \frac{1}{2}(g_{\mu\nu,\rho} - g_{\mu\rho,\nu} - g_{\nu\rho,\mu})v^\mu v^\nu - g_{\mu\rho}\frac{dv^\mu}{d\lambda}.$$

Contracting this with $g^{\sigma\rho}$ we get

$$g^{\sigma\rho}\left(\frac{1}{2}g_{\mu\nu,\rho}v^{\mu}v^{\nu}-\frac{d}{d\lambda}g_{\mu\rho}v^{\mu}\right)=-\left\{{}^{\sigma}_{\mu\nu}\right\}v^{\mu}v^{\nu}-\frac{dv^{\sigma}}{d\lambda}.$$

Therefore, (7.11) can be written

$$\frac{dv^{\sigma}}{d\lambda}+\left\{{}^{\sigma}_{\mu\nu}\right\}v^{\mu}v^{\nu}=-L\left(\frac{d}{d\lambda}\frac{1}{L}\right)v^{\sigma}. \tag{7.12}$$

This shows that an extremal curve is a geodesic with respect to the Christoffel connection. Furthermore, we identify

$$\alpha=-L\left(\frac{d}{d\lambda}\frac{1}{L}\right).$$

The dependence on L in (7.12) can be removed by solving the differential equation (7.9), which in this case becomes

$$\frac{d^2p}{d\lambda^2}-\frac{1}{L}\left(\frac{dp}{d\lambda}\right)^2=0,$$

and which is solved by $dp/d\lambda = L$. Due to (7.10) the new parameter p has the meaning of the arc-length s with $ds = Ld\lambda$. This works only if $L \neq 0$. If this is the case, one derives the equation for an auto-parallel

$$\frac{d^2x^{\sigma}}{ds^2}+\left\{{}^{\sigma}_{\mu\nu}\right\}\frac{dx^{\mu}}{ds}\frac{dx^{\nu}}{ds}=0 \tag{7.13}$$

which is identical to (7.7). Thus a test particle moves in a curved spacetime (aka a gravitational field) in such a way that it follows the shortest path. At this point, we see the distinction of how to explain the movements of planets in terms of Newton's and Einstein's gravity theories: According to Newton, planets move (in time) around the sun on ellipses in three-dimensional space, while according to Einstein, planets move on geodesics in the four-dimensional geometry curved by the gravitational field of the sun.

"Newtonian Limit"

In Newton's theory of gravity the equations of motion for N mass points moving under the mutual gravitational attraction is given by

$$m_i\ddot{\vec{x}}_i=-G\sum_{j=1,j\neq i}^{N}\frac{m_im_j(\vec{x}_i-\vec{x}_j)}{|\vec{x}_i-\vec{x}_j|^3}$$

where G is the Newtonian gravitational constant. These can also be expressed with the gravitational potential

$$\phi(\vec{x}):=-G\sum_{j=1}\frac{m_j}{|\vec{x}-\vec{x}_j|}=-G\int d^3\vec{x}'\,\frac{\rho(\vec{x}')}{|\vec{x}-\vec{x}'|}$$

with a mass density $\rho(\vec{x})$. We obtain the equations of motion of a point particle and the gravitational field equation in the form

$$\ddot{\vec{x}} = -\vec{\nabla}\phi(\vec{x}) \qquad \text{and} \qquad \triangle\phi(\vec{x}) = 4\pi G\,\rho(\vec{x}). \tag{7.14}$$

The terminology of Newton is in terms of 3-dimensional Euclidean space, absolute time, and gravitational potentials, whereas Einstein conceptualizes gravity in terms of 4-dimensional spacetime and field equations of a geometrical object (the metric). These are quite different mathematical structures. Therefore, a straightforward comparison is not that easy. Based on earlier work by E. Cartan and by K. Friedrichs, J. Ehlers found a framework [148] in which a precise definition of the limit from Newton's description of gravitation Einstein's version can be given. Instead of complying with Ehlers, I follow the argumentation you find in most textbooks. With some caution, the connection between the Newton's and Einstein's concepts is illustrated by taking the Newtonian limit of the geodesic Eq. (7.7), characterized by a mass point which moves "slowly" in a "static" and "weak" gravitational field:

- "slow" in comparison with the velocity of light: $|\vec{v}| \ll 1$, i.e. also $\gamma \approx 1$ and $d\tau \approx dt$, such that the (non-trivial) components of (7.7) become

$$\frac{d^2x^i}{dt^2} + \left\{ {i \atop 00} \right\} \approx 0.$$

- "static": $g_{\mu\nu,0} = 0$ with the consequence that

$$\left\{ {i \atop 00} \right\} = -\frac{1}{2} g^{i\mu} g_{00,\mu} = -\frac{1}{2} g^{ij} g_{00,j}.$$

- "weak": $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}; \quad | h_{\mu\nu} | \ll 1$ from which

$$g^{\mu\nu} \approx \eta^{\mu\nu} - h^{\mu\nu} \qquad \text{and} \qquad \left\{ {i \atop 00} \right\} \approx \frac{1}{2}\frac{\partial}{\partial x^i} h_{00}. \tag{7.15}$$

Eventually the equation of motion for the test particle is reduced to

$$\frac{d^2x^i}{dt^2} \approx -\frac{1}{2}\frac{\partial}{\partial x^i} h_{00}$$

which is to be compared with the Newtonian expression $d^2\vec{x}/dt^2 = -\vec{\nabla}\phi$, as derived before as (7.14). This comparison yields–because of

$$g_{00} = \eta_{00} + h_{00} \approx (1 + 2\phi) \tag{7.16}$$

–the result that the metric has the meaning of a gravitational potential, and that derivatives of the metric (and therefore the Christoffel symbols) are to be interpreted as gravitational forces.

The relation (7.16) reveals that a gravitational field influences the running of a clock: From

$$d\tau = \sqrt{g_{00}}dt \approx (1 + \phi)dt$$

one derives that the running of a clock is retarded in a strong gravitational field in comparison with a weak field[6].

7.2.4 The "Principle" of General Covariance

This "principle" was advocated in various publications by Einstein himself–but nowhere really completely defined. He believed and disbelieved in such a principle at various times in the course of developing GR; see e.g. [395]. Such a principle has been heavily discussed in philosophy of science. Its meaning and role are still not completely settled. Loosely, it states that the dynamical equations of GR should be form invariant (covariant) under general coordinate transformations. Some authors even seem to suggest that general covariance follows from the strong equivalence principle, or that the covariance principle is the mathematical formulation of the physical equivalence principle. A lot of confusion arose in the community by not clearly distinguishing between the concepts of invariance and covariance or from mixing up mere coordinate transformations and genuine diffeomorphism. More about this later (see Subsect. 7.5.3.)

In any case: For special relativity, the invariance of an expression under Lorentz transformations guarantees that the expression is a Lorentz tensor. Since in the context of general relativity we talk about general coordinate transformations, we need a generalization of the Lorentz tensor concept–and these are Riemann tensors.

7.3 Riemann-Cartan Geometry

You saw the nearly inevitable appearance of a metric from the equivalence principle. Thus one is lead to consider spacetime as a four-dimensional manifold which is equipped with a metric and the Christoffel symbol as a further structural element. This Riemannian geometry is the stage for Einstein' general relativity. In order to widen our horizon for extending Einstein's theory by the appropriate differential geometry, to understand the meaning of torsion for gravity theories, and to convey the gauge character of gravitational theories, I will introduce in this subchapter a slightly more general geometry, namely Riemann-Cartan geometry. Furthermore, instead of a metric and an affine connection I introduce tetrads and spin connections. Their use is often preferred because they are more appropriate from a modern geometry-rooted perspective, and since they are absolutely necessary if one wants to couple fermions to a gravitational field. In the following, I will often switch between four-dimensional

---

[6] Together with the influence of the motion on clocks according to special relativity, one can calculate how both relativity theories affect the relative time measurements in GPS satellites and on earth. If these effects were not taken into consideration, GPS systems would show deviations of more than 10 km already within a day.

and D-dimensional spacetimes. Although gravitational dynamics is quite different if one deviates from $D = 4$ most of the differential geometry holds in arbitrary dimensions.

7.3.1 Tensors

We met tensors in the context of Minkowski geometry. In Sect. 3.2.2 they are defined by their transformation properties with respect to Lorentz transformations. If one allows for general coordinate transformations, the notion of tensors must be extended. In a differentiable D-dimensional manifold X_D, tensors may be defined locally by their transformation properties with respect to general coordinate transformations.

- A scalar obeys

$$S'(x') = S(x) \quad \leftrightarrow \quad \delta S = 0.$$

- A contravariant vector transforms as

$$u'^{\mu}(x') = \frac{\partial x'^{\mu}}{\partial x^{\nu}} u^{\nu} = \mathcal{J}^{\mu}_{\nu}(x) u^{\nu}(x) \quad \leftrightarrow \quad \delta u^{\mu} = (\mathcal{J}^{\mu}_{\nu} - \delta^{\mu}_{\nu}) x^{\nu}.$$

  Observe that this is a generalization of Lorentz transformations, namely $\mathcal{J}^{\mu}_{\nu}(x)$ reduces to Λ^{μ}_{ν} in the case that $\mathcal{J}^{\mu}_{\nu}$ is a constant matrix.
- A covariant vector transforms as

$$v'_{\mu} = v_{\nu} \frac{\partial x^{\nu}}{\partial x'^{\mu}} = v_{\nu} \mathcal{K}^{\nu}{}_{\mu}(x).$$

- A generic n-fold contravariant and m-fold covariant tensor transforms as

$$T'^{\mu\dots}_{\varrho\dots} = \mathcal{J}^{\mu}_{\nu} \mathcal{J}^{\cdot}_{\cdot} \dots T^{\nu\dots}_{\sigma\dots} \mathcal{K}^{\sigma}{}_{\varrho} \dots \mathcal{K}^{\cdot}_{\cdot}$$

with the appropriate number of $\mathcal{J}$ and $\mathcal{K}$ matrices.

It is obviously possible to build linear combinations of tensors of the same rank (n, m). Products and contractions of tensors are again tensors.

While there are similarities in the transformation properties of Lorentz and Riemann tensors, there is no analogue for Lorentz spinors, because the diffeomorphism group does not admit finite-dimensional spinor representations[7]. Mostly spinors are assumed to transform as Riemann scalars.

New–compared to the story of Lorentz tensors–is the notion of a tensor density. Its definition is motivated by the transformation property of the determinant $g = \det(g_{\mu\nu})$ of the metric. Building the determinant of the transformed metric $g'_{\mu\nu} = g_{\rho\sigma} \mathcal{K}^{\rho}{}_{\mu} \mathcal{K}^{\sigma}{}_{\nu}$ we deduce $g' = g\,(\det \mathcal{K})^2 = g\,(\det \mathcal{J})^{-2}$, i.e. the

[7] According to Y. Ne'eman and Dj. Šijači infinite-dimensional representations do exist; see [380].

determinant is not a scalar[8]. This motivates the definition: A tensor density of weight w transforms as

$$T^{(w)'\mu\dots}{}_{\varrho\dots} = \mathcal{J}^{\mu}{}_{\nu}\mathcal{J}^{\cdot}{}_{\cdot}\dots T^{\nu\dots}{}_{\sigma\dots}\,\mathcal{K}^{\sigma}{}_{\varrho}\dots\mathcal{K}^{\cdot}{}_{\cdot}\,(\det\mathcal{J})^{-w}.$$

This amounts to assign to the determinant of the metric a weight (-2) such that $(-g)^{w/2}T^{(w)'\mu\dots}{}_{\varrho\dots}$ is a tensor. Since in an action functional, we integrate over a Lagrange density (*sic*!), and since

$$d^4x' = \left|\frac{\partial x'}{\partial x}\right| d^4x = (\det\mathcal{J})\,d^4x$$

the volume is a scalar density with weight $(+1)$ and $\sqrt{-g}\,d^4x$ is a scalar, that is $\sqrt{-g}\,d^4x = \sqrt{-g'}\,d^4x'$.

For later purposes we record the behavior of tensorial objects with respect to infinitesimal coordinate transformations

$$x'^{\mu} = x^{\mu} + \xi^{\mu} \qquad \rightarrow \qquad \mathcal{J}^{\mu}{}_{\nu} = \delta^{\mu}_{\nu} + \xi^{\mu}_{,\nu} \qquad \mathcal{K}^{\mu}{}_{\nu} = \delta^{\mu}_{\nu} - \xi^{\mu}_{,\nu}.$$

For example

$$\begin{aligned} \delta S &= 0 & \bar{\delta} S &= -S_{,\mu}\xi^{\mu} \\ \delta u^{\mu} &= \xi^{\mu}_{,\nu}u^{\nu} & \bar{\delta} u^{\mu} &= \xi^{\mu}_{,\nu}u^{\nu} - u^{\mu}_{,\nu}\xi^{\nu} \\ \delta v_{\mu} &= -\xi^{\lambda}_{,\mu}v_{\lambda} & \bar{\delta} v_{\mu} &= -\xi^{\lambda}_{,\mu}v_{\lambda} - v_{\mu,\lambda}\xi^{\lambda} \\ \delta g_{\mu\nu} &= -\xi^{\lambda}_{,\mu}g_{\lambda\nu} - \xi^{\lambda}_{,\nu}g_{\mu\lambda} & \bar{\delta} g_{\mu\nu} &= -\xi^{\lambda}_{,\mu}g_{\lambda\nu} - \xi^{\lambda}_{,\nu}g_{\mu\lambda} - g_{\mu\nu,\lambda}\xi^{\lambda} \\ \delta g &= -2g\xi^{\lambda}_{,\lambda} & \bar{\delta} g &= -2g\xi^{\lambda}_{,\lambda} - g_{,\lambda}\xi^{\lambda}. \end{aligned}$$

7.3.2 *Affine Connection and Covariant Derivative*

In the case of Minkowski geometry, the partial derivative ∂_{μ} transforms like a Lorentz vector. However, in generic geometries, the partial derivative of a tensor is not necessarily a tensor again. Let us take for instance the relation defining a contravariant vector, $u'^{\mu} = \mathcal{J}^{\mu}_{\nu}(x)u^{\nu}$, and take a partial derivative $\frac{\partial}{\partial x'^{\lambda}}$:

$$\frac{\partial u'^{\mu}}{\partial x'^{\lambda}} = \left(\frac{\partial}{\partial x'^{\lambda}}\,\mathcal{J}^{\mu}_{\nu}\right)u^{\nu} + \mathcal{J}^{\mu}_{\nu}\left(\frac{\partial}{\partial x'^{\lambda}}\,u^{\nu}\right) = \mathcal{J}^{\mu}_{\nu\varrho}\mathcal{K}^{\varrho}{}_{\lambda}\,u^{\nu} + \mathcal{J}^{\mu}_{\nu}\left(\frac{\partial u^{\nu}}{\partial x^{\varrho}}\right)\mathcal{K}^{\varrho}{}_{\lambda}.$$

[8] It is a scalar only if $\det\mathcal{K} = 1 = \det\mathcal{J}$. Indeed, since Einstein originally was not fully aware of the concept of a tensor density, he was inclined to assume that the general coordinate transformations are to be restricted to those for which $\det\mathcal{K} = 1 = \det\mathcal{J}$–even two weeks before he presented the ultimate theory to the Prussian Academy of Science on Nov. 25th, 1915.

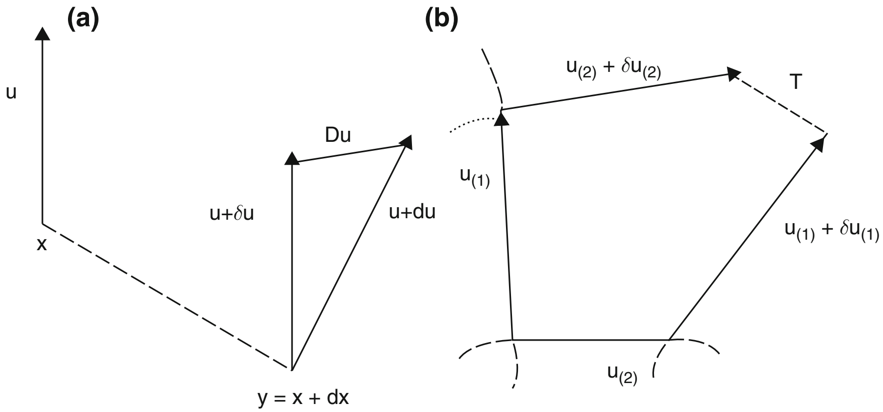


Fig. 7.1 Parallel transport, covariant derivative and torsion

The first term on the right-hand side spoils the expression $\left(\frac{\partial u'^{\mu}}{\partial x'^{\nu}}\right)$ being a tensor. In the case of Minkowski geometry this annoying term vanishes because the Lorentz matrices are constant. Taking a derivative of a tensor always means to compare tensors at different points in spacetime. However, a tensor transforms differently in two points; thus a direct comparison is not possible. Take for instance a vector u^{μ} at a point x and at a neighboring point $y = x + dx$, see Fig. 7.1a. In order to compare u_x with u_y, we must first parallel transport u_x to y and then compare the resulting object u_{pt} with u_y. The components of u_y are given by $u^{\mu}_y = u^{\mu}(x + dx) = u^{\mu} + du^{\mu}$ but these are, in a non-flat geometry, not necessarily identical to u_{pt} at y. The vector u_{pt} differs from $u^{\mu}(x)$ by a small amount: $u^{\mu}_{pt} = u^{\mu}(x) + \Delta u^{\mu}(x)$. This amount can only depend on the vector itself and on the distance dx and we might assume a linear dependence

$$\Delta u^{\mu} = -\Gamma^{\mu}{}_{\rho\sigma} u^{\rho} dx^{\sigma}.$$

Indeed, we are forced to require linearity, since it must be true that $\Delta(u^{\mu} + v^{\mu}) = \Delta u^{\mu} + \Delta v^{\mu}$. Thus the difference we are looking for is

$$Du^{\mu} := u^{\mu}(y) - u^{\mu}_{pt} = du^{\mu} - \Delta u^{\mu} = (\partial_{\sigma} u^{\mu} + \Gamma^{\mu}{}_{\rho\sigma} u^{\rho}) dx^{\sigma}$$

which is the definition of a tensor

$$D_{\sigma}(\Gamma) u^{\mu} := \partial_{\sigma} u^{\mu} + \Gamma^{\mu}{}_{\rho\sigma} u^{\rho} \tag{7.17}$$

called the covariant derivative[9] of the vector u^{μ}. The covariant derivative is defined with respect to the object Γ, called a linear or affine connection. It clearly has $4 \times 4 \times 4 = 64$ components in 4 dimensions. Despite its notation as $\Gamma^{\mu}{}_{\rho\sigma}$, it is

[9] The identically-named covariant derivatives in general relativity and in Yang-Mills theory have common differential-geometric roots. In both cases, the connection is defined as a section in a specific fibre bundle; for more of this advanced geometry see Appendix E.6 and [192].

not a tensor, but transforms in such a way that the troublesome term in the ordinary derivative of a tensor is absorbed, and the covariant derivative of a tensor becomes a tensor itself. Explicitly:

$$\Gamma'^{\mu}{}_{\lambda\nu} = \mathcal{J}^{\mu}_{\rho}\,\Gamma^{\rho}{}_{\sigma\tau}\,\mathcal{K}^{\sigma}{}_{\lambda}\mathcal{K}^{\tau}{}_{\nu} + \mathcal{J}^{\mu}_{\rho}\mathcal{K}^{\rho}{}_{\nu\lambda}. \tag{7.18}$$

If there would be only the first term on the right-hand side, this would be the transformation behavior typical of a one-fold contravariant and two-fold covariant tensor. Since the second term in (7.18) only depends on the coordinate transformations, but not on the connection, the difference of two affine connections is a tensor. And since it is symmetric in the two lower indices of Γ, the antisymmetrized connection transforms as a tensor; this defines the torsion of the connection, which is treated in the next section. For infinitesimal transformations $x'^{\mu} = x^{\mu} + \xi^{\mu}$ the connection transforms as

$$\delta\Gamma'^{\mu}{}_{\lambda\nu} = \Gamma'^{\rho}{}_{\lambda\nu}\xi^{\mu}_{,\rho} - \Gamma'^{\mu}{}_{\rho\nu}\xi^{\rho}_{,\lambda} - \Gamma'^{\mu}{}_{\lambda\rho}\xi^{\rho}_{,\nu} - \xi^{\mu}_{,\lambda\nu}.$$

The covariant derivative is–by construction–defined for every tensor. On scalars it is–by definition–identically equal to the the partial derivative. Due to the fact that $u^{\mu}v_{\mu}$ is a scalar, we immediately get for a covariant vector from the chain rule:

$$D_{\lambda}(\Gamma)v_{\mu} = v_{\mu,\lambda} - \Gamma^{\sigma}{}_{\mu\lambda}v_{\sigma}.$$

For a generic tensor $T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\nu_l}$

$$\begin{aligned} D_{\lambda}(\Gamma)T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\nu_l} &= \partial_{\lambda}T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\nu_l} \\ &\quad + \Gamma^{\mu_1}{}_{\lambda\rho}T^{\rho\ldots\mu_k}{}_{\nu_1\ldots\nu_l} + \ldots + \Gamma^{\mu_k}{}_{\lambda\rho}T^{\mu_1\ldots\rho}{}_{\nu_1\ldots\nu_l} \\ &\quad - \Gamma^{\rho}{}_{\nu_1\lambda}T^{\mu_1\ldots\mu_k}{}_{\rho\ldots\nu_l}\cdots - \ldots\Gamma^{\rho}{}_{\nu_l\lambda}T^{\mu_1\ldots\mu_k}{}_{\nu_1\ldots\rho}. \end{aligned} \tag{7.19}$$

A D-dimensional manifold equipped with a connection Γ is called a *linearly connected space*, and denoted as L_D.

A more-in-depth description of Riemann-Cartan geometry (which is nice to know, but which you do not need for the understanding of this subsection) is given in Appendix E. There the $\{\partial_{\mu}\}$ are recognized as the coordinate basis of the vector fields in the tangent bundle TM to a manifold M, and the $\{dx^{\mu}\}$ are the basis of one-forms in the cotangent spaces T^*M. Tensors are sections in the product of k tangent bundles and l cotangent bundles. A tensor-valued p-form has the structure

$$\mathcal{T}^{\mu\cdot\cdot}_{\nu\cdot\cdot} = \frac{1}{p!}T^{\mu\cdot\cdot}_{\nu\cdot\cdot\lambda_1\cdot\cdot\lambda_p}dx^{\lambda_1}\wedge\cdot\cdot\wedge dx^{\lambda_p}.$$

The connections $\Gamma^{\mu}{}_{\lambda\rho}$ are components of connection one-forms $\Gamma^{\mu}{}_{\rho} = \Gamma^{\mu}{}_{\lambda\rho}dx^{\lambda}$. If $\Gamma = (\Gamma^{\mu}{}_{\rho})$ is understood as a matrix, the covariant derivative as acting on a tensor-valued p-form $\mathcal{T}$ can compactly be written as the matrix equation $D(\Gamma)\mathcal{T} = d\mathcal{T} + [\Gamma, \mathcal{T}]$ where the commutator $[\Gamma, \mathcal{T}]$ is defined by

$$[\Gamma, \mathcal{T}]^{\mu\cdot\cdot}_{\nu\cdot\cdot} = \left(\Gamma^{\mu}{}_{\rho}\mathcal{T}^{\rho\cdot\cdot}_{\nu\cdot\cdot} + \text{all upper indices}\right) - (-1)^p\left(\mathcal{T}^{\mu\cdot\cdot}_{\rho\cdot\cdot}\Gamma^{\rho}{}_{\nu} + \text{all lower indices}\right).$$

7.3.3 Torsion and Curvature

As mentioned before, the affine connection itself is not a tensor. However its antisymmetric part

$$T^{\lambda}{}_{\mu\nu}(\Gamma) := \Gamma^{\lambda}{}_{\mu\nu} - \Gamma^{\lambda}{}_{\nu\mu} = 2\Gamma^{\lambda}{}_{[\mu\nu]} \tag{7.20}$$

transforms as a tensor and is called the Cartan *torsion* tensor (or torsion tensor–for short) belonging to the connection Γ. Its geometrical meaning is related to the fact that infinitesimal parallelograms in an L_D do not close in general, the closure failure being proportional to the torsion tensor: From Fig. 7.1b we find for vectors $u^{\mu}_{(i)}$ tangent to the coordinate lines x^i that $\delta_1 u^{\mu}_{(2)} - \delta_2 u^{\mu}_{(1)} = -T^{\mu}_{21} dx^1 dx^2$. The torsion tensor has $D \times \frac{D(D-1)}{2}$ components. Another useful geometric object is the *contortion* tensor[10] built from the torsion as

$$K^{\lambda}{}_{\mu\nu} := \frac{1}{2}(T^{\lambda}{}_{.\mu\nu} - T_{\mu}{}^{\lambda}{}_{.\nu} - T_{\nu}{}^{\lambda}{}_{.\mu}).$$

While the torsion is antisymmetric in the last pair of indices, the contortion is antisymmetric in the first two indices: $(K^{\lambda}{}_{\mu\nu}) = (K^{\mu}{}_{\lambda\nu})$.

Since parallel transport is path-dependent, the change of a vector v_{μ} due to a parallel transport around an infinitesimal closed path bounded by the surface $\triangle\sigma^{\lambda\rho}$ is

$$\triangle v_{\nu} = \oint \Gamma^{\mu}{}_{\nu\rho} v_{\mu} dx^{\rho} = +\frac{1}{2}\tilde{R}^{\mu}{}_{\nu\lambda\rho} v_{\mu}(\Gamma)\triangle\sigma^{\lambda\rho}$$

where $\tilde{R}^{\mu}{}_{\nu\lambda\rho}(\Gamma)$ is the *Riemann-Cartan curvature tensor*

$$\tilde{R}^{\mu}{}_{\nu\lambda\rho}(\Gamma) := \partial_{\lambda}\Gamma^{\mu}{}_{\nu\rho} + \Gamma^{\mu}{}_{\sigma\lambda}\Gamma^{\sigma}{}_{\nu\rho} - (\lambda \leftrightarrow \rho). \tag{7.21}$$

The torsion and the curvature obey various algebraic and differential identities, which–written in their components–can be found e.g. in [468].

More understanding of the torsion and the curvature tensor can be gained by considering the non-commutativity of covariant derivatives: For a field Φ^A transforming as a Lorentz object, i.e. $\Phi^A = \Sigma^A_B \Phi^B$:

$$[D_{\mu}, D_{\nu}]\Phi^A = \tilde{R}^A{}_{B\mu\nu}\Phi^B - T^{\sigma}{}_{\mu\nu} D_{\sigma}\Phi^A \tag{7.22}$$

where $\tilde{R}^A{}_{B\mu\nu}$ denotes the curvature tensor in the corresponding representation. For instance, for a scalar field ϕ we first find $D_{\mu}D_{\nu}\phi = D_{\mu}(\partial_{\nu}\phi) = \partial_{\mu}\partial_{\nu}\phi - \Gamma^{\sigma}{}_{\mu\nu}\partial_{\sigma}\phi$ from which

$$[D_{\mu}D_{\nu}]\phi = (-\Gamma^{\sigma}{}_{\mu\nu} + \Gamma^{\sigma}{}_{\nu\mu})\partial_{\sigma}\phi = -T^{\sigma}{}_{\mu\nu}\partial_{\sigma}\phi.$$

---

[10] Observe: contortion but not "contorsion" as you can find it in many papers dealing with the tetrad description of spacetime geometry.

In the case of a vector field, after some algebra (see e.g. Sect. 3.6 of [78]), one derives

$$[D_\mu, D_\nu]u^\varrho = \tilde{R}^\varrho{}_{\sigma\mu\nu}u^\sigma - T^\sigma{}_{\mu\nu}D_\sigma u^\varrho.$$

It is to be stressed (and sometimes indicated by the notation) that torsion and curvature are properties of the connection, and since a manifold can have different connections there can be different torsion and curvature tensors. In this sense, a spacetime strictly has neither curvature nor torsion.

In terms of differential form, the torsion and the curvature are two-forms:

$$\begin{aligned} T^\sigma &= D(\Gamma)dx^\sigma = ddx^\sigma + [\Gamma, dx]^\sigma = \Gamma^\sigma{}_\nu dx^\nu = \Gamma^\sigma{}_{\mu\nu}dx^\mu \wedge dx^\nu \\ &= \frac{1}{2}T^\sigma{}_{\mu\nu}dx^\mu \wedge dx^\nu \tilde{R}^\sigma{}_\rho = d\Gamma^\sigma{}_\rho + \Gamma^\lambda{}_\sigma\Gamma^\rho{}_\lambda = \frac{1}{2}\tilde{R}^\sigma{}_{\rho\mu\nu}dx^\mu \wedge dx^\nu. \end{aligned}$$

For the curvature matrix[11] $\tilde{R} = (\tilde{R}^\sigma{}_\rho)$ we can write compactly $\tilde{R} = d\Gamma + \Gamma\Gamma$.

7.3.4 Metric

From the roots of general relativity in the equivalence principle we assume that the manifolds we are dealing with are metric spaces: They have a metric tensor g, that is a non-degenerate symmetric tensor of type (0,2) that allows one to define the scalar product of two vectors as $g_{\mu\nu}u^\mu v^\nu$ and to raise and lower indices (as we already–a little sloppily–did in previous expressions). Observe that the metric and the linear connection are geometric objects, conceptually independent of each other. A metric serves to measure lengths, angles, areas, volumes, whereas a connection refers to properties which remain invariant under affine transformations, such as translations and parallelism. A D-dimensional differentiable manifold with linear connection and metric is denoted as (L_D, g).

With the availability of a metric, one constructs from the curvature tensor further tensors, relevant for gravitational theories:

(i) The *Ricci tensor*: $\tilde{R}_{\mu\nu} := \tilde{R}^\lambda{}_{\mu\lambda\nu}$

(ii) The *curvature scalar*: $\tilde{R} := g^{\mu\nu}\tilde{R}_{\mu\nu} = \tilde{R}^\mu{}_\mu$

(iii) *Einstein tensor*:

$$\tilde{G}^{\mu\nu} := \tilde{R}^{\mu\nu} - \frac{1}{2}\tilde{R}g^{\mu\nu}. \tag{7.23}$$

Define the tensor

$$Q_{\mu\nu\lambda}(\Gamma) := -D_\mu(\Gamma)g_{\nu\lambda} = -\partial_\mu g_{\nu\lambda} + \Gamma^\sigma{}_{\nu\mu}g_{\sigma\lambda} + \Gamma^\sigma{}_{\lambda\mu}g_{\nu\sigma}$$

[11] Not to be confused with the curvature scalar which is introduced below.

called the *non-metricity* tensor. The connection can be written as

$$\Gamma^{\mu}{}_{\lambda\nu} = \left\{{}^{\,\mu}_{\lambda\nu}\right\} + K^{\mu}{}_{\lambda\nu} + \frac{1}{2}\left(Q_{\lambda\nu.}{}^{\mu} + Q_{\nu\lambda.}{}^{\mu} - Q^{\mu}_{.\nu\lambda}\right) \tag{7.24}$$

where $\left\{{}^{\,\mu}_{\lambda\nu}\right\}$ is the Christoffel symbol

$$\left\{{}^{\,\mu}_{\lambda\nu}\right\} = \frac{1}{2} g^{\mu\rho}(g_{\lambda\rho,\nu} + g_{\nu\rho,\lambda} - g_{\lambda\nu,\rho})$$

which me met in the context of generic coordinate transformations and of the geodesic equation, and which is now recognized as a specific linear connection. The covariant derivative with respect to this connection is traditionally abbreviated by a semicolon: $D_\mu(\{\})T = T_{;\mu}$, e.g.

$$u^{\nu}_{;\mu} = u^{\nu}_{,\mu} + \left\{{}^{\,\nu}_{\mu\lambda}\right\} u^{\lambda}.$$

Due to the symmetry in the two lower indices the Christoffel symbols have 40 independent components in $D = 4$.

The splitting of a connection according to (7.24) allows for the distinction of various spaces: The most generic object is the metric affine manifold (L_4, g) itself. Even though metric affine geometries are investigated in the context of gauge approaches to gravitational theories, I will not treat this topic here; interested readers may consult [50], [265]. But let me mention a specific case, namely Weyl geometry for which the connection obeys

$$Q_{\mu\nu\lambda}(\overset{W}{\Gamma}) = a_\mu g_{\nu\lambda}$$

with a Weyl field a_μ. This geometry was investigated by H. Weyl in 1918 in an attempt to formulate a unified theory of gravitation and electromagnetism. Although Weyl's original formulation failed, Weyl spaces are still 'alive', for instance in the context of an axiomatic formulation of general relativity, and in conformal field theories.

Metric Compatibility

As emphasized, one of the key assumptions in the foundation of general relativity is the local Minkowski structure of spacetime. In compliance with this assumption, we assume that locally, in suitable coordinates, the metric assumes the $4D$ Minkowski form $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$. In order to preserve lengths and angles under parallel transport one imposes the condition

$$Q_{\mu\nu\lambda}(\Gamma) = 0. \tag{7.25}$$

In this case, the linear connection is called *metric-compatible*, or a *Levi-Civita connection*. The metricity condition (7.25) still allows for diverse geometries: The space with the most general metric-compatible linear connection is called a *Riemann-*

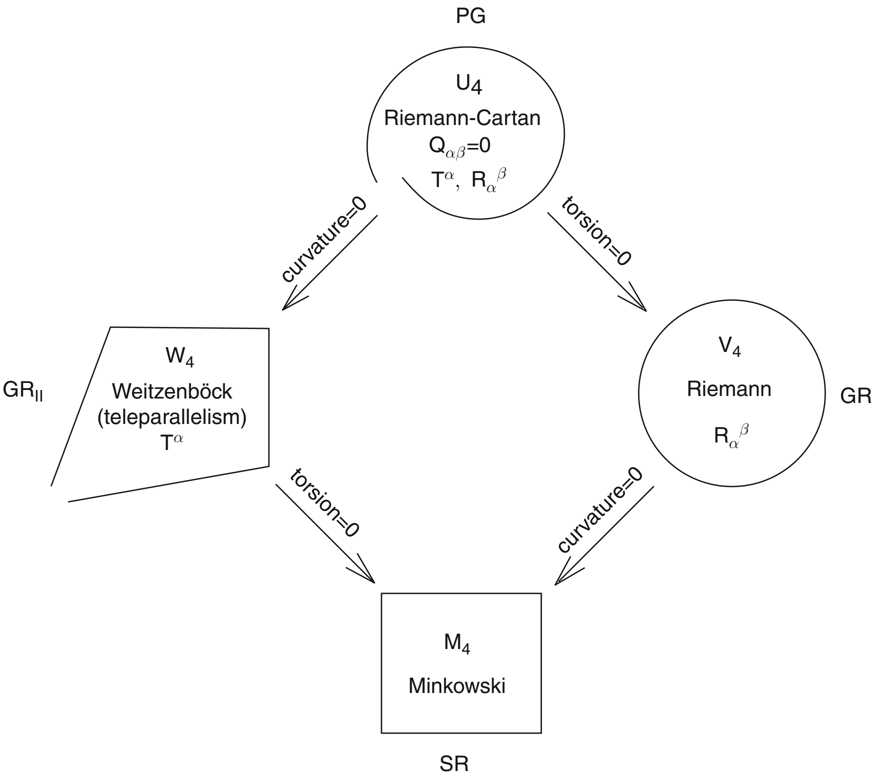


Fig. 7.2 Riemann-Cartan geometries

Cartan space U_4. If the torsion (and thus also the contortion) vanishes we have a *Riemann space* V_4, and if, alternatively, the curvature vanishes, we have *Weitzenböck's teleparallel space* T_4. If both curvature and torsion vanish identically, we are back to Minkowski space M_4, see Fig. 7.2. In a Riemannian geometry where–due to (7.24) - the connection is identical to the Christoffel symbol, we have indeed

$$Q_{\mu\nu\lambda}(\{\}) = -g_{\nu\lambda;\mu} = 0.$$

This makes sense according to the quest for strong equivalence: There is a local inertial system in which the metric is the Minkowski metric and in which the covariant derivative vanishes. Since the metric is a tensor, its covariant derivative vanishes in in every coordinate system.

In U_4 the relation (7.24) boils down to

$$K^{\mu}{}_{\lambda\nu} = \Gamma^{\mu}{}_{\lambda\nu} - \left\{ {\mu \atop \lambda\nu} \right\}. \tag{7.26}$$

which, due to the antisymmetry of the tensor K, reveals that in 4 dimensions the contortion has 24 components.

The curvature tensor, the Ricci tensor and the curvature scalar of U_4 can due to (7.26) be expressed by the respective Riemann space tensors plus terms depending on the contortion. If the Riemann objects $\tilde{R}_{\ldots}(\{\})$ are denoted by $R_{\ldots}$, one can generically make the partition $\tilde{R}_{\ldots}(\Gamma) = R_{\ldots} + \hat{R}_{\ldots}$. For the curvature scalar, in particular, one finds

$$\tilde{R}(\Gamma) = R + 2(K^{\mu\nu}{}_{\nu})_{;\mu} - K_{\mu\sigma\nu}K^{\sigma\nu\mu} + K^{\mu}{}_{\sigma\mu}K^{\sigma\nu}{}_{\nu}. \tag{7.27}$$

We saw that for Riemannian geometries the auto-parallels/geodesics defined with respect to a connection by (7.8) coincide with the curves of minimal length defined by (7.7). In a Riemann-Cartan space we need to distinguish both types of curves. Although the defining equation for a geodesic

$$\frac{d^2x^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0$$

contains only the symmetric part of the connection it is in general not an extremal curve since this must additionally obey $\delta\int ds = 0$. This leads to the geodesic equation in which the symmetric part of the connection is identical to the Christoffel symbol.

Riemann Geometry

Whereas in an U_4-geometry, the metric and the connection are still independent objects, the connection in a Riemann geometry can be calculated completely from the derivatives of the metric. Since the connection–being identical to the Christoffel symbol–has one derivative of the metric, the Riemann curvature tensor containing a derivative of the connection (according to 7.21) has second derivatives of the metric:

$$R_{\mu\nu\lambda\varrho} = g_{\mu\sigma}\left[(\left\{{}^{\sigma}_{\nu\rho}\right\}_{,\lambda} + \left\{{}^{\sigma}_{\tau\lambda}\right\}\left\{{}^{\tau}_{\nu\rho}\right\}) - (\lambda \leftrightarrow \rho)\right].$$

Due to the symmetries of the Christoffel symbol the Riemann curvature tensor obeys the algebraic relations

$$\begin{aligned}
&R_{\mu\nu\lambda\varrho} = R_{\lambda\varrho\mu\nu}\\
&R_{\mu\nu\lambda\varrho} = -R_{\mu\nu\varrho\lambda} = R_{\nu\mu\varrho\lambda} = -R_{\nu\mu\lambda\varrho}\\
&R_{\mu\nu\lambda\varrho} + R_{\mu\lambda\varrho\nu} + R_{\mu\varrho\nu\lambda} = 0.
\end{aligned}$$

Therefore it has in four dimensions only 20 independent components, instead of $4^4 = 256$[12]. And due to these index relations the Ricci tensor $R_{\mu\nu}$ is seen to be symmetric.

[12] In $D > 1$ dimensions, combinatorics yields $\frac{D^2(D^2-1)}{12}$ independent components.

Aside from the algebraic relations among the components of the Riemann curvature tensor there are differential relations

$$R_{\mu\nu\lambda\varrho;\sigma} + R_{\mu\nu\varrho\sigma;\lambda} + R_{\mu\nu\sigma\lambda;\varrho} = 0. \tag{7.28}$$

These are called *Bianchi identities*[13]. The proof makes again use of the equivalence principle and the fact that $R_{\mu\nu\lambda\varrho}$ is a tensor: Consider a point P, choose an inertial system at this point. Now, at P the connection components vanish and the terms with the derivatives of the connections enter the curvature as

$$R_{\mu\nu\sigma\varrho}\Big|_P = \frac{1}{2}[g_{\mu\sigma,\nu\varrho} - g_{\nu\sigma,\mu\varrho} - g_{\mu\varrho,\nu\sigma} + g_{\nu\varrho,\mu\sigma}].$$

Take the covariant derivative of this expression, which at P is the same as the partial one, with the result

$$R_{\mu\nu\sigma\varrho\,;\,\lambda}\Big|_P = \frac{1}{2}[g_{\mu\sigma,\nu\varrho\lambda} - g_{\nu\sigma,\mu\varrho\lambda} - g_{\mu\varrho,\nu\sigma\lambda} + g_{\nu\varrho,\mu\sigma\lambda}],$$

for which cyclical exchange yields (7.28). Because of the Bianchi identities, the Einstein tensor (7.23)–which in a Riemann geometry is also symmetric–obeys

$$G^{\mu\nu}{}_{;\mu} = 0. \tag{7.29}$$

For obvious reasons these equations are called the "contracted Bianchi identity" and we will see later that they are completely equivalent to the Noether identities for vacuum general relativity, due to the diffeomorphism symmetry of the gravitational theory.

Let me further state some relations often helpful in calculations because of the frequent occurrence of $\sqrt{-g} = \sqrt{-\det(g_{\mu\nu})}$ in gravitational theories:

$$\frac{\partial g}{\partial g_{\rho\sigma}} = g\, g^{\rho\sigma} \tag{7.30a}$$

$$\left\{ {\mu \atop \mu\nu} \right\} = \frac{1}{\sqrt{-g}}\partial_\nu\sqrt{-g} \tag{7.30b}$$

$$\sqrt{-g}\, v^{\mu}_{;\mu} = (\sqrt{-g} v^{\mu})_{,\mu}. \tag{7.30c}$$

7.3.5 Tetrads and Spin Connections

Instead of describing an L_4 with a metric and a linear connection, it is possible to characterize it by tetrads (in German: Vierbeine, to be taken literally as "four-legs"; in D dimensions there are D-beins–or "vielbeins") and so-called spin connections. As you will see, this approach is in a certain sense even more geometric as it fits directly to the description within the calculus of differential forms (see Appendix F),

[13] Named after the Italian mathematician Luigi Bianchi (1856–1928).

in which general covariance is trivial, because in this calculus there is no longer a reference to coordinates at all. And we are strongly encouraged to use the tetrad language if we want to describe the gravitational interaction of spinors (remember, these are the "matter" particles). Furthermore, tetrads are the starting point in Loop Quantum Gravity, one of the approaches to quantize the gravitational field.

Tetrads

Due to the equivalence principle, spacetime L_4 has locally the structure of a Minkowski space M_4. The idea is to select at every point p a set of coordinates z_p^I $(I = 0, 1, 2, 3)$ in M_4. Then the spacetime metric can be expressed locally as

$$g_{\mu\nu}(x) = e^I_\mu(x)\eta_{IJ}e^J_\nu(x). \tag{7.31}$$

where $e^I_\mu(x) = \partial_\mu z^I_p(x)$, and where η is the Minkowski metric. The $e^I_\mu(x)$ are called *tetrads*. The upper index numbers four vectors $(I = 0, 1, 2, 3)$ in M_4, the lower index labels the respective components referring to the coordinates $(\mu = 0, 1, 2, 3)$. The index μ is called holonomic (or world, or coordinate space) index, and the index I is the anholonomic (or Minkowski, or tangent frame) index.

Tetrads are defined only up to local Lorentz transformations, since the transformed tetrads

$$e'^I_\mu(x) = \Lambda^I{}_J(x)e^J_\mu(x)$$

leave the metric (7.31) invariant. The tetrads erect, what is called a local frame of references, or *tangent frame* for short. The set union over all spacetime points is called the frame bundle; more about this in Appendix E.5.

Denote by $\{\vartheta^I\}$ a set of four (or, in general, D) linear independent basis vectors in T_pL_4. These are related to the coordinate basis $\{dx^\mu\}$ in the form $\vartheta^I(x) = e^I_\mu dx^\mu$. Defining the inverse tetrad $e^\mu_I(x)$ by[14]

$$e^\mu_I(x) = \eta_{IJ}g^{\mu\nu}(x)e^J_\nu(x)$$

one easily verifies

$$e^I_\mu(x)e^\mu_J(x) = \delta^J_I, \qquad e^I_\mu(x)e^\nu_I(x) = \delta^\nu_\mu, \qquad e^\mu_I(x)g_{\mu\nu}e^\nu_J(x) = \eta_{IJ}.$$

It is possible to express the metric by tetrads, see (7.31). It is not possible, however, to express a tetrad by the metric. This is obvious because a tetrad has 16 components, whereas the metric tensor has 10 independent components. The difference

[14] At other places I will denote the inverse tetrad by E^μ_I; this is advantages in short-hand expressions in terms of matrix relations.

$(16 - 10) = 6$ is accounted for by the existence of six local Lorentz transformations on the tetrads.

To each tensor in the Riemann-Cartan geometry (labeled by world-indices) a frame-space tensor can be constructed by contractions with tetrads and/or inverse tetrads (and *vice versa*).

Spin Connections

The parallel transport of u^I and v_I obeys a rule

$$\Delta u^I = -\omega^I{}_{J\mu} u^J dx^\mu \qquad\qquad \Delta v_I = \omega^J{}_{I\mu} v_J dx^\mu$$

where $\omega^J{}_{I\mu}$ is the *spin connection*. Define covariant derivatives with respect to the spin connection as

$$D(\omega)u^I = D_\mu(\omega)u^I dx^\mu = (u^I{}_{,\mu} + \omega^I{}_{J\mu} u^J)dx^\mu$$

with similar expressions for other tensors with tangent-space indices.

Quite in analogy with the coordinate space formulation, we can write in the case of the tangent frame approach for the spin connection one-form ω:

$$\omega = \omega_I \vartheta^I, \qquad \text{i.e.} \qquad \omega^J{}_{KI}\vartheta^I = \omega^J{}_{K\mu} dx^\mu.$$

The covariant derivative is written compactly

$$\nabla := D(\omega) = d + [\omega, -] \qquad\qquad \nabla = \nabla_I \vartheta^I = \nabla_\mu dx^\mu.$$

Here and in the following, the symbol ∇_μ stands for the covariant derivative with respect to the spin connection, i.e $\nabla_\mu = D_\mu(\omega)$. The covariant derivative with respect to a connection Γ is denoted as $D_\mu := D_\mu(\Gamma)$ in those contexts where it is irrelevant which affine connection the derivative is built with.

What is the relation between the spin connection ω and the linear connection Γ? This can be found from the obvious requirement that the tetrad components of a vector $u^\mu(x)$, parallel transported from x to dx be equal to $u^I + \delta u^I$:

$$u^I + \delta u^I = e^I_\nu(x + dx)\,(u^\nu + \delta u^\nu).$$

After expanding $e^I_\nu(x + dx)$ as $e^I_\nu + (\partial_\mu e^I_\nu)dx^\mu$, and collecting all coefficients of dx^μ, we get

$$D_\mu(\Gamma, \omega)e^I_\nu := \partial_\mu e^I_\nu + \omega^I{}_{J\mu} e^J_\nu - \Gamma^\lambda{}_{\nu\mu} e^I_\lambda = 0 \tag{7.32}$$

where the covariant derivative $D(\Gamma, \omega)$ was introduced. This derivative "sees" all holonomic indices with the affine connection, and all anholonomic indices with the spin connection. The relation (7.32) can be written in two ways, namely as

$$\nabla_\mu e^I_\nu - \Gamma^\lambda{}_{\nu\mu} e^I_\lambda = 0$$

which allows one to express the torsion in terms of the covariant derivatives of the tetrads:

$$T^{\lambda}_{\ \nu\mu} = e^{\lambda}_{I}\, \nabla_{[\mu} e^{I}_{\nu]}$$

or as

$$D_\mu e^I_\nu + \omega^I_{\ J\mu} e^J_\nu = 0.$$

The latter enables us to derive expressions for the spin connection such as

$$\omega_{IJ\mu} = \gamma_{IJ\mu} + K_{IJ\mu} \tag{7.33}$$

where K is the contortion and γ the *Ricci rotation coefficients*

$$\gamma_{IJ\mu} := \frac{1}{2} e^K_\mu\, (\Omega_{IJK} + \Omega_{JKI} - \Omega_{KIJ}) \qquad \text{with} \qquad \Omega_{IJK} = e_{\mu I}\, (e^\nu_J e^\mu_{K,\nu} - e^\nu_K e^\mu_{J,\nu}), \tag{7.34}$$

the latter depending only on the tetrads and their derivatives.

Riemann-Cartan and Riemann Spaces

In the metric perspective, the condition of metric compatibility (7.25) $D_\mu(\Gamma) g_{\nu\lambda} = 0$ leads from a manifold L_4 to a Riemann-Cartan space U_4. What does the metricity condition amount to in the tetrad description? We derive

$$\begin{aligned} 0 = D_\mu(\Gamma) g_{\nu\lambda} = D_\mu(\Gamma, \omega) g_{\nu\lambda} &= D_\mu(\Gamma, \omega) \eta_{IJ} e^I_\nu e^J_\lambda \\ &= e^I_\nu e^J_\lambda D_\mu(\omega) \eta_{IJ} + \eta_{IJ} D_\mu(\Gamma, \omega)(e^I_\nu e^J_\lambda). \end{aligned}$$

Since the last term vanishes due to (7.32), the metric-compatibility condition is equivalent to the condition that the tensor field η_{IJ} obeys

$$D_\mu(\omega) \eta_{IJ} = (\omega^K_{\ I\mu} \eta_{KJ} + \omega^K_{\ J\mu} \eta_{KI}) dx^\mu = 0, \tag{7.35}$$

from which it follows that in a Riemann-Cartan space U_4, the spin connection is antisymmetric[15]

$$\omega^{IJ}_\mu + \omega^{JI}_{\quad \mu} = 0. \tag{7.36}$$

Therefore, it has in four dimensions $4 \times 6 = 24$ independent components.

[15] Here I anticipate that the antisymmetry is closely related to the fact that gravity, if formulated in a Riemann-Cartan geometry, can be interpreted as a gauge theory of the Poincaré group with its antisymmetric Lorentz generators; see Sect. 7.6.3.

Table 7.1 Coordinate system (CS) versus tangent frame (TF)

	CS	TF	CS $\Rightarrow$ TF	TF $\Rightarrow$ CS
Basis	∂_μ	e_I	$\partial_\mu = e^I_\mu e_I$	$e_I = e^\mu_I \partial_\mu$
Co-basis	dx^μ	ϑ^I	$dx^\mu = e^\mu_I \vartheta^I$	$\vartheta^I = e^I_\mu dx^\mu$
Metric	$g_{\mu\nu}(x), V_\mu = g_{\mu\nu} V^\nu$	$\eta_{IJ}, V_I = \eta_{IJ} V^J$	$g_{\mu\nu} = \eta_{IJ} e^I_\mu e^J_\nu$	$\eta_{IJ} = g_{\mu\nu} e^\mu_I e^\mu_J$
Connection	$\Gamma = \Gamma_\mu dx^\mu$	$\omega = \omega_I \vartheta^I$	$\Gamma = E(\nabla e)$	$\omega = e(\mathrm{DE})$
Covariant derivative	$D(\Gamma) = D = d + [\Gamma, -]$	$D(\omega) = \nabla = d + [\omega, -]$	$D = E \nabla e$	$\nabla = eDE$
Curvature	$R = d\Gamma + \Gamma^2$	$R = d\omega + \omega^2$	$R^\mu{}_\nu = e^\mu_I R^I{}_J e^J_\nu$	$R^I{}_J = e^I_\mu R^\mu{}_\nu e^\nu_J$
Torsion	$T^\mu = D dx^\mu$	$T^I = \nabla \vartheta^I$	$T^\mu = e^\mu_I T^I$	$T^I = e^I_\mu T^\mu$

Some further definitions concern the torsion and the curvature in the tetrad description of a geometry. By solving (7.32) for Γ in terms of the tetrads and the spin connections as $\Gamma(e, \partial e, \omega)$ and plugging this into the Riemann-Cartan expressions (7.20) and (7.21) we arrive at

$$\text{Torsion}: \quad T^I{}_{\mu\nu} = \frac{1}{2}(\nabla_\mu e^I_\nu - \nabla_\nu e^I_\mu) \tag{7.37a}$$

$$\text{Curvature}: \quad \tilde{R}^{IJ}{}_{\mu\nu} = (\partial_\mu \omega^{IJ}{}_\nu + \omega^{KJ}{}_\mu \omega^I{}_{K\nu}) - (\mu \leftrightarrow \nu). \tag{7.37b}$$

(Or, in a matrix notation: $T^I = \nabla \vartheta^I, \; R = d\omega + \omega^2$.) The related Ricci tensor and curvature scalar are

$$\tilde{R}^I_\mu := \tilde{R}^{IJ}{}_{\mu\nu} e^\nu_J \qquad\qquad \tilde{R} := \tilde{R}^I_\mu e^\mu_I.$$

The Riemann-Cartan curvature makes its appearance in the commutator of two covariant derivatives, for instance as

$$[\nabla_\nu, \nabla_\lambda] u^I = \tilde{R}^I{}_{K\nu\lambda} u^K.$$

Finally, we arrive again at a Riemann geometry V_4 if the torsion (and with it the contortion) vanishes. According to (7.33, 7.34) the spin connection can be expressed uniquely by tetrads and their derivatives, and the Riemann curvature can be expressed completely in terms of the tetrads and their first and second derivatives.

The Table (7.1) summarizes the previous results in a coordinate system on the tangent and cotangent bundle as compared to the frame bundle and its dual.

The relations between the connections and between the covariant derivatives in CS and TF hold only for vanishing torsion. Here $\Gamma = E(\nabla e)$ is to be interpreted as a matrix equation; in components $\Gamma^\mu_\nu = e^\mu_I \nabla e^I_\nu$. Furthermore, $D = E \nabla e$ reads explicitly $D^\mu{}_\nu \mathcal{T} = e^\mu_I \nabla (e^I_\nu \mathcal{T})$.

Transformations of Tetrads and Spin Connections

Since the tetrads and the spin connections carry both world- and tangent-space indices, it is important to know in each context how they transform with respect to coordinate transformations (diffeomorphisms D), infinitesimally characterized by ξ^μ, and with respect to Lorentz transformations (L) infinitesimally characterized by λ^{IJ}. Under coordinate transformations,

$$e^I_\mu \stackrel{D}{\rightarrow} \mathcal{K}^\nu_\mu\, e^I_\nu \qquad\qquad \delta_\xi\, e^I_\mu = -\,\xi^\nu_{,\mu}\, e^I_\nu \tag{7.38a}$$

$$\omega^{IJ}{}_\mu \stackrel{D}{\rightarrow} \mathcal{K}^\nu_\mu\, \omega^{IJ}{}_\nu \qquad\qquad \delta_\xi\, \omega^{IJ}{}_\mu = -\xi^\nu_{,\mu}\, \omega^{IJ}{}_\nu \tag{7.38b}$$

with the matrices $\mathcal{K}$ defined in (7.3), while under Lorentz transformations

$$e^I_\mu \stackrel{L}{\rightarrow} \Lambda^I_J\, e^J_\mu \qquad\qquad \delta_\lambda\, e^I_\mu = \lambda^I{}_J\, e^J_\mu \tag{7.39a}$$

$$\omega^I{}_{J\mu} \stackrel{L}{\rightarrow} \Lambda^I{}_N\, \omega^N{}_{M\nu} \Lambda^M{}_J + \Lambda^I{}_N \partial_\mu \Lambda^N{}_J \qquad \delta_\lambda\, \omega^I{}_{J\mu} = \omega^I{}_{K\nu}\lambda^K_J + \omega^K{}_{J\nu}\lambda^I{}_K - \lambda^I{}_{K,\mu}. \tag{7.39b}$$

Comparing the transformation behavior of the spin connection with respect to Lorentz transformations with the transformation of the Γ connection (7.18), you will recognize again that they live in different spaces, namely the affine connections Γ in the manifold and the spin connections ω in the tangent bundle to this manifold.

7.4 Physics in Curved Spacetime

The equivalence principle may be translated into a recipe for carrying "laws of physics" valid in flat spacetime (*vulgo* in the absence of gravitation) to curved spacetime (that is to link up gravitation). This recipe, known as *minimal coupling* scheme, is roughly as follows[16] (1) Take a law of physics, valid in inertial coordinates in flat spacetime, (2) Write it in a coordinate-invariant (tensorial) form (3) Assert that the resulting law remains true in curved spacetime. We saw 'minimal coupling' already appearing in Yang-Mills type gauge theories[17].

7.4.1 Mechanics, Hydrodynamics, Electrodynamics

For the equations in mechanics, hydrodynamics and electrodynamics the minimal coupling scheme amounts to start with the manifest Lorentz-invariant expressions and to make these manifestly covariant with respect to general coordinate transformations by the substitutions

---

[16] Here I use the wording of [78].

[17] This gives rise to the question why nature likes it mini: As simple as it can be, but not too simple.

$$\begin{aligned} \eta_{\mu\nu} &\mapsto g_{\mu\nu} \\ \partial_\mu &\mapsto D_\mu(\{\}) = {}_{;\mu} \\ \int d^4x \ldots &\mapsto \int \sqrt{-g} d^4x \ldots \\ d &\mapsto D \end{aligned} \tag{7.40}$$

Mechanics

A mass point moves in a force field according to the equation

$$m\frac{du^\mu}{d\tau} = f^\mu, \tag{7.41}$$

where $u^\mu = \frac{dx^\mu}{d\tau}$ is the four-velocity and τ is the eigen-time. As described before, the coupling to a gravitational field occurs through the substitutions $du^\mu \to Du^\mu$ and $d\tau \to D\tau = d\tau$, such that

$$\frac{Du^\mu}{D\tau} = \frac{du^\mu}{d\tau} + \Gamma^\mu{}_{\lambda\nu} u^\lambda \frac{dx^\nu}{d\tau} = \frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu{}_{\lambda\nu} \frac{dx^\lambda}{d\tau}\frac{dx^\nu}{d\tau}.$$

Inserted in (7.41), this gives the result

$$m\frac{d^2x^\mu}{d\tau^2} = f^\mu - m\,\Gamma^\mu{}_{\lambda\nu} \frac{dx^\lambda}{d\tau}\frac{dx^\nu}{d\tau}.$$

In case of a vanishing external force f^μ this is none other than the geodesic equation (7.7).

Hydrodynamics

Although hydrodynamics is not counted as belonging to fundamental physics, I will mention it here because in most of the applications of GR (for instance in stellar models or in cosmology), matter sources are modelled by fluids. Fluids are defined in terms of macroscopic quantities like density, pressure, entropy, viscosity, …The simplest models are so-called perfect fluids: fluids characterized by two parameters, namely density ρ and pressure p such that in the rest system of the fluid the energy-momentum tensor is

$$(\check{T}^{\mu\nu}) = \begin{pmatrix} c^2\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}.$$

One may boost this rest-frame *ansatz* in order to obtain the form of the tensor in arbitrary frames. However, there is an argument typical of (special) relativity: To construct this tensor, we have only two objects available, namely the Minkowski metric $\eta^{\mu\nu}$ and the four-velocities u^μ. The energy-momentum tensor must have the generic form $T^{\mu\nu} = \alpha(\rho, p)\eta^{\mu\nu} + \beta(\rho, p)u^\mu u^\nu$ with scalar functions α, β. In going to the rest system $(u^\mu) = (c, \vec{0})$, we need to compare this with $(\check{T}^{\mu\nu}) = \alpha\eta^{\mu\nu} + \beta\eta^{00}$, from which finally results

$$T^{\mu\nu} = (\rho + p/c^2)u^\mu u^\nu - p\eta^{\mu\nu}. \tag{7.42}$$

The conservation of the energy-momentum tensor $\partial_\mu T^{\mu\nu} = 0$ provide the dynamical equations (where now $c = 1$)

$$(\rho + p)u^\mu_{,\mu} + \rho_{,\mu}u^\mu = 0 \tag{7.43a}$$

$$(\rho + p)u^\mu u^\nu_{,\mu} + p_{,\mu}(u^\mu u^\nu - \eta^{\mu\nu}) = 0. \tag{7.43b}$$

If you are seeing these for the first time, they may seem baffling. But as a next step, you might go to the nonrelativistic limit $|v| \ll 1$, $p \ll \rho$ to recognize the previous expressions as the equation of continuity and as Euler's equations for the perfect fluid, *viz.*

$$\frac{\partial\rho}{\partial t} + \vec{\nabla}(\rho\vec{v}) = 0 \qquad\qquad \rho(\frac{\partial}{\partial t} + \vec{v}\cdot\vec{\nabla})\vec{v} = -\vec{\nabla}p.$$

But this is not why I added this part on hydrodynamics; what I wanted to stress is how easy it is to formulate hydrodynamics in a gravitational field: simply replace all derivatives in (7.43) by covariant derivatives and all occurrences of the Minkowski metric by the metric $g_{\mu\nu}$–that's it!

Electrodynamics

The manifest Riemann-covariant form of the Maxwell equations is

$$F^{\mu\nu}{}_{;\nu} = -j^\mu \tag{7.44a}$$

$$F_{\mu\nu;\lambda} + F_{\nu\lambda;\mu} + F_{\lambda\mu;\nu} = 0. \tag{7.44b}$$

The Eq. (7.44b) are solved by the *ansatz*

$$F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu,$$

and thus the field strength tensor is unchanged. The expression $F^{\mu\nu}{}_{;\nu}$ in equations (7.44a) is explicitly $F^{\mu\nu}{}_{;\nu} = F^{\mu\nu}{}_{,\nu} + \left\{ {\nu \atop \nu\rho} \right\} F^{\mu\rho}$. Now, using (7.30) the Maxwell Eq. (7.44a) can be written in terms of the ordinary partial derivative as

$$(\sqrt{-g}F^{\mu\nu})_{,\nu} = -\sqrt{-g}\, j^\mu$$

and the continuity equation becomes

$$j^\mu{}_{;\mu} = 0 \qquad \Leftrightarrow \qquad (\sqrt{-g}j^\mu)_{,\mu} = 0.$$

The Maxwell equations for the electromagnetic field coupled to gravity follow from the Lagrange density

$$\mathcal{L}_{EM}(g, A, \partial A) = \sqrt{-g}\, \mathcal{L}^f_{EM}(\eta \to g, A, \partial A \to DA) = -\frac{1}{4}\sqrt{-g}\, g^{\mu\varrho} g^{\nu\sigma} F_{\mu\nu} F_{\varrho\sigma}, \tag{7.45}$$

where the index f in $\mathcal{L}^f_{EM}$ denotes the Lagrange density in (f)lat Minkowski space.

Here we sense an analogy with the introduction of gauge fields in Yang-Mills theories. There, the gauge fields acted as connections serving to give the covariant derivatives a group invariant behavior. Here we see the Levi-Civita connections Γ in the same role as the gauge group connections. The analogy becomes even more pronounced in the tetrad formulation of gravity and in the calculus of differential forms, as will be shown later in this chapter and in Appendix F.

The preceding considerations can be directly adopted for non-Abelian gauge fields. However, in the next Sect. I describe fields in curved spacetime–alias–minimally-coupled to gravitation, using another notation.

7.4.2 *Coupling Relativistic Fields to Gravity*

Whereas bosonic fields transform as tensors with respect to general coordinate transformations, fermionic fields do not exhibit a specific transformation behavior. However, all fields transform in a definite way with respect to the Lorentz group. In order to incorporate spinors and to couple these minimally to a gravitational fields it is thus more appropriate to use the tetrad formalism instead of the metric approach. As a matter of fact, the recipe for minimal coupling becomes more comprehensible, since one recognizes already by the notation the objects in flat and in curved space. The minimal-coupling recipe means replacing

$$\begin{aligned} Q^I &\mapsto Q^\mu = e^\mu_I Q^I \\ \partial_I &\mapsto e^\mu_I D_\mu(\Omega) = e^\mu_I (\partial_\mu + \Omega_\mu) \\ \int d^4x \ldots &\mapsto \int e\, d^4x \ldots \end{aligned} \tag{7.46}$$

where $e := \det(e^I_\mu) \equiv \sqrt{-g}$ and Ω_μ is the Fock-Ivanenko connection

$$\Omega_\mu(x) := \frac{1}{2}\omega^{IJ}{}_\mu \Sigma_{IJ} \tag{7.47}$$

with $\Sigma_{IJ} = e^\mu_I e^\nu_J\, \Sigma_{\mu\nu}$ where $\Sigma_{\mu\nu}$ is the spin matrix.

Scalar Field

For a spin-0 field $\varphi(x)$ start with the Klein-Gordon equation in flat space

$$\mathcal{L}^f_{KG} = \frac{1}{2}(\eta^{IJ}\partial_I\varphi\,\partial_J\varphi - m^2\varphi^2),$$

which using the minimal-coupling recipe becomes

$$\mathcal{L}_{KG} = \frac{1}{2}e\left(\eta^{IJ}e^\mu_I(D_\mu\varphi)e^\nu_J(D_\nu\varphi) - m^2\varphi^2\right),$$

since for a scalar field $\Sigma_{IJ} = 0$, the D_μ derivatives are identical to the partial derivatives ∂_μ. Thus, the Klein-Gordon Langrangian in curved space can be written as

$$\mathcal{L}_{KG} = \frac{1}{2}\sqrt{-g}(g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - m^2\varphi^2).$$

This expression is the same as one would have obtained by using the "metric" recipe (7.40). And this is not a surprise, since the field $\varphi(x)$ is a tensorial object with respect to general coordinate transformations. The Euler derivatives of this Lagrangian are

$$[\mathcal{L}_{KG}]^\varphi = -\sqrt{-g}\,m^2\varphi - \partial_\mu\left(\sqrt{-g}\,g^{\mu\nu}\partial_\nu\varphi\right)$$

which due to the relation $\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\right) = -\sqrt{-g}g^{\mu\sigma}\Gamma^\nu{}_{\mu\sigma}$ lead to the field equations

$$(g^{\mu\nu}\partial_\mu\partial_\nu - g^{\mu\sigma}\Gamma^\nu{}_{\mu\sigma}\partial_\nu + m^2)\varphi(x) = 0.$$

These can also be written as

$$(D_\mu D^\mu + m^2)\varphi = 0$$

which of course is simply the generally covariant Klein-Gordon equation.

Just to give an example of a non-minimal coupling: In four dimensions $\mathcal{L}_{KG}$ may contain a term $\zeta R\varphi^2$ (R being the curvature scalar) with a dimensionless coupling ζ.

Spinor Field

Since the Lorentz group acts on the tangent-space indices, it is possible to define the covariant derivative of a spinor ψ as

$$D_\mu\psi = \left(\partial_\mu - \frac{i}{2}\Sigma^{IJ}\omega_{\mu IJ}\right)\psi \tag{7.48}$$

with $\Sigma^{IJ} := \frac{1}{2}[\gamma^I, \gamma^J]$, where $\gamma^I := \gamma^\mu e^I_\mu$. The Minkowski-space Lagrange density for a Dirac particle

$$\mathcal{L}^f_D = i\overline{\psi}\gamma^I\partial_I\psi - m\overline{\psi}\psi$$

becomes with the minimal-coupling recipe

$$\mathcal{L}_D = e(i\overline{\psi}\gamma^I e_I^\mu D_\mu \psi - m\overline{\psi}\psi) \tag{7.49}$$

in curved space. The field equations in curved space are

$$(i\gamma^\mu D_\mu - m)\psi = 0.$$

Gauge Vector Field

We might consider a Maxwell field again, since any group index from a non-Abelian gauge group is inert to the minimal-coupling recipe. The spin matrix in the Fock-Ivanenko connection (7.47) is

$$(\Sigma_{IJ})_K^{\ L} = \delta_I^L \eta_{JK} - \delta_J^L \eta_{IK}$$

corresponding to the $(\frac{1}{2}, \frac{1}{2})$ representation of the Lorentz group. Applying again the minimal coupling recipe yields initially

$$\begin{aligned} \partial_I A_K \quad \mapsto \quad e_I^\mu(\partial_\mu + \Omega_\mu)A_K &= e_I^\mu(\partial_\mu A_K + \omega^L{}_{K\mu} A_L) \\ &= e_I^\mu e_K^\nu \partial_\mu A_\nu + e_I^\mu[(\partial_\mu e_K^\nu) + \omega^L{}_{K\mu} e_L^\nu] A_\nu. \end{aligned}$$

From this–due to (7.32)–$F_{IK} = e_I^\mu e_K^\nu(\partial_\mu A_\nu - \partial_\nu A_\mu)$, so that

$$\mathcal{L}_{EM}^f = -\frac{1}{4}\eta^{IK}\eta^{JL}\, F_{KL}F_{IJ} \to -\frac{1}{4}\sqrt{-g}\, g^{\mu\varrho}g^{\nu\sigma}\, F_{\mu\nu}F_{\varrho\sigma},$$

and (7.45) is regained.

7.5 Geometrodynamics

The coupling of gravitation to matter was dealt with in the previous section. This section is about the dynamics of the gravitational field itself[18].

7.5.1 Field Equations

Einstein arrived at the field equations of general relativity on a meander-rich path. He very early considered the metric of a spacetime as being the appropriate field

[18] I have chosen to call this 'geometrodynamics', a term originally coined by J.A. Wheeler; however for an ambitious program geared to quantize GR together with the other fundamental interactions.

describing gravitation. With his autodidactic acquaintance of Riemann geometry and by assistance of M. Grossmann, it became clear to him that the field equations he was looking for must be expressed with the aid of the Riemann curvature tensor.

Deriving the Field Equations "The Easy Way"

Today in many textbooks, you can find the field equations "derived" straightforwardly from requiring that the equations (a) must have a tensorial form; (b) should be at most of second order; and (c) must approximate Newton's theory of gravity in the Newtonian limit.

Previously, we saw (compare (7.16)) that in the Newtonian limit the g_{00} component of the metric is related to the gravitational potential ϕ by

$$g_{00} \cong (1 + 2\frac{\phi}{c^2}).$$

where I re-introduced the c^2 in order to later get the correct dimensions of the gravitational coupling constant in the Lagrangian. Now with ϕ obeying the Laplace equation $\triangle\phi = 4\pi G\,\rho(\vec{x})$ we would like this to be part of a system of tensor equations. The left-hand side is proportional to $\triangle g_{00}$, which is a term in a second rank tensor. The mass density must be related to a term in an energy-momentum tensor $T_{\mu\nu}$ by

$$\triangle g_{00} = \kappa T_{00} \qquad\qquad \kappa = \frac{8\pi G}{c^2}.$$

Thus we are looking for a tensorial equation

$$E_{\mu\nu} = \kappa T_{\mu\nu}$$

with $E_{00} \cong \nabla^2 g_{00}$, and where the tensor $E_{\mu\nu}$ (i) is symmetric,[19] (ii) depends at most on second derivatives of the metric, and (iii) is covariantly conserved (since $T_{\mu\nu}$ is covariantly conserved). Now there is a theorem by Lovelock [350] which states that in four dimensions, there is only one tensor which contains at most second derivatives in the metric, and this is the Riemann curvature tensor. The most general two-component tensor that can be built from the Riemann curvature is therefore

$$E_{\mu\nu} = aR_{\mu\nu} + bRg_{\mu\nu} + cg_{\mu\nu}$$

with coefficients (a, b, c), to be determined. This tensor $E_{\mu\nu}$ already obeys the requirement of symmetry. Requiring that it is covariantly constant leads to

$$E^{\mu}_{\ \nu;\mu} \overset{!}{=} 0 = aR^{\mu}_{\ \nu;\mu} + bR_{;\nu}.$$

[19] This is merely an assumption, maybe put forward by Einstein (see pages 48 and 49 in [162]) in observing that the energy-momentum tensors for electrodynamics and for hydrodynamis are symmetric.

By contracting the Bianchi identity (7.28) twice one finds $R^{\mu}_{\ \nu;\mu} = \frac{1}{2}R_{;\nu}$, and the previous requirement reads $0 \stackrel{!}{=} (\frac{1}{2}a + b)R_{;\nu}$. Here, the solution $R_{;\nu} = 0$ is to be excluded, because due to

$$E^{\mu}_{\ \mu} = aR + 4bR + 4c = \kappa T^{\mu}_{\ \mu}$$
$$E^{\mu}_{\ \mu;\nu} = (a + 4b)R_{;\nu} = \kappa T^{\mu}_{\ \mu;\nu},$$

$R_{;\nu} = 0$ would enforce $T^{\mu}_{\ \mu;\nu} = 0$, which is not true in general. Thus $b = -\frac{1}{2}a$, and

$$E_{\mu\nu} = a(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}) + cg_{\mu\nu} = \kappa T_{\mu\nu}. \tag{7.50}$$

There is still the requirement for the correct Newtonian limit:

$$E_{00} = a(R_{00} - \frac{1}{2}Rg_{00}) + cg_{00} \stackrel{!}{\cong} \nabla^2 g_{00}.$$

Here, we could calculate the weak field expressions for the Ricci tensor and the curvature scalar, and then proceed. Calculations are eased by first taking the trace of (7.50) which is

$$a(R - 2R) + 4c = \kappa T \qquad \text{or} \qquad aR = -\kappa T + 4c,$$

so that

$$aR_{00} = \kappa(T_{00} - \frac{1}{2}Tg_{00}) + cg_{00}.$$

In the Newtonian limit: $|T_{ik}| \ll T_{00}$, $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ $|h_{\mu\nu}| \ll 1$. Hence

$$T = \eta^{\mu\nu}T_{\mu\nu} \cong T_{00}, \qquad T_{00} - \frac{1}{2}Tg_{00} \cong \frac{1}{2}T_{00};$$

therefore

$$aR_{00} \cong \kappa\frac{1}{2}T_{00} + cg_{00}. \tag{7.51}$$

Furthermore, R_{00} is given by

$$R_{00} = R^{\mu}_{\ 0\mu0} = R^{k}_{\ 0k0} = \partial_k\Gamma^{k}_{\ 00} - \partial_0\Gamma^{k}_{\ k0} + \Gamma^{k}_{\ k\mu}\Gamma^{\mu}_{\ 00} - \Gamma^{k}_{\ 0\mu}\Gamma^{\mu}_{\ k0}.$$

Again, in the Newtonian limit the terms quadratic in the connections drop out since they are of second order. Further the second term in R_{00} is zero because it is a time-derivative. Thus $R_{00} \cong \partial_k\Gamma^{k}_{\ 00}$. In using (7.15) (with the understanding that in the present context the connection is identical to the Christoffel symbol) $R_{00} \approx -\frac{1}{2}\triangle h_{00}$. Inserting this into (7.51), the comparison yields $a = -1, c = 0$, and therefore $E_{\mu\nu}$ is none other than the negative of the Einstein tensor (7.23). Einstein's field equations are

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -\kappa\, T_{\mu\nu}. \tag{7.52}$$

These equations were presented by Einstein in Berlin on Nov. 25th, 1915 at the end of ten years's research [152]: "*Damit ist endlich die allgemeine Relativitätstheorie als logisches Gebäude abgeschlossen.*"[20]

Einstein's field equations relate the geometry of a region of spacetime, encoded by the Einstein tensor on the left-hand side to the energy density in this region–expressed by the energy-momentum tensor on the right-hand side. The field equations exhibit the generic structure

$$\text{derivatives of fields} = \text{source term},$$

a structure that also holds for Maxwell and Yang-Mills theories. They also display what was phrased by J.A. Wheeler as: "Matter tells space how to curve, ..." in the sense that the energy density determines the geometry, and that "... space tells matter how to move"[21], where he refers to the equation of geodesics (7.7). And indeed, most of the applications of GR to astrophysics and cosmology start from plugging in the energy-momentum density (for instance in the form of the hydrodynamic EM-tensor (7.42)) into the right-hand side of (7.52) and aim to deduce the associated geometry. Exact solutions, however, are rare because of the complicated structure of Einstein's field equations. And the situation is even more obscure because of an inherit 'gauge freedom', or–less ambiguous–a freedom to choose different coordinates to describe the solutions.

Coordinate Conditions

The Einstein field Eq. (7.52) are 10 nonlinear partial differential equations with 10 unknown variables, namely the independent tensor components of $g_{\mu\nu}$. At first glance this seems to be satisfactory, but the number of independent differential equations is not the same as the number of unknowns. This is because of the four contracted Bianchi identities $\nabla_\mu G^{\mu\nu} = 0$ (7.29). Therefore the field equations do not uniquely determine the metric tensor: Together with $g_{\mu\nu}(x)$ also $g'_{\mu\nu}$ is a solution, where $g'_{\mu\nu}(x)$ results from $g_{\mu\nu}(x)$ by a general coordinate transformation. Such a coordinate transformation is characterized by four differentiable functions $x'^\mu(x)$. Therefore only 6 out of 10 differential equations are independent. This is comparable to the gauge freedom in Maxwell's theory. Here, in case of GR, the "gauge freedom" can be broken by imposing four conditions (mostly called "coordinate conditions") imposed on the metric tensor, like e.g.

$$\partial_\nu(\sqrt{-g}g^{\mu\nu}) = 0 \qquad \text{DeDonder gauge} - \text{or} - \text{harmonic coordinates} \tag{7.53a}$$

$$g_{00} = 1 \;\; g_{0i} = 0 \qquad \text{Gauß gauge} - \text{or} - \text{co} - \text{moving coordinates.} \tag{7.53b}$$

Indeed, in both cases coordinate transformations exist by which one can go from arbitrary coordinates to coordinates in which the "gauge conditions" are valid.

[20] "With this, we have finally completed the general theory of relativity as a logical structure."

[21] Written for instance in the introduction of [373]

Initial-Value Problem

The local symmetry in general relativity is the reason why the initial-value problem (more appropriately called the Cauchy problem) of the gravitational field equations becomes very intricate; for a full discussion see Chap. 10 in [526]. Briefly, the issue is as follows: Suppose we have a space-like hypersurface Σ on which the values of $g_{\mu\nu}$ and their first time derivatives are given. The task is to determine from the field equations the gravitational field in the neighborhood of Σ. The problem we face here is that we do not have a sufficient number of equations to determine the second and higher derivatives of the metric. The reason is the structure of the field equations: Write the contracted Bianchi identities (7.29) explicitly as

$$\partial_0 G^{\mu 0} = -\partial_k G^{\mu k} - \Gamma^{\mu}_{\varrho\sigma} G^{\varrho\sigma} - \Gamma^{\varrho}_{\varrho\sigma} G^{\mu\sigma}.$$

Since the right-hand side of this expression includes at most second derivatives with respect to the coordinate time x^0, the $G^{\mu 0}$ on the left-hand side have at most first derivatives with respect to the time variable. Therefore, the subset $G^{\mu 0} = -\kappa T^{\mu 0}$ of the field equations contains no information on the dynamics. These are to be imposed on the initial data as constraints[22]. Only the other six field equations

$$G^{ij} = -\kappa T^{ij} \tag{7.54}$$

are genuine dynamical equations. These determine the second time derivatives of only the six metric components g^{ij}. The second time derivatives of the four $g^{\mu 0}$ remain undetermined.

Again, this ambiguity arises because of the "gauge freedom" of the field equations, and again it can be removed by imposing coordinate conditions. Take for instance the harmonic coordinates (7.53a) and compute a time derivative on these conditions:

$$\partial_0\partial_0 \left(\sqrt{-g}\, g^{\mu 0}\right) = -\partial_i \partial_0 \left(\sqrt{-g}\, g^{\mu i}\right).$$

This relation, together with (7.54), then suffices to determine the second time derivatives of all ten metric tensor components.

7.5.2 Action Functionals for General Relativity

Hilbert Action

Five days before A. Einstein presented his GR field equations in Berlin, D. Hilbert in a seminar in Göttingen wrote down an action for metric gravity

$$S_H = \frac{1}{2\kappa} \int d^4x\, \sqrt{-g}\, R = \frac{1}{2\kappa} \int d^4x\, \mathcal{L}_H(g, \partial g, \partial\partial g) \tag{7.55}$$

[22] More about Constrained Dynamics in Appendix C.

which–as will be derived in the next section–entails as Euler-Lagrange expressions Einstein's field equations in empty space. The factor (2κ) is in this source-free situation irrelevant for the field equations. It gets its proper meaning in case that spacetime regions are present which contain energy densities. Then the factor takes care that the correct Newtonian limit is guaranteed. In $\kappa = 8\pi G$ appears Newton's gravitational constant G. Why it enters the action with $1/G$ becomes intelligible by a dimensional argument. As derived in Sect. 5.4.1 any action must always have (mass)-dimension zero. Now the metric components $g_{\mu\nu}$ do have dimension zero. Thus the curvature tensor containing second derivatives of the metric has dimension $(L)^{-2}$. In order that the Lagrange density $\sqrt{-g}\,R$ in 4D has dimension $(L)^{-4}$, it must be multiplied by a constant with dimension $(L)^{-2}$ or mass dimension $(+2)$. Indeed, Newton's constant G, with dimension $(\text{Length})^3(\text{Mass})^{-1}(\text{Time})^{-2}$ has mass dimension (-2).

Off course, the gravitational action (7.55) can also be written in terms of tetrads and spin connections. But at first let us stay with metric GR.

Euler-Lagrange Equations of Metric General Relativity

The complete Lagrange density of GR consists of two parts: One part $\mathcal{L}_G$ which contains only terms related to the gravitational field in terms of the metric and its derivatives. For this part we may allow aside from the Hilbert action (7.55) a further term containing the *cosmological constant* Λ:

$$S_G = \frac{1}{2\kappa}\int d^4x\,\sqrt{-g}\,(R - 2\Lambda).$$

The other part $\mathcal{L}_M$ describes the coupling of "matter" fields, which are–as before–collectively denoted by Q^α. In metric GR, these are either scalar or vector fields with Lagrangians $\mathcal{L}_Q$ stated in Sect. 7.4. (I write "matter" here with quotation marks, because in the sense of particle physics, the matter fields are fermionic.) The full action of GR is thus

$$S_{GR}[g, Q] = S_G[g] + S_M[g, Q]. \tag{7.56}$$

Now there are two options known to derive the field equations as Euler-Lagrange expressions, called second-order and first-order (or Palatini[23] variational) formalism. In the second-order approach, the Euler derivative is taken with respect to the metric, that is the curvature scalar is considered a function of the metric and its first and second derivatives: $R(g, \Gamma(g), \partial\Gamma(g))$. Instead, in the first-order approach, the metric and the connections are considered to be independent: $R(g, \Gamma, \partial\Gamma)$. Off course the number of Euler-Lagrange equations is different in both cases: There are 10 in the second-order formalism and $10 + 40$ in the first-order one.

[23] As W. Pauli states in his classical text [406], that this formalism was first investigated by Einstein [158].

In the following, let us work in the second-order formalism and take as independent variables the components $g^{\mu\nu}$ of the inverse metric. Remarks on the first-order treatment are made afterwards. The field equations derived from (7.56) have the generic form

$$[\mathcal{L}]_{\mu\nu} := \frac{\delta S_{GR}}{\delta g^{\mu\nu}} = \frac{\delta S_G}{\delta g^{\mu\nu}} + \frac{\delta S_M}{\delta g^{\mu\nu}} = 0 \tag{7.57a}$$

$$[\mathcal{L}]_{\alpha} := \frac{\delta S_{GR}}{\delta Q^{\alpha}} = \frac{\delta S_M}{\delta Q^{\alpha}} = 0. \tag{7.57b}$$

The latter equations for the non-gravitational fields are simply those derived from the Lagrangian with minimal coupling to gravity (as exemplified in Sect. 7.4).

Let us now deduce the explicit gravitational field equations. Instead of painfully determining the Euler derivatives, it is advantageous to argue directly from the variation of the action (7.56). We have initially:

$$\begin{aligned}
\delta S_G &= \frac{1}{2\kappa}\int d^4x\,\delta(\sqrt{-g}(R-2\Lambda)) \\
&= \frac{1}{2\kappa}\int d^4x\,\delta(\sqrt{-g})(R-2\Lambda) + \frac{1}{2\kappa}\int d^4x\sqrt{-g}\,\delta(g^{\mu\nu}R_{\mu\nu}) \\
&= \frac{1}{2\kappa}\int d^4x\,\delta(\sqrt{-g})(R-2\Lambda) + \frac{1}{2\kappa}\int d^4x\sqrt{-g}(\delta g^{\mu\nu})R_{\mu\nu} \\
&\qquad + \frac{1}{2\kappa}\int d^4x\sqrt{-g}\,g^{\mu\nu}\delta R_{\mu\nu}.
\end{aligned}$$

In order to find expressions for $\delta\sqrt{-g}$ and $\delta g^{\mu\nu}$ in the first and second integral, respectively, one can make use of the so-called Jacobi formula, $\ln\det M = \mathrm{tr}(\ln M)$, which holds for arbitrary matrices. Its differential can be rewritten as

$$d\det M = \det M\,\mathrm{tr}\,(M^{-1}dM).$$

Therefore

$$\delta g = -g g_{\mu\nu}\delta g^{\mu\nu} \qquad\qquad \delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}\,g_{\mu\nu}\delta g^{\mu\nu}. \tag{7.58}$$

Observe that $\delta g^{\mu\nu} = -g^{\mu\rho}g^{\nu\sigma}\delta g_{\rho\sigma}$. With these expressions, the variation of S_G becomes

$$\begin{aligned}
\delta S_G = &-\frac{1}{2\kappa}\int d^4x\sqrt{-g}\{\frac{1}{2}(R-2\Lambda)g_{\mu\nu} - R_{\mu\nu}\}\delta g^{\mu\nu} \\
&+ \frac{1}{2\kappa}\int d^4x(\sqrt{-g})g^{\mu\nu}\delta R_{\mu\nu}.
\end{aligned} \tag{7.59}$$

The variation of the Ricci tensor with respect to the metric could in principle be found by first varying the connection with respect to the metric and then inserting this into $\delta R_{\mu\nu}$. Another approach is as follows: Since the variation of the connection is the difference of two connections, it is itself a tensor, and as such it makes sense to take its covariant derivative:

$$D_\lambda(\delta\Gamma^\rho{}_{\nu\mu}) = \partial_\lambda(\delta\Gamma^\rho{}_{\nu\mu}) + \Gamma^\rho{}_{\lambda\sigma}(\delta\Gamma^\sigma{}_{\nu\mu}) - \Gamma^\sigma{}_{\lambda\nu}(\delta\Gamma^\rho{}_{\sigma\mu}) - \Gamma^\sigma{}_{\lambda\mu}(\delta\Gamma^\rho{}_{\nu\sigma}).$$

With some tedious but straightforward calculation one can then derive

$$\delta R^\rho{}_{\mu\lambda\nu} = D_\lambda(\delta\Gamma^\rho{}_{\mu\nu}) - D_\nu(\delta\Gamma^\rho{}_{\lambda\mu}).$$

Therefore,

$$g^{\mu\nu}\delta R_{\mu\nu} = D_\lambda\left[g^{\mu\nu}(\delta\Gamma^\lambda{}_{\nu\mu} - \delta^\lambda_\nu\delta\Gamma^\sigma{}_{\sigma\mu})\right] := D_\lambda\delta w^\lambda, \tag{7.60}$$

and because of (7.30c) the integrand in the second term of (7.59) is a divergence:

$$\sqrt{-g}\, g^{\mu\nu}\delta R_{\mu\nu} = \partial_\lambda(\sqrt{-g}\,\delta w^\lambda). \tag{7.61}$$

Thus the variation of the gravitational part of the action is

$$\delta S_G = -\frac{1}{2\kappa}\int d^4x\{\frac{1}{2}(R-2\Lambda)g_{\mu\nu} - R_{\mu\nu}\}\delta g^{\mu\nu} + \frac{1}{2\kappa}\int d^4x\,\partial_\lambda(\sqrt{-g}\delta w^\lambda). \tag{7.62}$$

For an appropriate behavior of the integrand at the boundary, the last term vanishes. If this is not the case, the original action (7.56) must be modified by a term compensating the non-vanishing surface term from the previous variation; more about this a little later. For the moment, we have derived

$$\frac{\delta S_G}{\delta g^{\mu\nu}} = -\frac{1}{2\kappa}\sqrt{-g}\{\frac{1}{2}(R-2\Lambda)g_{\mu\nu} - R_{\mu\nu}\} = \frac{1}{2\kappa}\sqrt{-g}(G_{\mu\nu} + \Lambda g_{\mu\nu}) \tag{7.63}$$

with the Einstein tensor $G_{\mu\nu}$.

Energy-Momentum Tensors: Hilbert …, Canonical …, Belinfante …

The Eq. (7.57a) rewritten with the previous result as

$$[\mathcal{L}]_{\mu\nu} = \frac{1}{2\kappa}\sqrt{-g}(G_{\mu\nu} + \Lambda g_{\mu\nu}) + \frac{\delta S_M}{\delta g^{\mu\nu}} = 0$$

turn into Einstein's field Eq. (7.52)–aside from the term with the cosmological constant, if the tensor

$$\overset{g}{T}_{\mu\nu} := \frac{2}{\sqrt{-g}}\frac{\delta S_M}{\delta g^{\mu\nu}} \tag{7.64}$$

is identically equal to a (matter) energy-momentum tensor, $T_{\mu\nu}$. Remember that in the context of Noether's theorems a 'canonical' energy-momentum tensor makes its appearance. Denoting the flat-space matter Lagrangian by $\mathcal{L}_M^f$ the canonical tensor is defined as

$$\Theta^{\mu\nu} = \frac{\partial \mathcal{L}_M^f}{\partial_\mu Q^\alpha} \partial^\nu Q^\alpha - \eta^{\mu\nu} \mathcal{L}_M^f$$

that is, it is specified in terms of derivatives with respect to the matter fields Q. On the other hand the tensor (7.64) is defined in terms of derivatives with respect to the metric. Thus, it is far from trivial that $\overset{g}{T}{}^{\mu\nu}$ is at all related to $\Theta^{\mu\nu}$. Amazingly enough, the tensor defined in terms of the metric is directly related to the Belinfante tensor, which as we saw in Sect. 3.3.2. is a relative of the canonical energy momentum tensor: If one defines the Hilbert tensor by

$$T_H^{\mu\nu} := \overset{g}{T}{}^{\mu\nu} \, (g_{\rho\sigma} \to \eta_{\rho\sigma})$$

then the Belinfante energy-momentum tensor is

$$T_B^{\mu\nu} = T_H^{\mu\nu} - [\mathcal{L}_M^f]_\alpha \mathcal{B}^{\alpha\mu\nu}. \tag{7.65}$$

Both tensors are identical on the solutions $[\mathcal{L}_M^f]_\alpha = 0$ of the flat matter field equations. The coefficients $\mathcal{B}^{\alpha\mu\nu}$ are to be read off from the transformation properties of the fields with respect to diffeomorphisms:

$$\delta_\xi Q^\alpha = \mathcal{A}_\mu^\alpha \xi^\mu + \mathcal{B}_\mu^{\alpha\nu} \partial_\nu \xi^\mu.$$

This is an astounding observation: gravity serves to define the "correct" energy-momentum tensor in flat spacetime. The original proof is due to L. Rosenfeld [450]. For a proof in a modern stance see the more accessible [420]; in Appendix F.2.3 the relation between the canonical and the Hilbert energy-momentum tensor is derived in terms of differential forms.

Metric GR in First-Order Form

As mentioned before, the Einstein-Palatini formalism starts from the action

$$S_{EP} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} \, R_{EP} = \frac{1}{2\kappa} \int d^4x \mathcal{L}_{EP}(g, \Gamma, \partial\Gamma).$$

The notation R_{EP} is introduced to indicate that the curvature scalar is built from an arbitrary connection Γ and not necessarily from the Christoffel connection. It is not at all self-evident that the field equations

$$\frac{\delta S_{EP}}{\delta g^{\mu\nu}} = 0 \qquad\qquad \frac{\delta S_{EP}}{\delta \Gamma^\lambda{}_{\mu\nu}} = 0$$

are equivalent to the field equations derived from the Hilbert action. Indeed, in many textbooks on GR you find a proof of equivalence under the assumption that the connection has no torsion. Only recently it was demonstrated that no such assumption is needed [108]. I particularly address these findings here, since symmetries are once again in play: Obviously a generic connection can be split as

$$\Gamma^{\rho}_{\mu\nu} = \left\{ {\rho \atop \mu\nu} \right\} + C^{\rho}{}_{\mu\nu}.$$

Then the Einstein-Palatini Lagrangian can be written as

$$\begin{aligned}\mathcal{L}_{EP} &= \sqrt{-g}\, g^{\mu\nu}(R_{\mu\nu} - C^{\rho}{}_{\rho\nu;\mu} + C^{\rho}{}_{\mu\nu;\rho} + C^{\sigma}{}_{\mu\nu}C^{\rho}{}_{\rho\sigma} - C^{\sigma}{}_{\rho\nu}C^{\rho}{}_{\mu\sigma}) \\ &= \mathcal{L}_H + \sqrt{-g}\, g^{\mu\nu}(C^{\sigma}{}_{\mu\nu}C^{\rho}{}_{\rho\sigma} - C^{\sigma}{}_{\rho\nu}C^{\rho}{}_{\mu\sigma}) + \text{divergence}.\end{aligned}$$

Here, $R_{\mu\nu}$ is the Ricci tensor for the Christoffel connection, and the covariant derivatives are also built with the Christoffel connection. The field equations for the $C^{\rho}{}_{\mu\nu}$ are mere algebraic equations

$$g_{\mu\nu}C^{\rho}{}_{\rho\sigma} + g_{\mu\sigma}C_{\nu}{}^{\rho}{}_{\rho} - C_{\mu\nu\sigma} - C_{\nu\sigma\mu} = 0.$$

It is shown in [108] that the most general solution is

$$C^{\rho}{}_{\mu\nu} = \delta^{\rho}_{\nu} U_{\mu}$$

with an arbitrary vector U_{μ}. That is, if this vector is chosen to be zero, the connection Γ dynamically becomes the Christoffel connection. The torsion $T^{\rho}_{\mu\nu} = C^{\rho}{}_{\mu\nu}$-$C^{\rho}{}_{\nu\mu}$ becomes

$$T^{\rho}_{\mu\nu} = \delta^{\rho}_{\nu} U_{\mu} - \delta^{\rho}_{\mu} U_{\nu}$$

and its trace is given by $U_{\mu} = \frac{1}{D-1} T^{\nu}{}_{\mu\nu}$. Thus it is not the vanishing of the torsion tensor, but the vanishing oft its trace that is required for establishing the equivalence of the first- and second-order variation for GR.

The appearance of the arbitrary vector U_{μ} signalizes the existence of a gauge symmetry in the Einstein-Palatini action

$$g_{\mu\nu} \ \to \ g_{\mu\nu} \qquad\qquad \Gamma^{\rho}_{\mu\nu} \ \to \ \Gamma^{\rho}_{\mu\nu} + \delta^{\rho}_{\nu} U_{\mu}$$

(called by the authors of [108] a projective or a $\mathbb{R}^D$ symmetry) which is not present in the Hilbert action. Therefore the "usual" choice with $U_{\mu} = 0$ is to be interpreted as a gauge condition.

Later in this chapter, modifications and extensions of/to general relativity will be dealt with. Among these are Lanczos-Lovelock theories (see Sect. 7.6). There is a proof that–assuming symmetry of the connection–the equivalence of first- and second order formalisms is a feature of just this type of theories [168]; for a generalization to arbitrary connections see [107]. And in four dimensions the Einstein-Hilbert action is the only Lanczos-Lovelock theory. The equivalence no longer holds in the presence of fermions, where the gravitational interaction must be described in terms of tetrads and spin-connections.

A Short Remark About Vacuum Geometries

But prior to this subject, a digression on the vacuum field equations,

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 0$$

is in order. Contracting these with $g^{\mu\nu}$ we get in D dimensions

$$R = \frac{2D}{D-2} \Lambda .$$

For vanishing cosmological constant thus $R = 0$ and the vacuum field equations become "simply" $R_{\mu\nu} = 0$. These innocent-looking equations[24] should not be underestimated: Only few exact solutions are known. One of them is obviously the Minkowski metric $\eta_{\mu\nu}$. The Minkowski metric is the solution with the maximal number of symmetries, that is the geometry with the maximal number $D(D+1)/2$ of independent Killing vectors; see also Appendix E.5.4. Other known exact solutions are less symmetric. The most important describe the metric outside spherical symmetric gravitational sources like the Schwarzschild geometry, or geometries outside charged and/or rotating black holes. Since this is not a book on macroscopic gravity I will not dwell upon these solutions. Instead let me come back to maximally-symmetric geometries for which the curvature tensor according to (E.24) can be expressed solely by the metric and the curvature scalar as

$$R_{\mu\nu\rho\sigma} = \frac{R}{D(D-1)} \{ g_{\mu\rho} g_{\nu\sigma} - g_{\mu\sigma} g_{\nu\rho} \} .$$

Obviously for the Minkowski metric, $R \equiv 0$. The case $R > 0$ characterizes the de Sitter geometry, $R < 0$ the anti-de Sitter geometry. Indeed from the de Sitter metric in "natural" coordinates (3.110), one calculates

$$R^{\mathrm{dS}}_{\mu\nu\rho\sigma} = \frac{1}{\mathcal{R}^2} \{ \eta_{\mu\rho} \eta_{\nu\sigma} - \eta_{\mu\sigma} \eta_{\nu\rho} \} .$$

In taking together the three previous relations, we find

$$\frac{1}{\mathcal{R}^2} = R \frac{1}{D(D-1)} = \frac{2D}{D(D-1)(D-2)} \Lambda .$$

---

[24] There is a photograph of Einstein taken on his visit at CalTech in 1931 where he is seen writing it on a blackboard. This became a widespread postcard.

Thus the cosmological constant and the de Sitter radius are interrelated by a numerical factor depending on the dimension of spacetime. For $D = 4$ this relation is $\Lambda = 3/\mathcal{R}^2$. Be aware that the Minkowski metric is not a solution of the vacuum field equations with a non-vanishing cosmological constant. Thus for $\Lambda \neq 0$ the equivalence principle in (one of) its original formulations is no longer valid. As tiny as the cosmological constant might be, it calls in principle for an extension of special relativity.

But how tiny is Λ? It is evidently small enough to have escaped observation in astronomy. With the observation in the 1990's of an accelerated expanding universe, one estimates within the concordance model of cosmology that $\Lambda = 10^{-52}m^{-2}$ in metrical units, or in natural units $\Lambda = 10^{-122}$.

Lagrange Formulation of Tetrad Gravity

The Hilbert action (7.55) as expressed in terms of tetrads reads:

$$\begin{aligned} S_H &= \frac{1}{2\kappa}\int d^4x\ eR = \frac{1}{2\kappa}\int d^4x\ eR^{IJ}{}_{\mu\nu}\, e^{\mu}_{I} e^{\nu}_{J} \\ &= \frac{1}{8\kappa}\int d^4x\ \epsilon^{\mu\nu\varrho\sigma}\,\epsilon_{IJKL}\, e^{I}_{\mu} e^{J}_{\nu}\, R^{KL}{}_{\varrho\sigma}, \end{aligned} \tag{7.66}$$

where the latter relation comes about because the determinant of the vierbeins can be written as

$$e = 4\,\epsilon^{\mu\nu\rho\sigma}\,\epsilon_{IJKL}\, e^{I}_{\mu} e^{J}_{\nu} e^{K}_{\rho}\, e^{L}_{\sigma}.$$

For the full theory, the action is augmented by the matter Lagrangians as they were deduced in Sect 7.4.2. In deriving the field equations, we are again free to employ either first- or second-order variation. At least as long as fermionic fields are absent, both variants lead to the same result. In the case of second order, the Lagrangian (7.66) is considered a function of the tetrads and its derivatives, that is the curvature tensor is understood as depending on $\omega[e, \partial e, \partial\partial e]$ (as in the Ricci rotation coefficients (7.34)). In this case there are sixteen field equations $([\mathcal{L}]^{\mu}_{I} = 0)$ in contrast to ten equations $([\mathcal{L}]_{\mu\nu} = 0)$ in second order metric gravity. The difference of six equations is again due to the additional Lorentz-invariance of tetrad gravity. If the tetrads and the spin connections are taken to be independent, that is $R = R(e, \omega, \partial\omega)$, one obtains from (7.66) aside from 16 field equations $\frac{\delta S_H}{\delta e^{\sigma}_{L}} = 0$ further 24 field equations $\frac{\delta S_H}{\delta \omega^{KL}_{\sigma}} = 0$, namely

$$\begin{aligned} \frac{\delta S_H}{\delta e} &= 0 \qquad \leftrightarrow \qquad \epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL} e^{J}_{\nu} R^{KL}{}_{\rho\sigma} = 0 \\ \frac{\delta S_H}{\delta \omega} &= 0 \qquad \leftrightarrow \qquad \epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL}\nabla_{\rho}(e^{I}_{\mu} e^{J}_{\nu}) = 0. \end{aligned}$$

The latter equations are equivalent to $\epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL}T^I_{\rho\mu}e^J_\nu = 0$ with the torsion given by (7.37a). If the torsion vanishes these are identically fulfilled, and the first set of equation is synonymous with those from the second-order approach. In [108], it is shown that (entirely in analogy with the case of metric gravity) the equivalence holds true for arbitrary spin-connections up to gauge transformations

$$e^I_\mu \to e^I_\mu \qquad\qquad \omega^I_{J\mu} \;\to\; \omega^I_{J\mu} + \delta^I_J V_\mu.$$

Furthermore, it is shown that the coupling of fermions requires non-zero torsion and metricity. This is for instance the Einstein-Cartan theory of gravity; dealt with in more detail in Sect. 7.6.2.

Optional Forms of Actions for GR

In recent years, essentially triggered by Ashtekar's "new variables" in the Hamiltonian formulation of GR (more details in Appendix C.3.3.), several other actions were investigated. All of these are equivalent in the sense that they yield Einstein's field equations. The relations among these actions (and their relation to the Hilbert action) are derived in full detail in the review [409], including the relations to the ADM and the Ashtekar Hamiltonian.

The selfdual Hilbert-Palatini Lagrangian
In working with tetrads in the first-order (Hilbert-Palatini) context, there is the Lorentz Lie algebra-valued field ω^{IJ}. Instead of its six real components, one may think of using three complex components. This is realized based on the group-theoretical fact that the complex Lorentz-algebra $\mathfrak{so}(3,1;\mathfrak{C})$ splits into two complex $\mathfrak{so}(3;\mathfrak{C})$ algebras as:

$$\mathfrak{so}(3,1;\mathfrak{C}) = \mathfrak{so}(3;\mathfrak{C}) \oplus \mathfrak{so}(3;\mathfrak{C}),$$

where the first component is self-dual and the second anti-self-dual. This means the following: Define for any $\mathfrak{so}(3,1)$ Lie algebra-valued field A^{IJ} its dual by

$$*A^{IJ} := \frac{1}{2}\epsilon^{IJ}{}_{KL}A^{KL}.$$

The dual of this is obeys $**A^{IJ} = -A^{IJ}$ (in the case of Lorentz signature of the metric). Now define

$$A^{\pm IJ} = \frac{1}{2}(A^{IJ} \mp i * A^{IJ})$$

and verify that A^{+IJ} is self-dual, and A^{-IJ} is anti-self-dual. These two types of fields decouple in the sense that

$$[A,B]^{IJ} = [A^+,B^+]^{IJ} + [A^-,B^-]^{IJ}.$$

With this, the self-dual curvature is the curvature of the self-dual spin connection:

$$R^{IJ}{}_{\mu\nu}(\omega^+) = R^{+IJ}{}_{\mu\nu}(\omega^+) = R^{+IJ}{}_{\mu\nu}(\omega).$$

Finally the self-dual Hilbert-Palatini action is found to be

$$S[e^\mu_I, \omega^{+IJ}] = \int d^4x\, \mathcal{L}^+_{HP} = \int d^4x\, e\Sigma^{+\mu\nu}_{\ \ IJ} R^{+IJ}_{\ \ \mu\nu}(\omega), \tag{7.67}$$

where $\Sigma^{\mu\nu}{}_{IJ} := \frac{1}{2} e^\mu_{[I} e^\nu_{J]}$. This action differs from the tetrad Hilbert action (7.66) by a term that does not change the field equations.

The Plebanski Lagrangian

The next manipulations can only be sketched here; for details see [409]: By taking projections of the objects Σ^{+IJ}, R^{+IJ} with respect to a time-like unit vector field N^I and to a space-like surface orthogonal to it, and by defining objects Σ^i and F^i proportional to the projected Σ^{+IJ} and R^{+IJ} it is found that the Lagrangian in (7.67) can be written as

$$\mathcal{L}^+_{HP} = i\epsilon^{\mu\nu\rho\sigma}\Sigma^i_{\mu\nu} F_{i\mu\nu}(A^k_\lambda),$$

where the curvature (or field strength) depends on the projected spin connections only. Thus far these are merely manipulations on $\mathcal{L}^+_{HP}$. To arrive at the Plebanski action the trick is to let the $\Sigma^i_{\mu\nu}$ become independent fields. It was shown in Sect. 3.3.4 that neither the equations of motion nor the symmetries are changed, if the original definitions of these fields are added to the Lagrangian as constraints with multiplier terms. A necessary and sufficient form of constraints can be shown to be (see again [409])

$$0 = \tilde{M}^{ij} = M^{ij} - \frac{1}{3}\delta^{ij} M^k_k \qquad\qquad M^{ij} := \epsilon^{\mu\nu\rho\sigma}\Sigma^i_{\mu\nu}\Sigma^j_{\rho\sigma}\,.$$

The property of $\tilde{M}$ being traceless can be shifted over to the Lagrangian multiplier (ψ_{ij}) with a further multiplier (μ), so that one arrives at the Plebanski action:

$$\begin{aligned} S_P[\Sigma^i_{\mu\nu}, A^i_\mu, \psi_{ij}, \mu] &= \int d^4x\, \mathcal{L}_P \\ &= \int d^4x\, i\,\epsilon^{\mu\nu\rho\sigma}\Sigma^k_{\mu\nu}(F_{k\rho\sigma}(A) + \psi_{kj}\Sigma^j_{\rho\sigma}) + \mu\psi^k_k. \end{aligned} \tag{7.68}$$

The CDJ Lagrangian

The CDJ-Lagrangian (named after its inventors R. Capovilla, J. Dell, and T. Jacobson [71]) was derived from the Plebanski Lagrangian in two steps, both of them amounting to the elimination of auxiliary fields. Remember from Sect. 3.3.4 that this does not change the dynamical content and the symmetries of the theory in question. First, the field Σ can be eliminated by solving its equation of motion

$$\frac{\delta S_P}{\delta \Sigma^k_{\mu\nu}} = i\epsilon^{\mu\nu\rho\sigma}(F_{k\rho\sigma} + 2\psi_{ij}\Sigma^j_{\rho\sigma}) = 0.$$

Inserting the solution $\Sigma^i_{\mu\nu} = -1/2(\psi^{-1})^{ij} F_{i\mu\nu}$ into $\mathcal{L}_P$, it becomes

$$\mathcal{L}_P \;\Rightarrow\; -\frac{i}{4}\epsilon^{\mu\nu\rho\sigma}(\psi^{-1})^{kj} F_{k\mu\nu} F_{j\rho\sigma} + \mu\psi^k_k =: -\frac{i}{4}\mathrm{Tr}(\psi^{-1}\Omega) + \mu\mathrm{Tr}\psi$$

where in the latter matrix notation, $\Omega^{kj} := \epsilon^{\mu\nu\rho\sigma} F^k_{\mu\nu} F^j_{\rho\sigma}$. From the characteristic polynomial for 3×3 matrices one derives

$$\mathrm{Tr}\,\psi = \frac{1}{2}\det\psi \left((\mathrm{Tr}\,\psi^{-1})^2 - \mathrm{Tr}\,\psi^{-2} \right).$$

And inserting this into the previous Lagrangian and next solving the equations of motion for ψ^{-1} allows one to eliminate this as an auxiliary field and finally arrive at the CDJ-action

$$S_{CDJ}[A^i_\mu, \eta] = \int d^4x\, \mathcal{L}_{CDJ} = \int d^4x\, \frac{\eta}{8} \left((\mathrm{Tr}\,\Omega^2 - \frac{1}{2}(\mathrm{Tr}\,\Omega)^2 \right). \tag{7.69}$$

with a redefined scalar η.

Observe the gradual shift from the Hilbert-Einstein action formulated in terms of a metric and its derivatives towards a formulation of the same dynamical theory in terms of spin-connections and a scalar field ("general relativity without the metric"). And notice that the Lagrangian in (7.69) is written in terms of $(F)^2$, a structure reminiscent to Yang-Mills type gauge theories.

Boundary Terms

In recent years it became more and more evident that "physics lurks" in boundary terms of an action–terms which are discarded in deriving the field equations. A theory is not only defined by its field equations but also by boundary conditions on the fields: If we believe in the prominent role of an action defining a theory, boundary terms arising in variations should be taken serious. In his contribution to ADM-50[25], Robert M. Wald concedes in a talk entitled "Lagrangians and Hamiltonians in Classical Field Theory": "... it had been my view that their role in classical field theory was not much more than that of a mnemonic device to remember the field equations.... The existence of a Lagrangian or Hamiltonian provides important auxiliary structure to a classical field theory, which endows the theory with key properties."

A boundary term

1. can serve as the generator of a canonical transformation: This was discussed in Sec. 2.1.4. In field theories one can use this feature in order to transit from one set of canonical variables to another one, an example being the transition from the ADM to the Ashtekar variables in canonical GR (see Sec. C.3.3).

[25] http://adm-50.physics.tamu.edu

2. must be included into the action in order to cancel terms which do not vanish due to non-trivial boundary conditions (York, Gibbons, Hawking); this is discussed in the present section.
3. leads to specific relations between pseudotensors and quasilocal expressions for the energy-momentum of gravitational systems; a topic explained in more detail in the next section on conservation laws.
4. may be necessary in order to allow for the definition of a functional derivative (Regge/Teitelboim).
5. signals a non-trivial topology.
6. may modify the quantum theory, because two versions of a theory which are related by a canonical transformations may not be unitarily equivalent.
7. can be interpreted thermodynamically as the entropy of horizons in gravity; see a short remark at the end of this section and in Appendix F.3.1.

Here are more details on item 2.): In deriving the gravitational field equations from the stationarity of the Hilbert action one finds a surface term according to (7.62)

$$B_H = \int d^4x \sqrt{-g} D_\lambda \delta w^\lambda = \int d^4x \, \partial_\lambda \big[\sqrt{-g} g^{\mu\nu} (\delta \Gamma^\lambda{}_{\nu\mu} - \delta^\lambda_\nu \delta \Gamma^\sigma{}_{\sigma\mu}) \big]$$

that must vanish. The term contains the variations of the connection, that is–in this second-order context–the variation of the metric and its derivatives. Thus setting $\delta g^{\mu\nu} = 0$ on the boundary is not enough. And requiring additionally $\delta(\partial_\lambda g^{\mu\nu}) = 0$ is not only rather *ad hoc*, but might even be too strong, in that this is not compatible with the field equations and/or initial conditions. J.W. York was among the first [575] with an answer to the question "What is fixed on the boundary in the action principles of general relativity?". The fact that in gravitational theories the variations of the derivatives of the metric show up is related to the Lagrangian's containing second derivatives of the metric. Since the field equations are only second order in the metric it is possible to write the action in a form in which the higher derivatives are due to a divergence, see e.g. [333]:

$$\begin{aligned} S_H = S_E + \int_\Omega d^4x \, \partial_\mu W^\mu &= \int_\Omega d^4x \, \mathcal{L}_{\text{bulk}} + \int_\Omega d^4x \, \mathcal{L}_{\text{surface}} \\ &= \int_\Omega d^4x \sqrt{-g} \, g^{\mu\nu} (\Gamma^\lambda{}_{\mu\nu} \Gamma^\sigma{}_{\lambda\sigma} - \Gamma^\lambda{}_{\mu\sigma} \Gamma^\sigma{}_{\lambda\nu}) \\ &\quad + \int_\Omega d^4x \, \partial_\mu \big[\sqrt{-g} \, g_{\lambda\rho,\nu} (g^{\mu\nu} g^{\lambda\rho} - g^{\mu\lambda} g^{\nu\rho}) \big]. \end{aligned} \tag{7.70}$$

The $\Gamma\Gamma$ (or "bulk") term corresponds to the action considered by Einstein in 1916 [153]:

$$S_E = \int_\Omega d^4x \sqrt{-g} \, G^* := \int_\Omega d^4x \sqrt{-g} \, g^{\mu\nu} (\Gamma^\lambda{}_{\mu\nu} \Gamma^\sigma{}_{\lambda\sigma} - \Gamma^\lambda{}_{\mu\sigma} \Gamma^\sigma{}_{\lambda\nu}). \tag{7.71}$$

The Hilbert and the Einstein Lagrangian differ in their behaviour under diffeomorphism:

$$\delta_\xi \mathcal{L}_H = -\partial_\lambda(\xi^\lambda \mathcal{L}_H) \qquad\qquad \delta_\xi \mathcal{L}_E = -\partial_\lambda(\xi^\lambda \mathcal{L}_H + W^\lambda).$$

Only the Hilbert Lagrangian transforms as a scalar density.

Both S_H and S_E entail the same dynamical equations: $[\mathcal{L}_H]^{\mu\nu} = 0 = [\mathcal{L}_E]^{\mu\nu}$, therefore they are–in a notion previously defined for finite-dimensional systems–solution-equivalent. But they are not boundary-equivalent.

The variations of the Hilbert and the Einstein action contain different boundary terms. Write

$$\delta S_H = \int_\Omega \frac{\delta S_H}{\delta g^{\mu\nu}} \delta g^{\mu\nu} + B_H, \quad \delta S_E = \int_\Omega \frac{\delta S_E}{\delta g^{\mu\nu}} \delta g^{\mu\nu} + B_E, \quad \delta S_H = \delta(S_E + S_B),$$

where S_B is the term that arises on going from the Hilbert action to the Einstein action, that is the second term in (7.70). Since

$$\frac{\delta S_H}{\delta g^{\mu\nu}} = \frac{\delta S_E}{\delta g^{\mu\nu}}$$

the boundary terms are related by

$$B_E + \delta S_B = B_H.$$

The aim is to construct a gravitational action S_{Grav} which by $\delta S_{Grav} \stackrel{!}{=} 0$ delivers Einstein's vacuum equations if the fields are fixed on the boundary of the integration domain. So we are looking for an action $S_{Grav} = S_H + S_{HB}$, where S_{HB} is a surface term to be determined. From

$$\delta S_{Grav} = \delta S_H + \delta S_{HB} = \int_\Omega \frac{\delta S_H}{\delta g^{\mu\nu}} \delta g^{\mu\nu} + B_H + \delta S_{HB}$$

a sufficient condition for the additional boundary term is

$$(B_H + \delta S_{HB})|_{\delta g=0} \stackrel{!}{=} 0$$

or, since B_H can be expressed by the boundary term from the Einstein action (and this already obeys $B_E|_{\delta g=0} = 0$):

$$(\delta S_B + \delta S_{HB})|_{\delta g=0} \stackrel{!}{=} 0.$$

This condition does not fix the surface term uniquely because $(\delta S_{EH} + \delta S_{HB})$ may differ from zero away from $\partial\Omega$. As was essentially shown by J. W. York [575], and by G. Gibbons and S. Hawking [216], a possible term for modifying the Hilbert action is

$$S_{HB} = 2 \int_{\partial\Omega} d^3x \, \sqrt{\epsilon h}\, K. \tag{7.72}$$

In this "Gibbons-Hawking-York boundary term", the trace of the extrinsic curvature K of the boundary $\partial\Omega$ and the determinant h of the induced metric on the boundary enter together with $\epsilon = +1$ if the normal to the boundary surface is spacelike, and $\epsilon = -1$ if it is timelike. (For the definition and meaning of induced metric and extrinsic curvature of a hypersurface see Appendix C.3.3) York furthermore demonstrated that it is not even necessary to require $\delta g^{\mu\nu} = 0$ on $\partial\Omega$, but that it suffices to fix the intrinsic three-metric on the boundary. The action $S_H + S_{HB}$ also goes by the name *trace-K action*. The GHY term is widely accepted as the correct modification of the Hilbert-Einstein action, and interestingly it appears at various other places in GR. For instance it is needed when it comes to derive the Hamiltonian equations for GR (the so-called ADM equations; see Appendix C.3.3.). Furthermore it yields a substantial contribution to the entropy of black holes.

Let me here finally mention a "triviality" which however in diffeomorphism invariant theories receives a meaning. Whenever one is dealing with a field theory in terms of a Lagrangian $\mathcal{L}(\phi^a, \partial\phi^a)$ for fields ϕ, a solution-equivalent Lagrangian is

$$\mathcal{L}' = \mathcal{L} - \partial_\lambda \left[\phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\lambda \phi^a)} \right].$$

This is the generalization of (2.24) from classical mechanics. In the case of GR, if we take $\mathcal{L}$ to be the bulk Lagrangian in (7.70), we identify the surface Lagrangian

$$\mathcal{L}_{surface} = -\partial_\lambda \left[g_{\mu\nu} \frac{\partial \mathcal{L}_{bulk}}{\partial(\partial_\lambda g_{\mu\nu})} \right].$$

Neither the bulk nor the surface term transform as scalar densities, but only their sum. This indicates that the surface Lagrangian could be found from the bulk Lagrangian by requiring manifest invariance under diffeomorphism. But, as exhibited by T. Padmanabhan [401], the bulk term can be derived alone from the surface term under the specific assumption that it represents the entropy of black-hole horizons. (This assumption is supported by various approaches to the thermodynamic laws for black holes.) This is a signal of the "holographic" nature of gravitation, and became part of understanding gravitation as thermodynamics of horizons [403]. These results hold true even for a much wider class of gravitational theories including the Lanczos-Lovelock gravity models [375].

7.5.3 Covariance, Invariance, and Symmetries

As mentioned before, in the genesis of general relativity arose a "principle" of general covariance. In the historical perspective two main aspects (which are still a source of confusion today) became important in this context: Einstein himself stumbled around 1913 by his hole argument (in German: "*Lochbetrachtung*"). And in 1917 E. Kretschmann demonstrated, that every physical equation can be rendered generally covariant, and therefore one cannot talk about a "principle". Here I refrain from entering the philosophical discussion about whether or not GR has an underlying "principle" of general covariance, or whether one needs principles at all in establishing or interpreting theories of gravitation; see [482, 395].

At other places in this book, the term "covariance", when used in the context of dynamical equations, was always meant as form invariance with respect to a group of transformations. This group is often termed *covariance group*. Thus the Newtonian equations of motion are covariant with respect to the Galilei transformations, and Maxwell's equations are covariant with respect to Lorentz transformations. The term "general covariance" refers to invariance of a theory under local invertible general coordinate transformations, these constituting the Einstein group (as sometimes called in older the physics literature) or the diffeomorphism group (as it is usually called today). The debate and confusion about general covariance arose from the question of whether GR plays a special role among other physical theories of being distinguished by being invariant under general coordinate transformations.

Recall that a diffeomorphism d is a C^∞ one-to-one mapping of a manifold M (in the context of GR a 4D pseudo-Riemannian spacetime) onto itself. Diffeomorphism covariance means that under the push-forward d_* the field equations for ϕ behave as

$$\mathcal{E}[\phi] = 0 \quad \curvearrowright \quad \mathcal{E}[d_*\phi] = 0.$$

The Einstein field equations $\mathcal{E}[g] = 0$ surely fulfill this criterion–by their very derivation from the Hilbert-Einstein action. But is this unique to GR? And is this the same as invariance under diffeomorphisms?

Take for example Maxwell's vacuum equations in a spacetime M^4 which is fixed. The field equations are:

$$F^{\mu\nu}{}_{;\nu} = 0 \qquad\qquad F_{\mu\nu;\lambda} + F_{\nu\lambda;\mu} + F_{\lambda\mu;\nu} = 0. \tag{7.73}$$

This system of equations is manifestly covariant under $\mathbf{Diff}(\mathbf{M^4})$-transformations. But the diffeomorphisms do not constitute its symmetry group. That is rather the conformal group, as derived in Subsect. 3.5.2. Only if the metric is considered as a dynamical field is the symmetry group of the combined Einstein-Maxwell system the diffeomorphism group. This example reveals that the relation of covariance to invariance hinges on which fields are considered dynamical and which fields are fixed from the outside.

It is admitted that J. L. Anderson (e.g. [7]) was the very first to undertake the endeavor of disentangling the notions of covariance, invariance and equivalence (principles). Central to Anderson's argumentation is the distinction between absolute and dynamical objects. In order to see his point, let me first follow [221] in distinguishing externally-specified background structures Σ and dynamical structures Φ. The field equations

$$\mathcal{E}[\Phi, \Sigma] = 0 \tag{7.74}$$

then define a relation between the two structures. According to the terminology of Anderson, the set of all possible Φ is called the space of *kinematically possible trajectories* $\mathcal{K}$. The field equations (7.74) then determine a subset $\mathcal{D} \subset \mathcal{K}$ of *dynamically possible trajectories* for given Σ.

Now suppose that a group[26] $\mathbf{G}$ acts on $\mathcal{K}$

$$\mathbf{G} \times \mathcal{K} \to \mathcal{K} \qquad\qquad (g, \Phi) \mapsto g \cdot \Phi.$$

The group $\mathbf{G}$ is the group of symmetries for the Eq. (7.74), if it leaves $\mathcal{D}$ invariant, that is:

$$\mathcal{E}[\Phi, \Sigma] = 0 \iff \mathcal{E}[g \cdot \Phi, \Sigma] = 0.$$

This is to be distinguished from covariance for which

$$\mathcal{E}[\Phi, \Sigma] = 0 \iff \mathcal{E}[g \cdot \Phi, g \cdot \Sigma] = 0.$$

These definitions exhibit that an invariance under a symmetry operation entails the covariance of the dynamical equations. Furthermore, covariance becomes an invariance if $g \cdot \Sigma = \Sigma$, that is, if the symmetry group $\mathbf{G}$ stabilizes the background structure.

Take again the example Maxwell's vacuum equations on a fixed background, and make the split into dynamical fields $\Phi = \{F_{\mu\nu}\}$ and kinematical fields $\Sigma = \{g_{\mu\nu}\}$. The stabilizer of the background is found by observing that the field equations depend on the metric through the covariant derivative and the operation of raising indices, which appears in the combination $\sqrt{-g}g^{\mu\rho}g^{\nu\sigma}$. This combination is left invariant under Poincaré transformations, and in four dimensions also under conformal transformations, in accordance with previous findings.

This and other examples give a hint that diffeomorphism covariance can be made trivial. Anderson provided the example of the heat equation

$$\frac{\partial}{\partial t} T(\vec{x}, t) = \kappa \nabla^2 T \tag{7.75}$$

for a temperature field $T(\vec{x}, t)$. This equation does not look as though it is diffeomorphism covariant (and not even Lorentz covariant). But by introducing a normalized constant four-vector field n^μ the same physical system is described by the "theory" on a fixed background g:

$$n^\mu T_{;\mu} = \kappa(n^\mu n^\nu - g^{\mu\nu})T_{;\mu\nu}$$

with $n^\mu \partial_\mu = \partial_t$. Now the equations are manifestly covariant with respect to general coordinate transformations. According to the previous definition, the invariance

[26] In the context of GR, this is the group of diffeomorphism, but the definition applies to any symmetry group.

group of the system is the subgroup of **Diff**(**M**) that stabilizes the background structure $\Sigma = \{n^\mu, g\}$. This is the group $E_3 \times \mathbb{R}$, that is the product of Euclidean spatial motions and time translations. And this is just the symmetry group of (7.75).

We conclude: Any equation that is written in a special coordinate system on a manifold M can be rendered **Diff**(**M**)-covariant by introducing the coordinate system as background structure. This essentially is the core of Kretschmann's critique on founding general relativity on a "principle" of general covariance.

But also invariance can be trivialized if the background is changed into a dynamical structure. For example, both for Maxwell's equations and for the heat equation, we might introduce dynamical structures by the field equations for g and (g, n), respectively:

$$R_{\mu\nu\rho\sigma}(g) = 0 \qquad \text{resp.} \qquad g_{\mu\nu}n^\mu n^\nu = 1 \qquad n^\mu_{;\nu} = 0.$$

This all demonstrates that for distinguishing covariance and invariance the distinction of dynamical and background field is not sharp enough. In order to get around the trivializations sketched above, Anderson introduced *absolute objects*, objects that are independent of dynamical objects and are part of the background. Many authors [198], [221], [489] tried to sharpen and improve Anderson's construction, but here I will not enter this discussion. In any case, it is now understood that the invariance group consists of those elements of $\mathcal{G}$ which leave the absolute objects invariant, and that non-trivial general covariance is present if absolute objects are absent.

In [489], Anderson's conceptualization is illustrated on three diffeomorphism-covariant "theories" for a symmetric tensor field g defined on $\mathbb{R}^4$:

$$\text{(a)} \qquad R_{\mu\nu\rho\sigma}(g) = 0$$
$$\text{(b)} \qquad R(g) = 0 \qquad C_{\mu\nu\rho\sigma}(g) = 0$$
$$\text{(c)} \qquad R_{\mu\nu}(g) = 0.$$

In the "theory" (a) the solution is the Minkowski metric, i.e. each solution of obeys $g = d_*\eta$ with the Minkowski metric η. There are no dynamical degrees of freedom; the Minkowski metric is the absolute object of this "theory". Since it is left invariant by the Poincaré group, this is according to Anderson the invariance group of (a).
"Theory" (b) is called the Einstein-Fokker theory. Here $C_{\mu\nu\rho\sigma}$ is the Weyl tensor, which can be characterized as the unique linear combination of the Riemann curvature tensor, the Ricci tensor and the curvature scalar being invariant under rescalings $g \mapsto \Omega^2 g$. The vanishing of the Weyl tensor implies that the metric is conformally flat: $g = d_*(\Omega^2\eta)$. The equation $R(g) = 0$ implies for the scalar field Ω the field equation $\Box_\eta \Omega = 0$. The invariance group of (b) is the conformal group.
Finally (c) are the field equations of vacuum general relativity. Here the metric is a fully dynamical object; there is no absolute object. Therefore the covariance and invariance groups are identical. In this sense GR is indeed diffeomorphism-invariant. This is a resolution of Kretschmann's objection that general covariance is not the "unique selling point" of general relativity.

These days, the term "general covariance" is hardly used in the scientific community, "background independence" is used instead. A vigorous debate about the proper approach towards quantum gravity is going on by the camps of 'loop quantum gravity (being background independent *per se*) and of string theory (in which it is hoped that ultimately one can get rid of background structures).

Active and Passive Diffeomorphisms

For metric gravity, the action (7.56) is invariant under the transformations[27]

$$x^\mu \to x'^\mu = x'^\mu(x) \tag{7.76a}$$

$$g_{\mu\nu} \to g'_{\mu\nu}(x); \qquad g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x(x')) \tag{7.76b}$$

$$Q_\alpha \to Q'_\alpha(x); \qquad Q'_\alpha(x') = t_\alpha^\beta Q_\beta(x) \tag{7.76c}$$

where t_α^β stands for the tensorial transformations of the fields, e.g. for a scalar field $\varphi'(x') = \varphi(x)$ and for a vector field $A'_\mu(x') = \frac{\partial x^\rho}{\partial x'^\mu} A_\mu(x)$. (Remember that fermionic fields are not treatable in metric gravity.)

Here obviously "invariance" is stated in terms of coordinates. However, the role of coordinates in GR needs to be defined in each context. (In the context of invariance, the most appropriate attitude is to avoid coordinates at all.) Let us consider pure gravity first. The transformation (7.76b) allows for two geometrical interpretations, known as *passive and active diffeomorphism* invariance[28].

A *passive diffeomorphism* is merely a change of coordinates: A specific object–here the metric tensor–is represented in different coordinate systems $\{x\}$ and $\{x'\}$. Then $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x')$ satisfying (7.76b) represent the same metric in M.

An *active diffeomorphism* relates different objects in M in the same coordinate system: Now $\{x'(x)\}$ is viewed as a map associating one point of the manifold to another point. Suppose that for two points $P, P' \in M$, the metric tensors are given by $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$, and suppose that they are both solutions of the field equations. If these two metrics are related by (7.76b) we have the case of active diffeomorphism invariance.

As described before, any theory can be made invariant under passive diffeomorphisms (coordinate changes) by adding a sufficient number of external/background/non-dynamical fields. Today it is understood that passive diffeomorphism invariance is a property of how a theory is formulated, while active diffeomorphism invariance is a property of the theory itself. In this sense general relativistic theories are invariant (in contrast to QED, QCD, or any other theory on a fixed background).

---

[27] The field equations derived from the Hilbert-Einstein action exhibit an additional scale invariance $g_{\mu\nu} \to \lambda^2 g_{\mu\nu}$, under which $R_{\mu\nu} \to R_{\mu\nu}$ and $R \to \lambda^{-2} R$. This is another example of a Lie symmetry not being a Noether symmetry.

[28] A nice example and illustration of the distinction between these two notions of diffeomorphism invariance can be found in [451].

Infinitesimally the transformations (7.76) become

$$\delta_\xi x^\mu = \xi^\mu \tag{7.77a}$$
$$\bar{\delta}_\xi g^{\mu\nu} = g^{\lambda\nu}\xi^\mu_{,\lambda} + g^{\mu\lambda}\xi^\nu_{,\lambda} - g^{\mu\nu}_{,\lambda}\xi^\lambda = D^\mu\xi^\nu + D^\nu\xi^\mu \tag{7.77b}$$
$$\bar{\delta}_\xi \varphi = -\varphi_{,\lambda}\xi^\lambda = -(D_\lambda\varphi)\xi^\lambda \tag{7.77c}$$
$$\bar{\delta}_\xi A_\mu = -A_\lambda\xi^\lambda_{,\mu} - A_{\mu,\lambda}\xi^\lambda = -(D_\lambda A_\mu)\xi^\lambda - A_\lambda(D_\mu\xi^\lambda). \tag{7.77d}$$

You may verify the relations with the covariant derivatives above either explicitly or follow the "comma goes to semicolon" rule in each of the expressions. From this you also may sense again that the $\bar{\delta}$ variation is "more" geometric than the δ variation. As a matter of fact, the $\bar{\delta}$ variation reflects the active and the δ variation the passive diffeomorphisms. For later purposes we also note

$$\bar{\delta}_\xi \Gamma^\lambda_{\ \mu\nu} = -D_{(\mu}D_{\nu)}\xi^\lambda + R^\lambda_{\ (\mu\nu)\sigma}\xi^\sigma. \tag{7.78}$$

7.5.4 Noether Identities and Conservation Laws

Already before the final version of general relativity was established in Nov. 1915, there was a problem regarding the conservation of energy. The Göttingen mathematician D. Hilbert had the suspicion that the failure of general relativity to deliver a conservation law for energy is intrinsic to the system. He asked E. Noether to help ensuring this suspicion. This initiated the work of Noether on what became known as her two theorems.

The "gravitational" Noether identities are in this section worked out solely for metric gravity. Tetrad gravity is treated in Appendix F.2.3. in terms of differential forms, not only for GR but for any diffeomorphism invariant theory formulated in terms of curvature and torsion.

Klein-Noether Identities

Generically, we expect for every symmetry transformation (7.77) with a vector field ξ an invariance identity

$$0 \equiv [\mathcal{L}]_A\bar{\delta}_\xi Q^A + \partial_\mu J^\mu_\xi = [\mathcal{L}]_{\mu\nu}\bar{\delta}_\xi g^{\mu\nu} + [\mathcal{L}]_\alpha\bar{\delta}_\xi Q^\alpha + \partial_\mu J^\mu_\xi \tag{7.79}$$

in accordance with (3.34). (In this notation, the Q^A comprise all dynamical fields, and the Q^α stand for the matter fields.) This relation alone, without specifying the Noether current J^μ_ξ or the dynamical theory itself, results for transformations of the form $\bar{\delta}_\xi Q^A = \mathcal{A}^A_\mu\xi^\mu + \mathcal{B}^{A\nu}_\mu(\partial_\nu\xi^\mu)$ in

$$\begin{aligned}0 &\equiv [\mathcal{L}]_A\,\bar{\delta}_\xi Q^A + \partial_\mu J^\mu_\xi = [\mathcal{L}]_A\,\mathcal{A}^A_\mu\xi^\mu + [\mathcal{L}]_A\,\mathcal{B}^{A\nu}_\mu(\partial_\nu\xi^\mu) + \partial_\mu J^\mu_\xi \\ &= [\mathcal{L}]_A\,\mathcal{A}^A_\mu\,\xi^\mu + \partial_\nu\{[\mathcal{L}]_A\,\mathcal{B}^{A\nu}_\mu\xi^\mu\} - \partial_\nu\{[\mathcal{L}]_A\,\mathcal{B}^{A\nu}_\mu\}\,\xi^\mu + \partial_\mu J^\mu_\xi.\end{aligned}$$

Now the first and the third term together vanish because of the Noether identity. Thus $\partial_\nu\{[\mathcal{L}]_A\, \mathcal{B}^{A\nu}_\mu \xi^\mu\} + \partial_\mu J^\mu_\xi \equiv 0$, and therefore the Noether current associated with any vector-field ξ can be written as

$$J^\mu_\xi = -[\mathcal{L}]_A \mathcal{B}^{A\mu}_\nu \xi^\nu + \partial_\nu U^{\mu\nu}_\xi. \tag{7.80}$$

that is, as an expression consisting of a term which vanishes when the field equations are satisfied and another term whose divergence vanishes identically. The Klein-Noether identities can be written down immediately, with no further specification of the theory.

Current Components

Since the Lagrangian of GR has second-order derivatives in the variables, the Noether current must be envisaged to have the generic form

$$J^\mu_\xi = j^\mu_\lambda \xi^\lambda + k^{\mu\rho}_\lambda \partial_\rho \xi^\lambda + l^{\mu\rho\sigma}_\lambda \partial_\rho\partial_\sigma \xi^\lambda.$$

Inserting this together with the generic form of the field variations into (7.79),

$$0 \equiv [\mathcal{L}]_A\, \mathcal{A}^A_\lambda \xi^\lambda + [\mathcal{L}]_A\, \mathcal{B}^{A\rho}_\lambda (\partial_\rho \xi^\lambda) + \partial_\mu \left(j^\mu_\lambda \xi^\lambda + k^{\mu\rho}_\lambda \partial_\rho \xi^\lambda + l^{\mu\rho\sigma}_\lambda \partial_\rho\partial_\sigma \xi^\lambda \right),$$

we get an expression that has derivatives of ξ up to order three. For this to vanish, we again require that because of the arbitrariness of the descriptors, each of the coefficients in front of a derivative of order zero to three has to vanish separately, with the resulting Klein-Noether identities

$$\begin{aligned}
0 &\equiv [\mathcal{L}]_A\, \mathcal{A}^A_\lambda + \partial_\mu j^\mu_\lambda \\
0 &\equiv [\mathcal{L}]_A\, \mathcal{B}^{A\rho}_\lambda + j^\rho_\lambda - \partial_\mu k^{\rho\mu}_\lambda \\
0 &\equiv k^{(\rho\sigma)}_\lambda + \partial_\mu l^{\mu(\rho\sigma)}_\lambda \\
0 &\equiv l^{(\mu\rho\sigma)}_\lambda .
\end{aligned}$$

Whereas these identities depend on the current components, the ensuing Noether identities only depend on the Euler derivatives and the symmetry transformations of the fields Q^A. The previous considerations and results are valid for all Lagrangians depending on at most second derivatives in the fields. What does this mean for diffeomorphism covariant theories, and specifically for GR?

Covariant Identities

To determine the explicit form of the currents and of the Klein-Noether identities we could start from an *ansatz* for the Noether current J^μ_ξ as of (3.86). This more general form is to be foreseen since the Lagrangian has second-order derivatives in the variables. As you can imagine, the calculations become rather tedious; you find them in [306]. One can streamline the calculations by intelligently guessing the current and by using covariant derivatives instead of ordinary ones. This will be done in the sequel; for a similar calculation see [413].

The current J^μ_ξ is determined by surface terms. In the case of metric gravity, they originate as follows:

- We derived previously (see (7.62)) that the variation of the Hilbert-Einstein Lagrangian yields

$$\delta\sqrt{-g}R = \sqrt{-g}(-G_{\mu\nu}\delta g^{\mu\nu} + D_\lambda \delta w^\lambda) \text{ with } \delta w^\lambda = g^{\mu\nu}(\delta\Gamma^\lambda{}_{\mu\nu} - \delta^\lambda_\nu \delta\Gamma^\sigma{}_{\sigma\mu}).$$

- Furthermore, the generic variation of the matter part yields

$$\delta\mathcal{L}_M = \frac{\partial\mathcal{L}_M}{\partial g^{\mu\nu}}\delta g^{\mu\nu} + [\mathcal{L}]^\alpha\, \delta Q_\alpha + D_\mu\left(\frac{\partial\mathcal{L}_M}{\partial(D_\mu Q_\alpha)}\delta Q_\alpha\right),$$

  where in the last term one may replace

$$\frac{\partial\mathcal{L}_M}{\partial(D_\mu Q_\alpha)} = \sqrt{-g}\frac{\partial\mathcal{L}^f_M}{\partial(\partial_\mu Q_\alpha)} = \sqrt{-g}\,\Pi^{\mu\alpha}.$$

- Now, in the case case of diffeomorphism $x^\mu \to x^\mu + \xi^\mu$, the Lagrangian transforms as a density of weight 1, i.e. $\bar{\delta}_\xi\mathcal{L} = -\partial_\mu(\mathcal{L}\xi^\mu)$, giving rise to a third surface term.
- A further term might come from the requirement of deriving the field equations together with reasonable boundary conditions on the fields and their derivatives, like example originating from the trace-K action (7.72). This is neglected in the following.

Taking into account the different contributions, we obtain the identity

$$[\mathcal{L}]_{\mu\nu}\,\bar{\delta}_\xi g^{\mu\nu} + [\mathcal{L}]^\alpha\,\bar{\delta}_\xi Q_\alpha + \sqrt{-g}D_\mu\left[\frac{1}{2\kappa}\bar{\delta}_\xi w^\mu + \Pi^{\mu\alpha}\bar{\delta}_\xi Q_\alpha + \frac{1}{\sqrt{-g}}\mathcal{L}\xi^\mu\right] \equiv 0. \tag{7.81}$$

In making use of $\sqrt{-g}D_\mu v^\mu = \partial_\mu(\sqrt{-g}v^\mu)$ we identify the currents in (7.79) as

$$J^\mu_\xi = \frac{1}{2\kappa}\sqrt{-g}\,\bar{\delta}_\xi w^\mu + \sqrt{-g}\,\Pi^{\mu\alpha}\,\bar{\delta}_\xi Q_\alpha + \mathcal{L}\xi^\mu.$$

Let us elaborate the different terms in these currents. First of all, with (7.77b) and (7.78),

$$\begin{aligned}\bar{\delta}_\xi w^\mu &= -(g^{\lambda\nu}\delta^\mu_\rho - g^{\lambda\mu}\delta^\nu_\rho)D_{(\nu}D_{\lambda)}\xi^\rho - \frac{3}{2}R^\mu{}_\rho\xi^\rho \\ &= -(D_\nu D^\nu\xi^\mu - D_\nu D^\mu\xi^\nu) - 2R^\mu{}_\rho\xi^\rho.\end{aligned} \tag{7.82}$$

Since both the variations of the fields and of the term $\bar{\delta}_\xi w^\mu$ are written in terms of covariant derivatives, we now seek for Klein-Noether identities expressed by covariant derivatives of currents. Thus generally, the currents J_ξ^μ contain terms proportional to $\xi,\ D\xi,\ DD\xi$:

$$J_\xi^\mu = \bar{j}_\nu^\mu \xi^\nu + \bar{k}_\nu^{\mu\lambda} D_\lambda \xi^\nu + \bar{l}_\nu^{\mu\rho\sigma} D_{(\rho} D_{\sigma)} \xi^\nu. \tag{7.83}$$

In order to obtain a unique partition of J_ξ^μ into these three parts, we may use the relation $D_\rho D_\sigma = D_{(\rho} D_{\sigma)} + D_{[\rho} D_{\sigma]}$ and can then express all terms proportional to $D_{[\rho} D_{\sigma]}$ by the curvature tensor. The different contributions to the currents are identified as

$$\bar{j}_\nu^\mu = -\frac{\sqrt{-g}}{2\kappa}\frac{3}{2} R_\nu^\mu + \sqrt{-g}\,\Pi^{\mu\alpha} \mathcal{U}_{\alpha\nu} + \delta_\nu^\mu \mathcal{L},$$

$$\bar{k}_\nu^{\mu\lambda} = \sqrt{-g}\,\Pi^{\mu\alpha} \mathcal{V}_{\alpha\nu}^\lambda, \qquad \bar{l}_\nu^{\mu\rho\sigma} = -\frac{\sqrt{-g}}{2\kappa}(g^{\rho\sigma}\delta_\nu^\mu - g^{\mu\rho}\delta_\nu^\sigma). \tag{7.84}$$

Furthermore, the variations of the matter fields Q_α are assumed to have the form

$$\bar{\delta}_\xi Q_\alpha = \mathcal{U}_{\alpha\lambda}\xi^\lambda + \mathcal{V}_{\alpha\lambda}^\mu D_\mu \xi^\lambda, \tag{7.85}$$

where the coefficients $\mathcal{U}$ and $\mathcal{V}$ depend on whether the fields are scalars, vectors, …

In the next step we exploit the identity (7.81), which then split into sets of Klein-Noether identities. For this purpose we need to execute (7.81) with the variation of the metric according to (7.77b) and of the matter fields according to (7.85). The result is

$$\begin{aligned} 2\,[\mathcal{L}]_{\mu\nu}\, D^\mu\xi^\nu + [\mathcal{L}]^\alpha\, \mathcal{U}_{\alpha\lambda}\xi^\lambda + [\mathcal{L}]^\alpha\, \mathcal{V}_{\alpha\lambda}^\mu D_\mu\xi^\lambda + (D_\mu \bar{j}_\nu^\mu)\xi^\nu + \bar{j}_\nu^\mu D_\mu\xi^\nu \\ + (D_\mu \bar{k}_\nu^{\mu\lambda}) D_\lambda\xi^\nu + \bar{k}_\nu^{\mu\lambda} D_\mu D_\lambda \xi^\nu + D_\mu\left(\bar{l}_\nu^{\mu\rho\sigma} D_{(\rho} D_{\sigma)}\xi^\nu\right) \equiv 0. \end{aligned} \tag{7.86}$$

In order to be able to untie this identity in terms of ξ and its derivatives, we need

$$\bar{k}_\nu^{\mu\lambda} D_\mu D_\lambda \xi^\nu = \bar{k}_\nu^{\mu\lambda} D_{(\mu} D_{\lambda)}\xi^\nu + \bar{k}_\nu^{\mu\lambda} D_{[\mu} D_{\lambda]}\xi^\nu = \bar{k}_\nu^{\mu\lambda} D_{(\mu} D_{\lambda)}\xi^\nu + \frac{1}{2}\bar{k}_\nu^{\mu\lambda} R^\nu{}_{\rho\mu\lambda}\xi^\rho$$

$$D_\mu\left(\bar{l}_\nu^{\mu\rho\sigma} D_{(\rho} D_{\sigma)}\xi^\nu\right) = -\frac{\sqrt{-g}}{2\kappa}\frac{1}{2}\left((D^\mu R_{\mu\rho})\xi^\rho + R_{\mu\rho} D^\mu \xi^\rho\right),$$

the latter showing that–in terms of covariant derivatives–the $\bar{l}_\nu^{\mu\rho\sigma}$ terms do not contribute (covariant) derivatives of order three to (7.86). Screening the identity (7.86), we observe that there is only one term containing $D_{(\mu} D_{\lambda)}\xi^\nu$. Since this is bound to vanish, we obtain the first set of identities

$$\bar{k}_\nu^{\mu\lambda} + \bar{k}_\nu^{\lambda\mu} \equiv 0. \tag{7.87}$$

Further identities are obtained from the terms proportional to $D\xi$ and ξ:

$$2\,[\mathcal{L}]_{\nu\lambda}\, g^{\nu\mu} + [\mathcal{L}]^{\alpha}\, \mathcal{V}^{\mu}_{\alpha\lambda} + \bar{j}^{\mu}_{\lambda} + D_{\nu}\bar{k}^{\nu\mu}_{\lambda} - \frac{\sqrt{-g}}{2\kappa}\frac{1}{2}R^{\mu}_{\ \lambda} \equiv 0 \tag{7.88a}$$

$$[\mathcal{L}]^{\alpha}\, \mathcal{U}_{\alpha\lambda} + D_{\mu}\bar{j}^{\mu}_{\lambda} + \frac{1}{2}\bar{k}^{\mu\rho}_{\nu}R^{\nu}_{\ \lambda\mu\rho} - \frac{\sqrt{-g}}{2\kappa}\frac{1}{2}D_{\mu}R^{\mu}_{\ \lambda} \equiv 0. \tag{7.88b}$$

Let us introduce the current

$$c^{\mu}_{\lambda} = \bar{j}^{\mu}_{\lambda} + D_{\nu}\bar{k}^{\nu\mu}_{\lambda} - \frac{\sqrt{-g}}{2\kappa}\frac{1}{2}R^{\mu}_{\ \lambda}. \tag{7.89}$$

Its covariant derivative can be written as

$$D_{\mu}c^{\mu}_{\lambda} = D_{\mu}\bar{j}^{\mu}_{\lambda} + \frac{1}{2}\bar{k}^{\mu\rho}_{\nu}R^{\nu}_{\ \lambda\mu\rho} - \frac{\sqrt{-g}}{2\kappa}\frac{1}{2}D_{\mu}R^{\mu}_{\ \lambda}$$

because, due to (7.87), $D_{\mu}D_{\nu}\bar{k}^{\nu\mu}_{\lambda} = D_{[\mu}D_{\nu]}\bar{k}^{\nu\mu}_{\lambda}$. Therefore the identities (7.88) become

$$2\,[\mathcal{L}]_{\nu\lambda}\, g^{\nu\mu} + [\mathcal{L}]^{\alpha}\, \mathcal{V}^{\mu}_{\alpha\lambda} + c^{\mu}_{\lambda} \equiv 0 \tag{7.90a}$$

$$[\mathcal{L}]^{\alpha}\, \mathcal{U}_{\alpha\lambda} + D_{\mu}c^{\mu}_{\lambda} \equiv 0. \tag{7.90b}$$

From these we derive the Noether identities

$$[\mathcal{L}]^{\alpha}\, \mathcal{U}_{\alpha\lambda} - D_{\mu}\left([\mathcal{L}]^{\alpha}\, \mathcal{V}^{\mu}_{\alpha\lambda}\right) - 2D_{\mu}[\mathcal{L}]_{\nu\lambda}g^{\nu\mu} \equiv 0. \tag{7.91}$$

Specific Cases

- In the absence of a gravitational field[29], the previous identities reduce to those for the matter part as derived in Sect. 3.3.3, namely the Klein-Noether cascades (3.72) and the Noether identities (3.75).
- In the absence of matter fields, that is for vacuum gravity, we observe that because of

  $$c^{\mu}_{\nu} = -2\frac{\sqrt{-g}}{2\kappa}(G^{\mu}_{\nu} + \Lambda\delta^{\mu}_{\nu}) = 2\,[\mathcal{L}]_{\nu\lambda}\, g^{\lambda\mu}$$

  the identity (7.90a) is verified to hold, and the identity (7.90b) is

  $$0 = D_{\mu}c^{\mu}_{\nu} = D_{\mu}(2[\mathcal{L}_{\nu\lambda}]g^{\lambda\mu}) = \frac{1}{\kappa}\sqrt{-g}D_{\mu}(G_{\nu\lambda} + \Lambda g_{\nu\lambda})g^{\lambda\mu}$$

  which is proportional to the "contracted" Bianchi identity $D_{\mu}G^{\mu\nu} = 0$. As a matter of fact, since the Bianchi identity holds for the Einstein tensor as a geometric identity independent of the field equation, (7.81) can be written

[29] Strictly speaking, we need to distinguish the case of no gravitation at all–which amounts to $g_{\mu\nu} = \eta_{\mu\nu}$, $D_{\mu} = \partial_{\mu}$–from the case of a fixed gravitational background field for which there is no (dynamical) part for the gravitational field in the total Lagrangian, but where covariant derivatives–now defined with respect to the background metric–still remain in the expressions.

$$\begin{aligned}0 &\equiv D_\lambda \left\{2g^{\lambda\nu}[\mathcal{L}]_{\mu\nu}\xi^\mu + \delta^\lambda_\mu \frac{1}{2\kappa}\bar{\delta}_\xi w^\mu + \delta^\lambda_\mu \mathcal{L}\xi^\mu\right\} \\ &= D_\lambda \left\{2R^\lambda{}_\mu \xi^\mu + \delta^\lambda_\mu \frac{1}{2\kappa}\bar{\delta}_\xi w^\mu\right\} =: D_\lambda \bar{\mathcal{J}}^\lambda_\xi,\end{aligned} \tag{7.92}$$

exhibiting a covariantly conserved current $\bar{\mathcal{J}}^\mu_\xi$.

- Let us take as a further example the case of gravity coupled to a scalar field φ. Now $\mathcal{U}_\nu = -\partial_\nu\varphi$, and because of $\mathcal{V} = 0$ also $\bar{k}^{\mu\lambda}_\nu = 0$; see (7.84). Thus the identities (7.90) are specialized to

$$2\,[\mathcal{L}]_{\nu\lambda}\, g^{\nu\mu} + c^\mu_\lambda \equiv 0 \qquad\qquad -[\mathcal{L}]^\varphi \partial_\lambda\varphi + D_\mu c^\mu_\lambda \equiv 0 \tag{7.93}$$

with

$$\begin{aligned} c^\mu_\lambda = j^\mu_\lambda - \frac{\sqrt{-g}}{2\kappa}\frac{1}{2}R^\mu_\lambda &= \left(-\frac{\sqrt{-g}}{2\kappa}\frac{3}{2}R^\mu_\lambda + \delta^\mu_\lambda \mathcal{L}_G\right) \\ &+ \left(-\sqrt{-g}\frac{\partial\mathcal{L}^f_M}{\partial(\partial_\mu\varphi)}\partial_\lambda\varphi + \delta^\mu_\lambda\sqrt{-g}\mathcal{L}^f_M\right) - \frac{\sqrt{-g}}{2\kappa}\frac{1}{2}R^\mu_\lambda \\ &= -2\frac{\sqrt{-g}}{2\kappa}(G^\mu_\lambda + \Lambda\delta^\mu_\lambda) - \sqrt{-g}\Theta^\mu{}_\lambda \\ &= -2\,[\mathcal{L}]_{\lambda\nu}\, g^{\nu\mu} + 2\frac{\delta S_M}{\delta g^{\lambda\nu}}g^{\nu\mu} - \sqrt{-g}\Theta^\mu{}_\lambda. \end{aligned}$$

Here Θ is the canonical energy-momentum tensor for the scalar field. If this expression for c^μ_λ is inserted into the first identity of (7.93) one finds that according to (7.64) $T^{\mu\nu}_H = \Theta^{\mu\nu}$. Notice what happened in this example of a scalar field: identities following from Noether's second theorem prove the equality of the Hilbert and the canonical energy-momentum tensor. This holds true for other fields as well, as shown in Appendix F.2.3. Because of the Bianchi identity $D_\mu G^{\mu\nu} \equiv 0$, the second identity in (7.93) is equivalent to

$$D_\mu\Theta^\mu{}_\lambda = -\frac{1}{\sqrt{-g}}[\mathcal{L}]^\varphi \partial_\lambda\varphi,$$

revealing that the matter energy-tensor is covariantly conserved whenever the matter field equations are satisfied.

Conservation Laws

Is energy conserved in general relativity? As strange as this question might seem, it has not been fully answered up to today. The answer partially depends on what you understand by "energy" and by "conserved"[30]. For instance you may ask for the total energy (in a closed system), for a quasi-local energy (in a bounded region

[30] see the appendix of [174] for misunderstandings about these notions.

of spacetime) or for the localized energy (in other words: the energy density). And of course the quest for an appropriate definition of energy applies to all the ten components of the energy-momentum tensor.

There is not yet a consensus on how energy, momentum and other conserved quantities for gravitational theories should be defined in an inherent way. This is astounding, since in previous chapters energy-momentum was shown to be associated with the fundamental structure of spacetime. In Minkowski spacetime, the conservation of the energy-momentum tensor is a consequence of the Noether symmetries related to translations and Lorentz transformations. Since these are due to global transformations, the first Noether theorem applies. In GR we are dealing with arbitrary coordinate transformations and, as will be seen, Noether's second theorem yields a plethora of conserved charges whose interpretation is not immediately always obvious.

As elucidated in Sect. 7.5.2, it is by the coupling of matter/energy fields to gravitation which allows to define the "correct" matter energy-momentum tensor. Now it is most idiosyncratic that a similar definition for the gravitational field leads to an identically vanishing energy-momentum tensor by virtue of the field equations.

Let us resume what immediately follows from diffeomorphism symmetry. The variation of the full action of GR, i.e. $S_{GR}[g, Q] = S_G[g] + S_M[g, Q]$, is

$$\delta S = \int d^4x \, \frac{\delta S_G}{\delta g_{\mu\nu}} \delta g_{\mu\nu} + \int d^4x \, \frac{\delta S_M}{\delta g_{\mu\nu}} \delta g_{\mu\nu} + \int d^4x \, \frac{\delta S_M}{\delta Q^\alpha} \delta Q^\alpha .$$

This entails the gravitational and matter field equations. Assume that the matter field equations are satisfied, that is, $\frac{\delta S_M}{\delta Q} = 0$. Now ask for diffeomorphism symmetries:

$$\begin{aligned} 0 = \delta_\xi S &= -2 \int d^4x \, \frac{\delta S_G}{\delta g_{\mu\nu}} (D_\mu \xi_\nu) - 2 \int d^4x \, \frac{\delta S_M}{\delta g_{\mu\nu}} (D_\mu \xi_\nu) \\ &= +2 \int d^4x \, D_\mu (\frac{\delta S_G}{\delta g_{\mu\nu}}) \, \xi_\nu + 2 \int d^4x \, D_\mu (\frac{\delta S_M}{\delta g_{\mu\nu}}) \, \xi_\nu . \end{aligned}$$

The first integrand vanishes identically because of the contracted Bianchi identity $G^{\mu\nu}{}_{;\nu} = 0$. Requiring invariance, the second term is bound to vanish for arbitrary descriptors ξ. Thus $D_\mu(\frac{\delta S_M}{\delta g_{\mu\nu}}) = 0$ or $D_\mu T^{\mu\nu} = 0$ provided that the matter field equations are satisfied. (We saw this in the previous section as following from the Klein-Noether identities; there applied to scalar matter.) On the other hand, the vanishing of the covariant derivative of the matter energy-momentum tensor $T^{\mu\nu}$ follows from the gravitational field equations $G^{\mu\nu} = -\kappa T^{\mu\nu}$. This assures consistency of GR!

Pseudotensors and Superpotentials

Since the energy-momentum tensor T^μ_ν is covariantly conserved, i.e.

$$D_\mu \sqrt{-g}\, T^\mu_{\ \nu} \;=\; \partial_\mu \sqrt{-g}\, T^\mu_{\ \nu} - \Gamma^\lambda_{\ \nu\mu} \sqrt{-g}\, T^\mu_{\ \lambda} \;=\; 0,$$

there is no continuity equation from which a globally conserved quantity can be derived. This reflects, that in the presence of gravitation, energy and momentum of matter alone are not conserved, in general. Indeed, it seems quite natural to expect that since matter fields locally exchange energy-momentum with the gravitational field, the latter itself should act as a source of gravity. Thus very early came the search for a quantity $t^\mu_{\ \nu}$–meant to represent the energy-momentum of the gravitational field–such that if this is added to the matter energy-momentum tensor the full expression has a vanishing divergence.

$$\partial_\mu \sqrt{-g}\, (T^\mu_{\ \nu} + t^\mu_{\ \nu}) = 0.$$

Einstein for instance [154], preferring "his" first-order action (7.71), defined the energy-momentum complex for the gravitational field by

$$(2\kappa)\ {}_E t^\mu_{\ \nu} = \frac{\partial G^*}{\partial g_{\rho\sigma,\mu}} g_{\rho\sigma,\nu} - \delta^\mu_\nu G^*$$

–as would seem natural, being constructed similar to the canonical energy-momentum tensor for other field theories. The field equations can be written as

$${}_E \mathcal{T}^\mu_{\ \nu} = T^\mu_{\ \nu} + {}_E t^\mu_{\ \nu} = -\partial_\rho \left(\frac{\partial G^*}{\partial g^{\sigma\nu}_{\ ,\rho}} g^{\sigma\mu} \right),$$

and indeed the partial derivative ∂_μ of the right-hand side vanishes identically, and thus $\partial_\mu\, {}_E \mathcal{T}^\mu_{\ \nu} \equiv 0$. But, obviously the quantity ${}_E \mathcal{T}^\mu_{\ \nu}$ is not a tensor. The meaning of this *pseudotensor* decomposition became a matter of debate between F. Klein and A. Einstein–and it more or less triggered the work of Noether on "her" theorems.

Various other pseudotensors $\mathcal{T}^\mu_{\ \nu}$ were proposed and investigated (by Papapetrou, Bergmann, Moeller, Landau-Lifschitz, Weinberg, Penrose), some of them having the property of being symmetric and thus amenable to be used to define angular momenta. As a matter of fact, there exists an infinity of pseudotensors as seen by the following argumentation: Make a choice of a *superpotential* $U^{\mu\lambda}_{\ \ \ \nu} = U^{[\mu\lambda]}_{\ \ \ \ \nu}$ and define from this a gravitational energy-momentum tensor by

$$\kappa \sqrt{-g}\, t^\mu_{\ \nu} := \sqrt{-g}\, G^\mu_{\ \nu} + \frac{1}{2} \partial_\lambda U^{\mu\lambda}_{\ \ \ \nu}.$$

Then from the field equations $G^\mu_{\ \nu} = -\kappa T^\mu_{\ \nu}$ we find that the divergence of the total energy-momentum pseudotensor

$$2\kappa \sqrt{-g}\, \mathcal{T}^\mu_{\ \nu} = 2\kappa \sqrt{-g}\, (T^\mu_{\ \nu} + t^\mu_{\ \nu}) = \partial_\lambda U^{\mu\lambda}_{\ \ \ \nu} \tag{7.94}$$

vanishes identically due to the antisymmetry of the superpotential in its upper indices. The previous construction can be generalized by introducing superpotentials defined from objects $U^{\mu\lambda\rho}_{\ \ \ \ \nu}$ with appropriate index symmetries. These are needed to define symmetric pseudotensors.

Some superpotentials were specifically investigated in the past: In 1939, P. von Freud (who "invented" the idea of superpotentials) realized that Einstein's pseudotensor can be derived from

$$ {}_F\mathcal{U}_\lambda^{\ \mu\nu} := (-g)^{\frac{1}{2}} g_{\lambda\rho}\partial_\sigma \left[(-g)(g^{\mu\rho}g^{\nu\sigma} - g^{\nu\rho}g^{\mu\sigma})\right]. $$

And A. Komar [326] discovered the "covariant" superpotential[31]

$$ {}_K U_\xi^{\mu\nu} := \frac{1}{2\kappa}\sqrt{-g}(D^\mu\xi^\nu - D^\nu\xi^\mu). \tag{7.95} $$

Indeed this expression is covariant, but at the cost of being dependent on arbitrary vector fields. Another one is named after the initials of its inventors (J. Katz, J. Bicak, D. Lynden-Bell) KBL-pseudopotential. It is defined with reference to a background metric[32], designated as $\bar{g}_{\mu\nu}$. A covariant derivative $\bar{D}_\mu$ is defined with respect to this background metric. The background metric makes it possible to define tensors from the difference of connections like $\Delta\Gamma = \Gamma - \bar{\Gamma}$. The KBL-superpotential can then be phrased in terms of the Komar superpotential and an additional term as

$$ \begin{aligned} {}_{KBL}U_\xi^{\mu\nu} &= {}_K U_\xi^{\mu\nu} - {}_K \bar{U}_\xi^{\mu\nu} + S^\mu\xi^\nu - S^\nu\xi^\mu \\ S^\mu &:= \frac{\sqrt{-g}}{2\kappa}(\Delta\Gamma^\mu_{\rho\sigma}g^{\rho\sigma} - \Delta\Gamma^\sigma_{\rho\sigma}g^{\mu\rho}). \end{aligned} \tag{7.96} $$

As observed by P. Bergmann and J. Goldberg in the early 1950's, superpotentials are rooted in Noether's invariance condition. Indeed, previously we derived that if the symmetry variations of the fields are of the form $\bar{\delta}_\xi Q^A = \mathcal{A}^A_\mu\xi^\mu + \mathcal{B}^{A\nu}_\mu(\partial_\nu\xi^\mu)$, the Noether currents can according to (7.80) be written as

$$ J_\xi^\mu = -[\mathcal{L}]_A\mathcal{B}^{A\mu}_\nu\xi^\nu + \partial_\nu U_\xi^{\mu\nu} $$

that is, as an expression consisting of a term which vanishes on-shell and another term whose divergence vanishes identically. This is a superpotential. Since, on the other hand, we know the coefficients in the expansion $J_\xi^\mu = \bar{j}^\mu_\nu\xi^\nu + \bar{k}^{\mu\lambda}_\nu D_\lambda\xi^\nu + \bar{l}^{\mu\rho\sigma}_\nu D_{(\rho}D_{\sigma)}\xi^\nu$ from (7.84), we are able to identify a superpotential in this expression. Take vacuum gravity as an example: Diffeomorphism symmetry results for the previous expression in

$$ {}_G J_\xi^\mu = -\frac{\sqrt{-g}}{\kappa}(G^\mu_\nu + \Lambda\delta^\mu_\nu)\xi^\nu + \partial_\nu({}_G\mathcal{U}_\xi^{\mu\nu}). \tag{7.97} $$

On the other hand one can derive from (7.84)

[31] In the 1980's there was some quarrel about a factor $\frac{1}{2}$ compared to the original Komar superpotential and the related Moeller pseudotensor. This is settled in the meantime; see e.g [312].

[32] To my knowledge, these by now very common techniques originate from [448] and were rediscovered and refined in [1].

$$
\begin{aligned}
{}_G J^\mu_\xi &= -\frac{\sqrt{-g}}{2\kappa}\left(D_\nu D^\nu \xi^\mu - D_\nu D^\mu \xi^\nu) + 2(G^\mu_\nu + \Lambda \delta^\mu_\nu)\xi^\nu\right) \\
&= -\frac{\sqrt{-g}}{\kappa}(G^\mu_\nu + \Lambda \delta^\mu_\nu)\xi^\nu + \frac{1}{2\kappa}\partial_\nu\left(\sqrt{-g}(D^\mu \xi^\nu - D^\nu \xi^\mu)\right).
\end{aligned}
$$

By this we can identify the superpotential with the Komar superpotential. This is also shown in [21]; where the authors furthermore rederive the Papapetrou pseudotensor by the help of Noether's theorem and a Belinfante symmetrization procedure.

In other contexts we saw that the Noether theorems offer a procedure of constructing conserved charges for any physical system provided its symmetries have been identified. In Subsect. 3.3.3 it was shown that both the Noether current J^μ_ξ and the current component j^μ_r in $J^\mu_\xi = j^\mu_r \epsilon^r +$ (derivatives of ξ) are conserved on-shell. In the case of vacuum GR one can determine the current j^μ_r from (7.97) as

$$
j^\mu_\nu = \frac{\sqrt{-g}}{2\kappa}\left((R - 2\Lambda)\delta^\mu_\nu + \Gamma^\lambda_{\lambda\rho,\nu}\, g^{\rho\mu} - \Gamma^\mu_{\lambda\rho,\nu}\, g^{\lambda\rho}\right).
$$

This current is not coordinate-invariant, which of course is reflected in that it obeys $\partial_\mu j^\mu_\nu \doteq 0$ and not $D_\mu j^\mu_\nu \doteq 0$. Thus, in general, it is not suited to define a conserved charge. Indeed this pseudocurrent is proportional to the Einstein energy-momentum pseudotensor. What's about the Noether current itself which on-shell is given by the derivative of a superpotential? Consider a region M in four-dimensional spacetime, bounded by ∂M, taken to be a timelike cylinder together with two spacelike surfaces at t_0 and t_1. From the on-shell conservation of the gravitational Noether currents ${}_G J^\mu_\xi$ (the index G will be dropped in the following expressions) we get by Stokes' theorem

$$
\begin{aligned}
\int_M d^4x\, \partial_\mu J^\mu_\xi = 0 \iff & \int_{\partial M} d^3x\, n_\mu J^\mu_\xi = 0 \\
\iff & \int_{t=t_1} d^3x\, J^0_\xi - \int_{t=t_0} d^3x\, J^0_\xi + \int_{t_0}^{t_1} dt \int_S d^2x\, n_i J^i_\xi = 0.
\end{aligned}
$$

Here S is the 2-dimensional sphere at infinity, and n_i is its unit normal vector pointing towards infinity. If the last term vanishes we derive conserved charges

$$
Q_\xi = \int_S d^2x\, n_i\, U^{0i}_\xi, \tag{7.98}
$$

since $J^0_\xi = \partial_i U^{0i}_\xi$ with the superpotential U. Since both the currents ${}_G J_\xi$ and the superpotential (7.95) depend on the vector field ξ, also the charges Q_ξ depend on the vector fields. You may interpret this either as dealing with an infinite set of conservation laws, or with the failure to obtain genuine conservation laws at all. For certain choices of ξ, the charges, formally expressed by (7.98), may be trivial (numbers), for other choices they may not exist at all. Ultimately one is interested to identify conserved charges corresponding for instance to energy or angular momentum. We know that in flat space the existence of these charges is related to invariances under time translations and rotations, respectively. Therefore the consideration of (7.98) is often restricted to cases where Killing symmetries exist; these imposing conditions on the vector fields ξ.

Pseudotensors depend on the reference frame chosen (or - in the covariant version– on the arbitrary vector fields ξ) and thus cannot describe the local gravitational energy-momentum tensor. The origin of this failure is the equivalence principle which inhibits the detection of a gravitational field at a point. (This is true in general: One can prove that Einstein's theory has no local gauge invariant observables.) For this reason, the pseudotensor approach has been largely abandoned. The idea was heavily criticized in the influential textbook by Misner, Thorne and Wheeler: "Anyone who looks for a magic formula for 'local gravitational energy-momentum' is looking for the right answer to the wrong question. Unhappily, enormous time and effort were devoted in the past to trying to 'answer this question' before investigators realized the futility of the enterprise" [373] (p. 467). However, as described later and in more details in Appendix F.3.1, there is a direct relationship between pseudotensors and quasi-local energy expressions.

Quasi-local Energy

If energy-momentum cannot be defined locally, what's about defining it in finite regions? The idea is to enclose a region of spacetime with a membrane (i.e. a closed spacelike two-surface) and to attach an energy-momentum four-vector to it. The idea became popular by the work of J. Brown and J.W. York; see e.g. [60]. They introduced quasilocal quantities by a Hamilton-Jacobi approach. Since the techniques are rather involved, I do not go into details here. The expressions they obtained turned out to be useful in many applications, especially in black-hole physics. Today, a very large number of expressions for quasi-local energy is known. The euphoria about this approach got damped after it was found that there is an infinite number of quasilocal expressions for the gravitational field–even under reasonably required properties of these objects; for a review see [495].

Energy and Boundary Terms

A clear and consistent picture on gravitational energy-momentum conservation emerged only after the importance of boundary conditions had been recognized and understood better. The role of boundary terms was already described before: We saw that the proper adjustment of the values of the gravitational field and/or its derivatives on a 3D-hypersurface is only possible if appropriate surface integrals (Gibbons-Hawking-York terms) are added to the Hilbert-Einstein action. Appendix C.3.3 contains a section about surface integrals needed in the Hamiltonian formulation of general relativity (Regge-Teitelboim term). Different boundary conditions need different surface terms in order for the internal consistency of the theory. Here one learns that what is called "energy" of a gravitational system depends on boundary conditions as well. This is not at all a peculiarity of gravitation. In thermodynamics for instance, there are different notions of energy as well, such as the internal energy, free energy, etc., each of them defined only with respect to distinct boundary conditions. And all of them are physically meaningful.

A related approach is the one by B. Julia and S. Silva [306]. They remove the arbitrariness in the superpotential by an extra condition which relates the variation of the superpotential to the variations of the fields on the boundary.

It has been shown by J.M. Nester and collaborators (see e.g. [88], [382]) that various forms of pseudotensors and quasilocal expressions of gravitational energy-momentum are related to Lagrangian and Hamiltonian boundary terms. Nester et al. start from the observation that for every pseudotensor, derived from a superpotential according to (7.94), it is not unreasonable to interpret the quantity

$$P_\nu = \int_V \sqrt{-g}\,\mathcal{T}^\mu_\nu (d^3x)_\mu = \frac{1}{2\kappa}\int_V \partial_\lambda U^{\mu\lambda}{}_\nu (d^3x)_\mu = \frac{1}{2\kappa}\oint_{\partial V} U^{\mu\lambda}{}_\nu \frac{1}{2}\, dS_{\mu\lambda}.$$

as the total conserved energy-momentum within a volume V. This is a quasi-local expression: It depends on the fields only on the boundary–and of course also on the reference frame because the pseudopotential is not a genuine tensor. But how could the conserved quantity P_ν have any relation to energy-momentum? Nester et al. find that the Hamiltonian is "hidden" in the previous derivation. Consider the object

$$\begin{aligned}
H(N) &:= \int_V \sqrt{-g}\, N^\nu \mathcal{T}^\mu_\nu (d^3x)_\mu \\
&= \int_V \sqrt{-g}\, N^\nu \left[\left(T^\mu_\nu + \frac{1}{\kappa}G^\mu_\nu\right) + \left(-\frac{1}{\kappa}G^\mu_\nu + t^\mu_\nu\right)\right](d^3x)_\mu \\
&= \int_V \sqrt{-g}\, N^\nu \left(T^\mu_\nu + \frac{1}{\kappa}G^\mu_\nu\right)(d^3x)_\mu - \frac{1}{2\kappa}\int_V \partial_\lambda\left(N^\nu U^{\mu\lambda}{}_\nu\right)(d^3x)_\mu \\
&= \int_V N^\mu \mathcal{H}_\mu + \int_{\partial V} \mathcal{B}.
\end{aligned}$$

This is the Hamiltonian of vacuum gravity in the ADM form. As elucidated in Appendix C.3.3, the canonical variables in the phase-space description of metric gravity are the lapse and the shift functions N^μ together with their ("weakly" vanishing) momenta, and the three-geometry metric components g_{ij} together with their canonically conjugate momenta, which are directly related to the extrinsic curvature of the three-geometry. The Hamiltonian becomes a sum of ("weakly" vanishing) constraints $\mathcal{H}_\mu$ which only depend on the three-geometry and its canonically conjugate momenta. Observe that the energy–being the numerical value of the Hamiltonian–is completely determined by the surface term. Since for a closed spacetime with compact boundary no surface term arises, its energy vanishes. For asymptotic flat spaces, the surface term becomes the ADM energy (C.112)

$$E_{ADM}\left[h_{ij}\right] = \oint d^2 s_k (h_{ik,i} - h_{ii,k})$$

as originally postulated heuristically by B. DeWitt, or derived from internal consistency by T. Regge and C. Teitelboim, and later shown by S. Hawking and G. Horowitz to directly follow from the boundary terms in the action. For stationary spacetimes the ADM energy becomes the same as the Komar energy–this ensures consistency. The nine other conservation laws follow from the Regge/Teitelboim surface terms (respectively their later improvements/refinements as described in Appendix C.3.3).

For any pseudotensor, the associated superpotential is found to be a Hamiltonian boundary term. In [384], Nester et al. give a complete listing of comparisons of pseudotensors and Hamiltonian formulations not only for energy-momentum, but also for angular momentum and center-of-mass.

Similar in spirit of Nester et al. is the approach by R. Wald [527]; see also [294]. He recognizes that the total derivative in the variation of the Lagrangian can be used to define a symplectic structure ("symplectic current"). By this Hamiltonians might exist, the circumstances depending on boundary conditions. For every Hamiltonian one then can associate a charge/energy; see more details in Appendix F.3.1. The article [527] carries the title "Black hole entropy is Noether charge". This provokes curiosity, to say the least. Indeed, when E. Noether formulated her theorems, only one person knew about black holes in GR–namely K. Schwarzschild–although he did not name "his" solution of the source-free Einstein equations by this name. And combining the thermodynamic concept of entropy in conjunction with the gravitational concept of black holes was one of the great enterprises rather slowly evolving in theoretical physics beginning in the 1970's. To give more details here, would require an explanation of Killing horizons, for which this book is not the right place. But anyhow, you agree with me that this is an exciting development: taking serious otherwise simply "thrown away" total derivatives, one would not be able to derive the entropy of black holes from diffeomorphism symmetry.

Total Energy

Since a local gravitational energy density does not exist, the total energy of an isolated system can be defined in a reasonable way only for an asymptotically flat spacetime[33]. The asymptotic condition replaces the homogeneity of time in "ordinary" relativistic field theories. Surely the integrals over the pseudotensorial energy densities, or the asymptotically extended quasi-local energy expressions should result in "reasonable" total energies, like for instance the ADM mass (for spatial infinity), the Bondi energy (for null infinity), and the Komar energy.

The Komar energy is defined for spacetimes with a timelike Killing vector field[34] K_μ and can be derived by defining at first the current

$$J^\mu_K = K_\nu R^{\mu\nu}.$$

[33] Here I refrain from a mathematically proper defintion which you find in any advanced textbook on GR (e.g. [526]). Informally, an asymptotically flat spacetime is a manifold for which at large distances from some region the geometry becomes indistinguishable from that of Minkowski spacetime or–referring to solutions of the field equations–the gravitational field, as well matter fields become negligible in magnitude.

[34] The existence of a timelike Killing vector field for a metric means that coordinates can be found such that the metric components are independent of the time variable. These are also called stationary spacetimes, the most well-known representatives being the Schwarzschild black holes, and the rotating and charged ones.

This current is conserved: $D_\mu J^\mu_K = (D_\mu K_\nu) R^{\mu\nu} + K_\nu D_\mu R^{\mu\nu} = 0$, the first term vanishing because of the Killing vector condition, and the second one vanishing because $K_\nu D_\mu R^{\mu\nu} = \frac{1}{2} K_\nu D^\nu R = 0$, since the directional derivative of the curvature scalar vanishes along a Killing vector. Since $D_\mu D_\nu K^\mu = R_{\mu\nu} K^\mu$, the current itself can be written as $J^\mu_K = D_\nu(D^\mu K^\nu)$. To this conserved Komar current (remember that covariant conservation of a current can be rewritten as 'ordinary' conservation: $\sqrt{-g}\, D_\mu J^\mu = \partial_\mu \sqrt{-g} J^\mu$) can be defined a conserved charge

$$E_K = N_K \int_\Sigma d^3x \sqrt{\gamma}\, n_\mu J^\mu_K = N_K \int_\Sigma d^3x \sqrt{\gamma}\, n_\mu D_\nu (D^\mu K^\nu).$$

Here N_K is a normalization constant, γ is the induced metric on the spacelike hypersurface Σ, and n^μ the unit normal to Σ. The Komar energy E_K (in another way originally derived in [326]) can be rewritten as an integral over a surface at spatial infinity; a two-sphere, say:

$$E_K = N_K \int_{\partial\Sigma} d^2x \sqrt{\tilde{\gamma}}\, n_\mu \sigma_\nu (D^\mu K^\nu), \tag{7.99}$$

with $\tilde{\gamma}$ being the metric on the sphere, and σ_ν the outward-pointing normal vector. That it makes sense to identify E_K with the total energy of a stationary spacetime is supported by evaluating it for specific metrics. For a Schwarzschild metric with a mass M, for instance, $E_K = (4\pi G\, N_K) M$ which fixes the normalization constant as $N_K = (4\pi G)^{-1}$.

Positivity of Energy

It might come as a surprise, that the question about the positivity of the energy in gravitational systems was only definitely answered in the 1980's. In other classical field theories, the total energy is the integral of a positive definite energy density. In gravity, the situation is different, because there is no energy density at all. Nevertheless there were some hints that the total energy is always strictly positive, except for flat Minkowski space, which has zero energy. The positivity of the ADM energy was proven[35] by R. Schoen and S.-T. Yau in 1981. Their proof became greatly simplified by E. Witten from an energy expression which is an integral over a space-like surface of an explicitly positive-definite quantity[36]. By this expression a number of related problems were solved (for references see p. 295 in [526]). The positivity of the energy in general relativity ensures the stability of the "ground state", namely Minkowski space.

---

[35] Strictly speaking, their proof needs the assumption that the "dominant energy condition" is valid.

[36] In Witten's energy expression occurs technically a spinor quantity. In hindsight it was observed that this comes about because the generator of diffeomorphism is the square of the generator of supersymmetry transformations [281].

7.6 Modifications and Extensions of/to General Relativity

In the genesis of conceptual and/or experimental improvements to Newton's theory of gravitation, general relativity was not unrivaled, but it competed in the beginning of the 20th century with ideas and drafts by for instance Max Abraham, Gustav Mie and Gunnar Nordström; see [439]. Einstein's general relativity "won the race" at that time. But–as history of physics reveals–so far no theory survives at the very end, but is surpassed by a more general theory. This might also be the fate of GR[37].

The urge for modifying or extending general relativity is driven by several motives. One originates from intellectual curiosity concerning the geometric stance of GR. Another conceptual motive relates to the success of the standard model of particle physics as being from its very structure a Yang-Mills theory: It seems natural to ask, whether GR can be understood as a gauge theory. And still another one comes from the extraordinarily successful concordance model of cosmology with its price of requiring an unexpected large dark matter and dark energy contribution. The efforts at explaining the "darkness" have provoked hundreds of speculations, among others searches for alternatives to GR–some of them being rather drastic in the sense of shaking tenets stressed in this book, like e.g. Lorentz invariance or the equivalence principle. It seems that the basic axioms of a theory of gravitation are still unknown [480], and that the results from cosmology and astrophysics are still to meager to restrict the arena of gravitational theories [72]. In a quite recent review article on modifications of GR [93] the authors write: "Lovelock's theorem means that to construct metric theories of gravity with field equations that differ from those of General Relativity, we must do one (or more) of the following: (i) Consider other fields, beyond (or rather than) the metric tensor. (ii) Accept higher than second derivatives of the metric in the field equations. (iii) Work in a space with dimensionality different from four. (iv) Give up on either rank (2, 0) tensor field equations, symmetry of the field equations under exchange of indices, or divergence-free field equations. (v) Give up locality."

Before dealing directly with modifications and extensions, let us contemplate a re-interpretation of GR.

7.6.1 Interpreting GR as a Spin-2 Field Theory

Why is it that GR is formulated in geometric terms (i.e. Riemann geometry) in contrast to the well-established theories of the other fundamental interactions, which are formulated as field theories on the background of Minkowski space? As already mentioned in Sect. 5.3.5, M. Fierz and W. Pauli in their work on wave equations for higher spin fields [186] took a closer look at spin-2 fields and found a Lagrangian with a symmetric field $h_{\mu\nu}$ for which the field equations for $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ are

[37] This touches the question of whether there will ever exist a final theory of everything, a point to be discussed in the conclusion, Sect. 9.2.

the same as Einstein's vacuum equations in the weak-field limit. With later work by S. Gupta (1954) and R.H. Kraichnan (1955), it seemed as if it is possible to uniquely derive full GR (and not only the weak-field limit) by starting from the linear theory and successively generating higher-order corrections by coupling the gravitational field to its own Hilbert energy-momentum tensor. This bootstrapping gravity point of view became popular in the early sixties, with R. Feynman's Caltec lectures on gravitation [183]. He introduced GR not in the otherwise common geometrodynamic view, but directly in the spirit of a field theory, which eventually can also be reinterpreted in terms of Riemann curvatures. This apparent equivalence of Einstein's geometrodynamic approach with a self-consistent spin-2 field approach is meanwhile widely quoted in textbooks on GR. Only recently T. Padmanabhan [402] pinpointed a number of weak points in the seemingly stringent chain of arguments in deriving gravity from gravitons.

Start with an *ansatz* for the action of non-interacting massless fields $h_{\mu\nu}$ built from derivatives $\partial_\lambda h_{\mu\nu}$. Thus look for the most generic expression

$$S = \int d^4x\, (\partial_\lambda h_{\mu\nu})(\partial_\tau h_{\rho\sigma}) M^{\lambda\mu\nu\tau\rho\sigma}(\eta)$$

where–as indicated–the tensor M depends solely on the Minkowski metric η. Indeed M must be cubic in the Minkowski metric in order to yield a scalar action. Not all possible permutations of indices lead to independent terms, and some combinations of terms can be written as total derivatives. As a result the most generic action can be written (using the conventions of Appendix A in [402])

$$\begin{aligned} S_h = \frac{1}{4}\int d^4x \big[&c_1(\partial_\lambda h^\mu_\mu)(\partial^\lambda h^\nu_\nu) + c_2(\partial_\lambda h_{\mu\nu})(\partial^\lambda h^{\mu\nu}) \\ &+ (c_3+c_4)(\partial_\lambda h_{\mu\nu})(\partial^\nu h^{\mu\lambda}) + c_5(\partial_\lambda h^\mu_\mu)(\partial_\nu h^{\nu\lambda})\big] \end{aligned}$$

plus a boundary term which is dropped here (details again in [402]). The coefficients c_i in this expression become fixed (up to an overall factor) by the requirement that the theory is invariant under transformations

$$h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu. \tag{7.100}$$

Then the action can be written as

$$\begin{aligned} S_h = \frac{1}{4}\int d^4x \big[&(\partial_\lambda h)(\partial^\lambda h) - (\partial_\lambda h_{\mu\nu})(\partial^\lambda h^{\mu\nu}) \\ &+ 2(\partial_\lambda h_{\mu\nu})(\partial^\nu h^{\mu\lambda}) - 2(\partial_\lambda h)(\partial_\nu h^{\nu\lambda})\big], \end{aligned} \tag{7.101}$$

with the abbreviation $h := h^\mu{}_\mu$.

Let us for comparison calculate the field equations of linearized GR. The starting point is the decomposition of the metric $g_{\mu\nu}$ as

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x) \qquad \text{with} \qquad |h_{\mu\nu}| \ll 1. \tag{7.102}$$

From this $g^{\mu\nu} \simeq \eta^{\mu\nu}$-$h^{\mu\nu}$, where the $\simeq$ symbol denotes that terms of order $(h^{\mu\nu})$ were neglected. Observe that the raising and lowering of indices of tensors proportional to $h^{\mu\nu}$ occurs with the Minkowski metric. Take it as an exercise to derive that in the linearized approximation,

$$\begin{aligned}
R_{\mu\nu} &\simeq \frac{1}{2}(\Box h_{\mu\nu} - h^{\lambda}_{\ \nu,\lambda\mu} - h^{\lambda}_{\ \mu,\lambda\nu} + h_{,\mu\nu}) \\
R \ &\simeq \Box h - h^{\lambda\mu}_{\ \ ,\lambda\mu} \\
G_{\mu\nu} &\simeq \frac{1}{2}(\Box h_{\mu\nu} - h^{\lambda}_{\ \nu,\lambda\mu} - h^{\lambda}_{\ \mu,\lambda\nu} + h_{,\mu\nu}) - \frac{1}{2}(\Box h - h^{\lambda\mu}_{\ \ ,\lambda\mu})\eta_{\mu\nu},
\end{aligned}$$

and that the vacuum field equations $G_{\mu\nu} = 0$ are the stationary points of the action (7.101).

The gauge invariance (7.100) signals that the fields $h_{\mu\nu}$ are redundant and are not pure spin-2 fields. Make this explicit by writing the tensor $h_{\mu\nu}$ in terms of its transverse traceless part $Q_{\mu\nu}$ (with $\partial_\mu Q^{\mu\nu} = 0$ and $\eta^{\mu\nu} Q_{\mu\nu} = 0$), a transverse vector A_μ (with $\partial^\mu A_\mu = 0$) and two scalars α and β as

$$h_{\mu\nu} = Q_{\mu\nu} + \partial_\mu A_\nu + \partial_\nu A_\mu + \left\{\partial_\mu \partial_\nu - \frac{1}{4}\eta_{\mu\nu}\partial^2\right\}\alpha + \frac{1}{4}\eta_{\mu\nu}\beta.$$

Plugging this into the action S_h it becomes

$$S_h = -\frac{1}{2}\int d^4x \left[(\partial_\lambda Q_{\mu\nu})(\partial^\lambda Q^{\mu\nu}) - \frac{3}{8}(\partial_\mu \gamma)(\partial^\mu \gamma)\right] \quad \text{with} \quad \gamma := \beta - \partial^2 \alpha.$$

Thus the vector degree of freedom completely drops out and the Q- and γ-fields are decoupled from each other. However, the remaining scalar field appears in the action with the wrong sign to constitute a kinetic energy term. Therefore we need to require that it drop out, which together with the requirement on Q to be transverse-traceless, amounts to the conditions on $h_{\mu\nu}$ as

$$h^{\mu}_{\ \mu} = 0 \qquad\qquad \partial_\mu h^{\mu}_{\ \nu} = 0 \tag{7.103}$$

which are exactly the Fierz-Pauli conditions (5.77b) and (5.77c). R. Feynman, in his Caltech lectures on gravitation, arrives at the same result, however in a somewhat different chain of arguments.

7.6.2 *Altering the Geometry*

The route towards GR suggests that gravitation is to be described in terms of a Riemannian spacetime, which is locally a Minkowski manifold with a symmetric metric and a metric-compatible connection. The request that the laws of special relativity are to hold locally already restricts the choice of other geometries. Before having a look on alternate geometries we ask which kinetic terms can possibly be part of the gravitational Lagrangian.

The essential building block in GR is the Riemann curvature tensor (RCT). Since it is a fourth rank tensor, a great many contractions and multiplications are possible. It is known ([576]) that one can construct 14 real algebraic invariants from the RCT. The simplest one is just the curvature scalar R. Others contain up to nine powers in the RCT. If one additionally allows in the construction of scalars also covariant derivatives of the RCT, there are 60 invariants "of the second kind" (an example being $D_\mu R^{\mu\nu} D_\nu R$) and 126 "of the third kind". However, by far not all of these mathematically sensible scalars do qualify as possible kinetic terms in a Lagrangian for gravitation. Those containing derivatives of the curvature would lead to field equations of order higher than four, so they are reasonably taken out of consideration. What about the condition of ending with second-order field equations? The answer relates to the so-called Lanczos-Lovelock gravity theories: These are the most general actions for a metric theory satisfying the criteria of general covariance and second-order field equations in D dimensions. A Lanczos-Lovelock Lagrangian in a D-dimensional Riemannian manifold is a finite polynomial, where each term is a definite expression of powers in the curvature, and where the number of these terms ranges up to $[\frac{D}{2}]$. Although these kind of theories for various reasons received a lot of attention in arbitrary dimensions, let us see its implication for $D = 4$. The Lanczos-Lovelock action is

$$S_{LL} = \int d^4x \sqrt{-g}\{\alpha_0 + \alpha_1 R + \alpha_2 G\}. \tag{7.104}$$

Here the α_i are constants, R is the curvature scalar and

$$G = R_{\mu\nu\varrho\sigma} R^{\mu\nu\varrho\sigma} - 4 R_{\mu\nu} R^{\mu\nu} + R^2 \tag{7.105}$$

is the so-called Gauß-Bonnet (topological) invariant, which can be written as the derivative of another tensor[38] and thus does not contribute to the field equations. As a boundary term it possibly plays a role in defining the energy of a gravitational system. With an appropriate choice of the constants the action (7.104) is none other than the Einstein action including a cosmological constant Λ:

$$S_E = \frac{1}{2\kappa} \int d^4x \sqrt{-g}\{R - 2\Lambda\}. \tag{7.106}$$

Einstein augmented his original field equation by a term with the cosmological constant in order to achieve a static solution for the universe. After E. Hubble discovered the expansion velocity of the galaxies, this constant was assumed to be zero, but in the late 1990's it became re-established as one possible cause of dark energy.

Already in 1925 Einstein investigated geometries with non-symmetric metrics and came back to it later with his collaborators E. Straus and B. Kaufman, see e.g. [162]

[38] Astoundingly it is highly non-trivial to find the explicit expression, various incorrect ones have appeared in the literature. For a thorough analysis see [569]; see also Appendix F.3.2 for the appropriate expressions in terms of differential forms.

(Appendix II, "Relativistic theory of the non-symmetric field"). The asymmetric part of the metric defines an extra field in the theory and there are two possible contractions of the curvature, whereas for symmetric metrics there is only the Ricci tensor. This opens some more freedom in building gravitational kinetic terms in the Lagrangian. Einstein's hope was to integrate by this the electromagnetic interaction into a unified field theory. For a recent approach of non-symmetric metrics to the explanation of dark matter, see [301].

Einstein-Cartan Gravity

In 1922 the French mathematician Elié Cartan investigated manifolds with a non-symmetric connection [79]. He pointed out that since the metric and the torsion (i.e. the antisymmetric part of the connection) are independent objects on a manifold, they both qualify as dynamical fields in a theory of gravitation. Allowing for non-vanishing torsion means to ponder on Riemann-Cartan geometries, and obvious generalizations of GR are actions of the form

$$S_T = \frac{1}{2\kappa} \int d^4x \; \mathcal{L}_{TG}(g, \partial g, T, \partial T) + \int d^4x \; \mathcal{L}_M(Q, DQ, g) \tag{7.107}$$

where $\mathcal{L}_{TG}$ is a gravitational scalar density depending on the metric, the torsion T and their derivatives, and $\mathcal{L}_M$ is the matter part, where the matter fields Q^α are coupled minimally to gravitation by the covariant derivative D. In introducing the metric energy-momentum tensor $\sigma^{\mu\nu}$ and the so-called spin-momentum potential $\mu_\lambda^{\ \mu\nu}$ (following the established notation of the F.W. Hehl School [262], [263], [265] except for the positioning of indices and a definition of torsion and contortion differing by a factor of 2)

$$\sigma_{\mu\nu} := \frac{2}{\sqrt{-g}} \frac{\delta S_M}{\delta g^{\mu\nu}} \qquad\qquad \mu_\lambda^{\ \mu\nu} := \frac{2}{\sqrt{-g}} \frac{\delta S_M}{\delta T^\lambda_{\ \mu\nu}} \tag{7.108}$$

the $10 + 24$ gravitational field equations become

$$\frac{\delta S_{TG}}{\delta g^{\mu\nu}} = -\frac{1}{2}\sqrt{-g}\,\sigma^{\mu\nu} \qquad\qquad \frac{\delta S_{TG}}{\delta T^\lambda_{\ \mu\nu}} = -\frac{1}{2}\sqrt{-g}\,\mu_\lambda^{\ \mu\nu}. \tag{7.109}$$

Further useful tensors are the spin-momentum tensor

$$\tau_\lambda^{\ \mu\nu} := \frac{1}{\sqrt{-g}} \frac{\delta S_M}{\delta K^\lambda_{\ \mu\nu}} \qquad \text{or} \qquad \tau^{\lambda\mu\nu} = \mu^{[\mu\lambda]\nu} \tag{7.110}$$

and the total energy-momentum tensor

$$\Sigma^{\mu\nu} = \sigma^{\mu\nu} - \overset{*}{\nabla}_\lambda \mu^{\lambda\mu\nu} \quad \text{with} \quad \overset{*}{\nabla}_\lambda := \nabla_\lambda + T_\lambda \quad \text{and} \quad T_\lambda = T^\mu_{\ \lambda\mu}. \tag{7.111}$$

Which explicit forms of $\mathcal{L}_{TG}$ are allowed in order that the field equations are at most of second-order? The Lovelock theorem was generalized by P. Von der Heyde [523] from Riemann to Riemann-Cartan theories, with the result that in four dimensions this is again essentially the curvature scalar, i.e. $\sqrt{-g}\tilde{R}$, aside from a cosmological term, and assuming that no other fundamental constants of mass dimension $\neq 1$ except κ are allowed. With the choice of the action

$$S_{TG} = \frac{1}{2\kappa}\int d^4x\ \sqrt{-g}\tilde{R} := S_{EC}$$

one arrives at the Einstein-Cartan theory. Although formally looking like the Hilbert action, one must have in mind that here the curvature scalar is not built with the Levi-Civita connection, but with a connection that is independent of the metric (therefore as a reminder the "tilde" over the R). The relation between these two curvature scalars is given by (7.27). In varying the action S_{EC} with respect to $g^{\mu\nu}$ and $T^{\lambda}_{\ \mu\nu}$ one obtains with a calculation essentially similar to the one for the GR case the result:

$$\frac{1}{\sqrt{-g}}\frac{\delta\sqrt{-g}\,\tilde{R}}{\delta g^{\mu\nu}} = \tilde{G}_{\mu\nu} + 2S_{\lambda}^{\ \rho\sigma}\frac{\delta\Gamma^{\lambda}{}_{\sigma\rho}}{\delta g^{\mu\nu}} \qquad \frac{1}{\sqrt{-g}}\frac{\delta\sqrt{-g}\,\tilde{R}}{\delta T^{\lambda}_{\ \mu\nu}} = 2S_{\tau}^{\ \rho\sigma}\frac{\delta\Gamma^{\tau}{}_{\sigma\rho}}{\delta T^{\lambda}_{\ \mu\nu}}.$$

Here the tensor S is defined in terms of the torsion as

$$S^{\lambda}_{\ \mu\nu} = \frac{1}{2}(T^{\lambda}_{\ \mu\nu} + \delta^{\lambda}_{\mu}T^{\rho}_{\ \nu\rho} - \delta^{\lambda}_{\nu}T^{\rho}_{\ \mu\rho}).$$

The variations of the connections appearing in these expressions can be derived using the relation (7.26) as

$$\frac{\delta\Gamma^{\lambda}{}_{\sigma\rho}}{\delta g^{\mu\nu}} = \frac{\delta\left\{{\lambda \atop \sigma\rho}\right\}}{\delta g^{\mu\nu}} \qquad \frac{\delta\Gamma^{\tau}{}_{\sigma\rho}}{\delta T^{\lambda}_{\ \mu\nu}} = -\frac{\delta K^{\tau}{}_{\sigma\rho}}{\delta T^{\lambda}_{\ \mu\nu}}.$$

After a quite lengthy calculation (see e.g. [262]) one eventually arrives at the field equations

$$\tilde{G}^{\mu\nu} = -\kappa\Sigma^{\mu\nu} \tag{7.112a}$$

$$S^{\mu\nu\lambda} = -\kappa\tau^{\mu\nu\lambda} \tag{7.112b}$$

for the Einstein-Cartan theory in the presence of matter - with the total energy momentum (7.111) and the spin momentum (7.110) as source terms[39]. These field equations were proposed by T.B.W. Kibble and by D. Sciama in 1962. Therefore the theory is sometimes found under the acronym ECKS.

Whereas (7.112a) is a set of genuine differential equations, the part (7.112b) represent algebraic relations for the torsion components. Thus in the Einstein-Cartan theory, torsion does not propagate, but is locally determined by the value of the spin-momentum. This allows one to eliminate the torsion completely, and another lengthy calculation then transforms (7.112a) into an expression

$$G^{\mu\nu} = -\kappa\hat{T}^{\mu\nu}$$

[39] Observe that neither the Einstein(-Cartan) tensor $\tilde{G}$ nor the energy-momentum tensor Σ are presumed to be symmetric.

with the V_4 Einstein tensor on the left hand side (the one without the tilde) and with a modified energy-momentum tensor depending on the spin-momentum potentials in an intricate manner.

What is the physical interpretation of torsion, that is which matter fields are a source of torsion? Without going into details we observe that the spin-momentum tensor (7.110) vanishes if the matter action does not depend on the contortion. The contortion is according to (7.33) identical to the spin connection up to an additional term depending on the tetrads. Thus those (microscopic) matter actions which do not depend on the spin connection can not be sources of torsion. According to the results presented in Sect 7.4.2. the scalar and gauge vector field Lagrangians are independent of the spin connection. This is not the case for spinor fields; see (7.49). (The same applies to spin-$\frac{3}{2}$ fields, but this is another story–told in the chapter on Supergravity.) Possible experimental indications of torsion come up for discussion in [386] within the broader context of whether and how spin is coupled to gravity.

7.6.3 Gravitation as a Gauge Theory

In virtue of the successful description of the particle world in terms of gauge-field theories, one may reasonably ask whether the gravitational interaction can also be formulated as a gauge theory, or to be more careful, whether suitably defined gauge principles can be implemented or discovered in gravitational theories.

We saw that the interaction of gravitation with other fields is accomplished by minimal coupling. Like in Yang-Mills theories, one needs to introduce a covariant derivative. This in turn necessitates a connection A, that is a gauge field. The analogy to the familiar gauge theories would be complete if eventually one finds kinetic terms for these "gravitational" connection fields. On a more formal and abbreviated level, the global symmetry of the matter field Lagrangian $\mathcal{L}_M(\Phi, d\Phi)$ gives rise to a conserved current via Noether's first theorem. Enhancing the rigid symmetry to a local symmetry becomes possible by replacing derivatives in the Lagrangian by covariant derivatives $D(A)$ defined in terms of the connection. This amounts to a coupling of the current with the connection and leads in turn to the covariant conservation of the current $J = \partial\mathcal{L}_M/\partial A$ as $D(A)J = 0$. It immediately suggests itself to use as the "gravitational" gauge fields a geometric connection in one of the Riemann-Cartan manifolds described in Sect. 7.3. But already at this stage we perceive a disparity between the Yang-Mills type approach in the standard model and a gauge approach for gravitation: It is not only a question of the symmetry group, but also a question of the geometry, since different geometries are possible and within each geometry we can have different connections. Aside from these many options, it turns out that the choice of a kinetic term is less restricted than in the Yang-Mills case. By this you realize that the topic of "Gravitation as a gauge theory" is full of subtleties, and here only the essentials can be dealt with. For the essential publications on the fields and their positions in a classification of gauge theories of gravitation, see the Reader [50]. Appendix E. 5.4. addresses the gauge aspect of gravity from the fibre bundle perspective, which more clearly exhibit why and where a gauge theory of gravitation differs from a Yang-Mills gauge theory.

Poincaré Gauge Theories

All matter Lagrangians in relativistic field theories are invariant with respect to global Poincaré transformations. Thus it is quite natural to ask whether and how the Poincaré group can serve as a gauge group. And indeed, 'Poincaré gauge theory' became a widely-investigated field. It started in 1956 when R. Utiyamah considered only the homogeneous part, that is the Lorentz group. With some mild assumptions and by choosing the most simple kinetic term (namely the Riemann curvature scalar), he recovered essentially GR. This procedure is not completely convincing, since the conserved current is the angular momentum, which is not alone representing the source of gravity. Thus it was stressed by T. Kibble [315] that one should consider the full Poincaré group. This meant to introduce also gauge fields associated to translations together with their field strength. As it turned out–and will be shown explicitly below–the latter are related to torsion. Choosing a Lagrangian linear in the (Riemann-Cartan) curvature scalar, Kibble arrived at the Einstein-Cartan version of gravity. These pioneering approaches of gauging the Poincaré group are now conceptually satisfactorily understood and go by the name 'Poincaré gauge theories' (PGT). For an early overview talk see [264], and for a textbook exposition see [49], whom I largely follow in its logic approach towards PGT.

The starting point is field theory defined in terms of a Lagrangian $\mathcal{L}_M(\Phi, \partial\Phi)$ with a set of fields Φ which realizes a representation of the Poincaré group. As expounded before, global Poincaré transformations, defined by $\delta x^\mu = \epsilon^\mu{}_\nu x^\nu + \epsilon^\mu$, where $\epsilon^{\mu\nu} = -\epsilon^{\nu\mu}$ and ϵ^μ are ten constant infinitesimal parameters. The matter fields transform according to[40]

$$\delta\Phi = \frac{1}{2}\epsilon^{\mu\nu}(\Sigma_{\mu\nu}\Phi) \qquad\qquad \bar{\delta}\Phi = (\epsilon^\mu P_\mu + \frac{1}{2}\epsilon^{\mu\nu}M_{\mu\nu})\Phi$$

where $\Sigma_{\mu\nu}$ is the spin matrix, and $P_\mu = -\partial_\mu$ and $M_{\mu\nu} = (x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu}$ are the group generators.

In the following, it turns out to be more suitable to consider as infinitesimal parameters $\epsilon^{\mu\nu}$ and $\xi^\mu = \epsilon^\mu{}_\nu x^\nu + \epsilon^\mu$ instead of the pair $(\epsilon^{\mu\nu}, \epsilon^\mu)$. For x-independent parameters $(\epsilon^{\nu\mu},\ \epsilon^\mu)$, this distinction is of course unimportant, but for local Poincaré transformations it is conceivable that $\xi^\mu = 0$ whereas $\epsilon^{\mu\nu} \neq 0$. This split of the parameter set allows to distinguish between coordinate transformations (characterized by ξ^μ) and Lorentz rotations $(\epsilon^{\mu\nu})$. To make this even more recognizable, holonomic and anholonomic indices are introduced–although in Minkowski space this distinction is superfluous. Thus the Lorentz transformations are explicitly given by ϵ^{KL} and field derivatives are understood as $\partial_K\Phi$. Again, at this stage, things are

[40] This differs from previous expressions, since I adopt the conventions of M. Blagojević [49]. His definitions of Σ, L, P differ from those used elsewhere in this text by factors (-i).

trivial since everything is in Minkowski space where it is always possible to choose $\partial_\mu = \delta^K_\mu \partial_K$. With this understanding the fields vary as

$$\bar{\delta}\Phi = (\xi^\mu P_\mu + \frac{1}{2}\epsilon^{IJ}\Sigma_{IJ})\Phi := \mathcal{P}\Phi. \tag{7.113}$$

Since $\bar{\delta}\partial_K\Phi = \partial_K\bar{\delta}\Phi$, we find

$$\bar{\delta}\partial_K\Phi = \mathcal{P}\partial_K\Phi + (\partial_K\xi^\mu)P_\mu\Phi \tag{7.114}$$

where for global transformations $\partial_K\xi^\mu = \epsilon^\mu{}_K = \delta^\mu_L\epsilon^L{}_K$.

The invariance identity (3.43) on the Lagrange density can in case of Poincaré symmetry be written as

$$0 \equiv \bar{\delta}\mathcal{L}_M + (\partial_\mu\mathcal{L}_M)\xi^\mu + \mathcal{L}_M\partial_\mu\xi^\mu. \tag{7.115}$$

For global transformations, the last term vanishes because of $\partial_\mu\delta x^\mu = \partial_\mu\xi^\mu = \epsilon^\mu{}_\mu = 0$. The identity (7.115) implies the conservation of energy-momentum and angular-momentum currents (see Subsect. 3.3.2). It also expresses the fact that the Lagrangian is a Lorentz scalar, namely $\delta\mathcal{L}_M = 0$.

Localizing Poincaré Symmetry by Introducing Gauge Fields: The Approach by Kibble and Sciama

In the transition from global to spacetime-dependent Poincaré transformations, the Lagrangian $\mathcal{L}_M$ is no longer invariant. This phenomenon sounds familiar, and we can imagine its source: derivatives no longer transform in a desired way. Indeed, (7.114) contains terms with derivatives of $\epsilon^{IJ}(x)$ and $\xi^\mu(x)$. And we also remember the Yang-Mills recipe in introducing potentials A_K in order to define a covariant derivative $D_K = \partial_K + A_K$ which behaves properly, namely just like (7.114), as

$$\bar{\delta}(D_K\Phi) \stackrel{!}{=} \mathcal{P}D_K\Phi + (\partial_K\xi^\mu)P_\mu\Phi.$$

This condition is then used to determine the transformation behavior of the potentials:

$$\bar{\delta}A_K = -\partial_K\sigma + [\mathcal{P}, A_K] - (\partial_K\xi^L)A_L \qquad \text{with} \qquad \sigma := \frac{1}{2}\epsilon^{IJ}\Sigma_{IJ}.$$

The gauge potentials $A_K{}^\mu$ and $A_K{}^{IJ}$ are to be identified as components with respect to the symmetry generators

$$A_K = A_K{}^\mu P_\mu + \frac{1}{2}A_K{}^{IJ}\Sigma_{IJ}. \tag{7.116}$$

At this stage of the "gauging game" there is a source of confusion and/or interpretation: On the one hand, $A_K{}^\mu$ looks like the gauge potential for translations (mediated by P_μ). On the other hand, the infinitesimal transformation in (7.113) goes with ξ^μ, i.e. with general coordinate transformations. So it is debatable of whether we arrive at a gauge theory of the Poincaré group, or whether the final theory is a gauge theory of the Lorentz and the diffeomorphism groups. This confusion and/or interpretation issue, having its origin in the fact that localized translations are diffeomorphisms, is up to today not completely resolved in the debate over in which sense gravity is a gauge theory.

The explicit $\bar{\delta}$-variations of the gauge potentials are found to be

$$\bar{\delta}A_K^{\ \mu} = \epsilon^L{}_K h_L^{\ \mu} - h_K^{\ \lambda}\partial_\lambda \xi^\mu - \xi^\lambda \partial_\lambda A_K^{\ \mu} \tag{7.117a}$$

$$\bar{\delta}A_K^{\ IJ} = -h_K^{\ \mu}\partial_\mu \epsilon^{IJ} - \epsilon^L{}_K A_L^{\ IJ} + \epsilon^I{}_K A_K^{\ KJ} + \epsilon^J{}_K A_K^{\ IJ} - \xi^\lambda \partial_\lambda A_K^{\ IJ} \tag{7.117b}$$

where $h_K^{\ \mu}$ arises as an object which for the moment only serves as an abbreviation:

$$h_K^{\ \mu} := \delta_K^\mu - A_K^{\ \mu} \tag{7.118}$$

with

$$\bar{\delta}h_K^{\ \mu} = -\epsilon^L{}_K h_L^{\ \mu} + h_K^{\ \lambda}\partial_\lambda \xi^\mu - \xi^\lambda \partial_\lambda h_K^{\ \mu}.$$

The covariant derivative can be expressed by using the $h_K^{\ \mu}$ as

$$D_K = h_K^{\ \mu}\partial_\mu + \frac{1}{2}A_K^{\ IJ}\Sigma_{IJ} \qquad \text{from which} \qquad \partial_\nu = b^K{}_\nu\left(D_K + \frac{1}{2}A_K^{\ IJ}\Sigma_{IJ}\right) \tag{7.119}$$

with the inverse $b^K{}_\nu$ of $h_K^{\ \mu}$ fulfilling $b^K{}_\nu h_K^{\ \mu} = \delta^\mu_\nu$, $b^K{}_\mu h_L^{\ \mu} = \delta^K_L$. For later purposes we register

$$\bar{\delta}b^K{}_\mu = \epsilon^K{}_L b^L{}_\mu - b^K{}_\lambda \partial_\lambda \xi^\mu - \xi^\lambda \partial_\lambda b^K{}_\mu. \tag{7.120}$$

Let $\mathcal{L}_M := \mathcal{L}_M(\Phi, D\Phi)$ be the original Lagrangian in which the ordinary derivatives are replaced by the covariant derivatives. It obeys $\bar{\delta}\mathcal{L}_M + \xi^\mu \partial_\mu \check{\mathcal{L}}_M = 0$. This does not suffice to fulfill (7.115), because there is an extra term $\check{\mathcal{L}}_M \partial_\mu \xi^\mu$ missing, or in other words, the Lagrangian is a scalar but not a scalar with the proper density. Thus we modify the $\check{\mathcal{L}}_M$ as

$$\tilde{\mathcal{L}}_M = \beta \check{\mathcal{L}}_M$$

where β is found to obey $\bar{\delta}\beta + \partial_\mu(\xi^\mu \beta) = 0$ in order to satisfy (7.115). By using (7.120) it can be shown that

$$\beta = [\det b^K{}_\mu] := b$$

up to a multiplicative factor that can be chosen in such a way that $\beta \to 1$ if $h_K^{\ \mu} \to \delta_K^{\ \mu}$.

Take as an example the Lagrangian for a free scalar field in Minkowski space: $\mathcal{L}_\varphi = \frac{1}{2}(\eta^{IJ}\partial_I \varphi \partial_J \varphi - m^2\varphi^2)$. The previous recipe for localizing the Poincaré symmetry leads to

$$\tilde{\mathcal{L}}_\varphi = \frac{1}{2}b(\eta^{IJ}D_I\varphi D_J\varphi - m^2\varphi^2) = \frac{1}{2}\sqrt{-\tilde{g}}(\tilde{g}^{\mu\nu}\tilde{\partial}_\mu\varphi\tilde{\partial}_\nu\varphi - m^2\varphi^2) \tag{7.121}$$

with $\tilde{g}^{\mu\nu} := \eta^{IJ} h_I^{\ \mu} h_J^{\ \nu}$, $\tilde{g} := \det(\tilde{g}^{\mu\nu})$ and $D_K\varphi := h_K^{\ \mu}\tilde{\partial}_\mu\varphi$. This looks like a theory of a scalar field in a spacetime with metric $\tilde{g}^{\mu\nu}$!

For a full-fledged Yang-Mills type gauge theory, we need to introduce kinetic terms for the gauge potentials into the Lagrangian. These are to be built from field strength which for instance can be found by calculating the commutator of two covariant derivatives, each of them expressed as (7.119)

$$[D_K, D_L]\Phi = h_K^{\ \mu}(\partial_\mu h_L^{\ \nu})\partial_\nu\Phi + \frac{1}{2} h_K^{\ \mu}(\partial_\mu A_K^{\ IJ})\Sigma_{IJ}\Phi + \frac{1}{2} A_K^{\ IJ} A_L^{\ NM}\Sigma_{IJ}\Sigma_{NM}\Phi \quad - (K \leftrightarrow L).$$

Instead of performing the full calculation, let us inspect the general structure: The first term gives rise to terms containing derivatives of the $h_K^{\ \mu}$. In expressing the derivative $\partial_\nu\Phi$ by $D_K\Phi$ further terms proportional to $A_K^{\ IJ}$ do arise. The second term in the commutator above gives rise to terms proportional to Σ_{IJ} with derivatives of the $A_K^{\ IJ}$. The last term in the commutator is quadratic in the $A_K^{\ IJ}$ and (together with the $(K \leftrightarrow L)$ piece) constitutes the commutator of two Σ 's which is again a linear combination of the Σ's. The result is

$$[D_K, D_L]\Phi = \frac{1}{2} F^{IJ}{}_{KL}\Sigma_{IJ}\Phi - F^I{}_{KL} D_I\Phi \tag{7.122}$$

where

$$F^{IJ}{}_{KL} = h_K^{\ \mu} h_L^{\ \nu} F^{IJ}{}_{\mu\nu}, \qquad F^{IJ}{}_{\mu\nu} = \partial_\mu A^{IJ}{}_\nu + A^I{}_{L\nu} A^{LJ}{}_\nu - (\mu \leftrightarrow \nu) \tag{7.123a}$$

$$F^I{}_{KL} = h_K^{\ \mu} h_L^{\ \nu} F^I{}_{\mu\nu} F^I{}_{\mu\nu} = D_\mu b^I{}_\nu - D_\nu b^I{}_\mu. \tag{7.123b}$$

The rotational field strength $F^{IJ}{}_{KL}$ and the translational field strength $F^I{}_{KL}$ are constituents in the Lagrangian $\mathcal{L}_P$ for the Poincaré gauge fields. The Lagrangian for the full theory has the form

$$\tilde{\mathcal{L}} = b\mathcal{L}_P(F^{IJ}{}_{KL}, F^I{}_{KL}) + \tilde{\mathcal{L}}_M(\Phi, D_K\Phi).$$

Relating the Gauge Structures to Geometries

Up to this point, the previous Lagrangian defines a gauge theory in Minkowski space with gauge fields $b^K{}_\mu$ and $A^{IJ}{}_\nu$. However, with regard to gravitational theories, there is striking analogy between the Poincaré group gauge structure and Riemann-Cartan geometries. This analogy (visible for instance in comparing (7.123) with (7.37) leads to the identifications of the $b^K{}_\mu$ with the tetrad and of the $A^{IJ}{}_\nu$ with the spin connection:

$$b^K{}_\mu \leftrightarrow e^K{}_\mu \qquad\qquad A^{IJ}{}_\nu \leftrightarrow \omega^{IJ}{}_\nu.$$

The field strengths corresponding to the gauge fields are consequently the curvature $\tilde{R}^{IJ}{}_{\mu\nu}$ and the torsion $T^I{}_{\mu\nu}$, with the result that this gauge theory is non-distinguishable from a U_4 geometry.

Nevertheless, we ask again whether this final theory is "really" a gauge theory of the Poincaré group. A straight application of the gauge recipe would amount to the following: Given the Lie-algebra of the Poincaré group generators $X^a = \{P_I, M_{IJ}\}$. Introduce Lie-algebra valued vector fields:

$$\hat{A}_\mu = \hat{A}^a_\mu X_a = \hat{e}^I_\mu P_I + \frac{1}{2}\hat{\omega}^{IJ}{}_\mu M_{IJ}$$

with gauge potentials $\hat{A}^a_\mu$ and covariant derivatives $\hat{D}_\mu = \partial_\mu + \hat{A}_\mu$. Each gauge potential has associated local gauge generators

$$\theta = \theta^a X^a = \epsilon^I P_I + \frac{1}{2}\lambda^{IJ} M_{IJ}$$

and the gauge potentials transform as

$$\delta_\theta \hat{A}^a_\mu = \hat{D}_\mu \theta^a = \partial_\mu \theta^a + \hat{A}^b_\mu \theta^c f^{cba} = \partial_\mu \theta^a + \hat{A}^b_\mu \epsilon^N f^{Nba} + \hat{A}^b_\mu \epsilon^{NM} f^{(NM)ba}$$

where f^{cba} are the structure constants of the Poincaré algebra. Explicitly these become for the translation and the Lorentz rotation gauge potentials

$$\begin{aligned}
\delta_\theta \hat{e}^I_\mu &= \partial_\mu \epsilon^I + \epsilon^K \hat{\omega}^I{}_{K\mu} + \lambda^I{}_K \hat{e}^K_\mu \\
\delta_\theta \hat{\omega}^{IJ}{}_\mu &= \partial_\mu \lambda^{IJ} + \lambda^I{}_K \hat{\omega}^{KJ}{}_\mu + \lambda^J{}_K \hat{\omega}^{IK}{}_\mu .
\end{aligned}$$

We notice that this transformation behavior of the gauge algebra objects $\hat{e}^I_\mu$ and $\hat{\omega}^{IJ}{}_\mu$ is clearly different from the transformation behavior of the tetrads and the spin connections as per (7.38, 7.39). Another hint of a disparity is the transformation of the metric under translations, namely as

$$\begin{aligned}
\delta g_{\mu\nu} = \delta(e^I_\mu e^J_\nu \eta_{IJ}) &= \Big[(D_\mu \epsilon^I) e^J_\nu + (D_\nu \epsilon^I) e^J_\mu\Big]\eta_{IJ} \\
&= \Big[(D_\mu \epsilon^\lambda) g_{\lambda\nu} + \epsilon^\lambda (D_\mu e^I_\lambda) e^J_\nu \eta_{IJ}\Big] + (\mu \leftrightarrow \nu) \\
&= (D_\mu \xi_\nu + D_\nu \xi_\mu) + \xi_\lambda (T_{\nu\mu\lambda} + T_{\mu\nu\lambda}).
\end{aligned}$$

Thus, in this kind of "gauge theories of the Poincaré group" one has to request that the torsion vanishes if one wants to identify local translations with diffeomorphisms. However, now the Riemann-Cartan geometric interpretation is lost[41]. Obviously there are different notions of "gauge theory of gravity". Observe that in the first "Kibble-Sciama" approach the translation potential is A^μ_K, related to the geometric tetrad field b^K_μ, which is the inverse of $h^\mu_K = \delta^\mu_K - A^\mu_K$, see (7.118). On the other hand, in the latter approach, the tetrad is directly treated as the gauge potential of translations. This approach being only compatible with a geometry if the torsion vanishes, means that the Kibble-Sciama approach is more general.

[41] I thank M. Blagojević for pointing this out to me.

The $R + T + T^2$ Family of PGT Lagrangians

The Poincaré approach opens a large arena for building alternative gravitational theories. What are its restrictions?

- The theory should obey the principle of equivalence. This, as shown in [524], does not restrict the U_4 theory.
- Any kinetic term built from the curvature and torsion tensor must be a Lorentz scalar. Contrary to Yang-Mills theories which for compact and semi-simple groups yield a unique kinetic term in field strength (aside from a possible Θ-term), in Poincaré gauge field theories because of the qualitative different character of the group we are faced with a variety of possible terms, e.g. $\tilde{R}_{\mu\nu}{}^{IJ} e^{\mu}_{I} e^{\nu}_{J}$, $\tilde{R}_{\mu\nu}{}^{I} \tilde{R}^{\mu\nu}{}_{I}, \ldots$
- Field equations should be at most of second order for the tetrads and the spin connection. The most general Lagrangian (respecting also parity invariance) was shown [259] to have the generic form

$$\mathcal{L}_{\mathrm{PGT}} \sim (\text{const}) + (\text{curvature}) + (\text{curvature})^2 + (\text{torsion})^2. \tag{7.124}$$

  Here I refrain from giving the detailed expression. You yourself will be able to write down all those scalars that can be built by contracting (powers of) the curvature and the torsion tensor. Bookkeeping of all possible terms is eased by the use of differential forms, in any case. An example is treated in App. F.3.2. Aside from the (gravitational) constant there is one term linear in the curvature scalar, six terms containing the square of the curvature tensor, and three terms quadratic in the torsion. As a matter of fact the Lagrangian can be written as a linear combination of only ten terms plus a derivative term representing the Gauß-Bonnet invariant.
- The theory should have Einstein' theory as an approximation. This essentially amounts to requiring that the term linear in the curvature scalar is present.
- The field equations should have a well-posed initial value problem. For a theory as complicated and nonlinear as (7.124) this property cannot easily be proven or disproven.

Notice that the constants multiplying the ten terms in (7.124) are to be interpreted as new fundamental constants of nature, some of them even with unusual mass dimensions. Since the tetrads have mass dimension $[e] = 0$ and the spin connections $[\omega] = +1$ the curvature and the torsion have $[\text{curvature}] = [\partial\omega] = +2$, $[\text{torsion}] = [\partial e] = +1$. Therefore, the coefficients of the term linear in the curvature and those quadratic in the torsion do have the same mass dimension, namely (D-2), the terms quadratic in the curvature do have mass dimension (D-4). In $D = 4$ the mass dimension of the coefficient in front of curvature is $+2$, which is the mass dimension of κ^{-1}. The coefficients of $(\text{curvature})^2$ have zero mass-dimensions: This looks promising for renormalizability, but a model with only $(\text{curvature})^2$ terms can not reproduce GR.

The ten-parameter family of PGT-Lagrangians (7.124) is rich in subtleties, depending on the choice of the constants in front of the nine invariants, and not even all aspects are understood. This concerns especially the unfolding of extra symmetries for specific choices of constants, their interpretation as gauge symmetries, and the appearance of massless and massive torsion modes (*tordions*) in the particle spectrum of the theories [51], [493]. In the late 1990's PGT came under heavy attack among other by indications that some PGT's may lack a well-defined initial-value problem. This even led to abandon PGT's at all (except for GR and the EC theory.) However, for some parameter choices PGT's could be "saved", see [261]. A Hamiltonian analysis [572] supports the conclusion that only two dynamic torsion modes (so-called "scalar torsion" modes) are physically acceptable.

If one takes in (7.124) only the term linear in the curvature one is back at the Einstein-Cartan (EC) theory

$$\mathcal{L}_{\mathrm{EC}} \propto \epsilon^{\mu\nu\rho\sigma} \epsilon_{IJKL} e^K{}_\rho e^L{}_\sigma \tilde{R}^{IJ}{}_{\mu\nu}.$$

But here again, we can observe that this is not a Lagrangian for the gauged Poincaré group: The Einstein-Cartan Lagrangian is indeed (quasi-)invariant under diffeomorphism and local Lorentz transformations, but under translations its change comes out to be

$$\delta_\epsilon \mathcal{L}_{\mathrm{EC}} \propto \partial_\rho \left(\epsilon^{\mu\nu\rho\sigma} \epsilon_{IJKL} \epsilon^K e^L{}_\sigma \tilde{R}^{IJ}{}_{\mu\nu}\right) + \epsilon^{\mu\nu\rho\sigma} \epsilon_{IJKL}\, e^K\, T^L_{\lambda\rho}\, \tilde{R}^{IJ}{}_{\mu\nu}.$$

This is peculiar to even dimensions. In odd dimensions, the Einstein-Cartan theory is a genuine gauge theory of either the Poincaré or the de Sitter groups since it can be formulated as a Chern-Simon gravity theory; more about this in Appendix F.3.2.

Another class of theories are the $R + T^2$ models. In these, like in the EC theory, torsion does not propagate, but is algebraically given by the local spin momentum potential, similar to (7.112b). In order to obtain propagating torsion one needs to include the R^2 terms. The viable scalar-torsion mode model mentioned before corresponds to specific parameter choices for the R, R^2, and T^2 terms, where the relation among the three T^2 terms is fixed. This specific PGT was shown to be able to explain the accelerated expansion of the universe [473]! Among the Lagrangians (7.124) is also a Yang-Mills type term $\tilde{R}_{IJKL}\tilde{R}^{IJKL}$, investigated as the 'Stephenson-Kilmister-Yang' theory of gravity in the literature.

Teleparallel Models

Teleparallel (or coframe) models are defined in T_4 spacetimes, characterized by connections with vanishing curvature (sometimes called Weitzenböck connections). The teleparallel or Weitzenböck Lagrangian is

$$\begin{aligned}
\mathcal{L}_{\mathrm{W}} &= b\mathcal{L}_{\mathrm{T}} + \lambda_{IJ}{}^{\mu\nu}\, \tilde{R}^{IJ}{}_{\mu\nu} + \tilde{\mathcal{L}}_{\mathrm{M}} \\
\mathcal{L}_{\mathrm{T}} &= \alpha(A\, T_{IJK}T^{IJK} + B\, T_{IJK}T^{JIK} + C\, T^K_{KT})
\end{aligned}$$

where $\lambda_{IJ}{}^{\mu\nu}$ are Lagrange multipliers, which are introduced to ensure that the curvature tensor vanishes, and α, A, B, C are numerical constants.

Because of the vanishing curvature, one can–by a gauge choice–transform the spin connections away (see e.g. Sect. 3.3 in [49]), so that the torsion (7.37a) becomes

$F^I{}_{\mu\nu} = \overset{\circ}{T}{}^I{}_{\mu\nu} = \frac{1}{2}(\partial_\mu e^I_\nu - \partial_\nu e^I_\mu)$. This can be interpreted as the field strength of an Abelian gauge theory with gauge potentials e^I_ν. Indeed, the teleparallel models can consistently be formulated as a gauge theory of the translation group with Lagrangian

$$b\mathcal{L}_T = \mathcal{L}_{Tr} = \frac{1}{2\kappa}\frac{1}{4}F^I{}_{\mu\nu}F^J{}_{\rho\sigma}C^{\mu\nu\rho\sigma}_{IJ} \qquad \text{with} \qquad C^{\mu\nu\rho\sigma}_{IJ} = \eta_{IJ}\eta^{\mu\rho}\eta^{\nu\sigma}.$$

Requiring that $\mathcal{L}_W$ gives the same results as GR in the linear weak-field approximation, the condition on the coefficients is $C = -1, 2\,A + B = 1$; see [392]. Now, by using (7.27), the first term in $\mathcal{L}_\mathrm{W}$ can be rewritten as

$$b\mathcal{L}_T = b\alpha\left\{A\tilde{R} - AR - 4AT^\rho_{\ ,\rho} + (4\,A - 1)T^K_{KT} + (B - 2\,A)T_{IJK}T^{JIK}\right\}.$$

In a teleparallel geometry, by definition $\tilde{R} = 0$, and thus the first term in this expression vanishes. With the parameter choice $A = \frac{1}{4}$, $B = \frac{1}{2}$ the last two terms vanish, also. And with $\alpha A = -1$ one arrives at

$$b\mathcal{L}_\| := bR - 4\partial_\rho(bT^\rho).$$

In other words, for a specific choice of parameters the theory defined by the particular teleparallel Lagrangian $b\mathcal{L}_\|$ is simply GR, called $GR_\|$, the *teleparallel equivalent of GR*. This is quite remarkable: instead of formulating Einstein's theory[42] as a theory in which objects follow geodesics in a curved spacetime with vanishing torsion it can equivalently be considered as a theory in flat spacetime in which objects follow force laws; for this aspect see [3], [10]. The teleparallel theory ceases to be equivalent to GR when microscopic matter fields are present. Since the spin connections can be gauged away, there are no obvious means available to couple spinorial matter.

De Sitter Gauge Theory

Gauge theories based on the de Sitter group $\mathbf{dS_4} = \mathbf{SO(2,3)}$ (see Sect. 3.5.2) also received attention by various authors. The benefit of considering $\mathbf{dS_4}$ is to be seen in that it has a non-degenerate Killing metric, and thus one is able to build YM-type actions quadratic in the field strength. Of course it is possible to formulate a de Sitter gauge type theory by performing steps similar to those in PGT, assuming right from the beginning that spacetime has a de Sitter (instead of a Minkowski structure) and that matter field actions are invariant with respect to global de Sitter transformations. Introducing then according to the YM recipe appropriate covariant derivatives with gauge potentials and field strength, one would eventually arrive at theories structurally similar to the PGT. However, here I present another approach, initiated by S.W. MacDowell and F. Mansouri [356]: Spacetime is still given by Minkowski space, and the de Sitter group $\mathbf{SO(2,3)}$ acts on the matter fields as a

[42] Einstein himself was not only aware of teleparallel gravitation but actively followed this approach in what he called "*Fernparallelismus*" [464].

group of internal symmetries[43]. Let the matter-field Lagrangian be invariant under de Sitter transformations. The Lie algebra of the de Sitter group is[44]

$$[M_{AB}, M_{cd}] = \eta_{ad} M_{bc} - \eta_{AC} M_{bd} + \eta_{bc} M_{ad} - \eta_{bd} M_{AC}$$

with $(a, b, \cdots = 0, 1, 2, 3, 4)$ and $\eta_{ab} = \mathrm{diag}(+, -, -, -, +)$. According to the gauge recipe, introduce gauge fields $A_\mu = \frac{1}{2} A_\mu^{ab} M_{ab}$ (where I use the same symbol for the representation matrices and the gauge algebra generators) and covariant derivatives $D_\mu = \partial_\mu + A_\mu$.

In order to prepare the Inönü-Wigner contraction to the Poincaré group we distinguish as in Sect. 3.5.2 the fourth components of all Lie-algebra valued objects $T_I := \frac{1}{\mathcal{R}} M_{4I}$, $A^I_\mu := \mathcal{R} A_\mu^{4I}$ (where $\mathcal{R}$ carries the meaning of the de Sitter radius), so that the gauge potentials become

$$A_\mu = A^I_\mu T_I + \frac{1}{2} A^{IJ}_\mu M_{IJ} \tag{7.125}$$

where now $I, J = 0, \ldots, 3$. From the de Sitter algebra, we calculate the field strength components

$$F^I_{\mu\nu} = \partial_\mu A^I_\nu + A^I{}_{K\mu} A^K_\nu - (\mu \leftrightarrow \nu) =: \mathcal{R} F^{4I}_{\mu\nu} \tag{7.126a}$$

$$F^{IJ}_{\mu\nu} = [\partial_\mu A^{IJ}_\nu + A^I{}_{K\mu} A^{KJ}_\nu] - \frac{1}{\mathcal{R}^2} A^I_\mu A^J_\nu - (\mu \leftrightarrow \nu). \tag{7.126b}$$

When compared to (7.123), the translational field strength is identical to the one that would result from the Poincaré algebra, whereas the rotational field strength receives an additional part that vanishes if the de Sitter radius goes to infinity:

$$F^{IJ}_{\mu\nu} = \hat{F}^{IJ}_{\mu\nu} - \frac{1}{\mathcal{R}^2} (A^I_\mu A^J_\nu - A^I_\nu A^J_\mu), \tag{7.127}$$

and where $\hat{F}$ is the rotational field strength part (7.123a) from the PGT.

A kinetic term for the gauge fields might be built in the Yang-Mills manner as

$$S = \int d^4x \, G^{\mu\nu\rho\sigma}_{AB} \, F^A_{\mu\nu} F^B_{\rho\sigma}$$

with the coefficients G chosen such that the Lagrangian is (quasi-)invariant. Now $G^{\mu\nu\rho\sigma}_{AB}$ can be constructed from $\eta^{\mu\nu}$ and $\epsilon^{\mu\nu\rho\sigma}$ in order to get a Lorentz scalar, from a group metric g_{AB} and from ϵ_{abcde}. It turns out that a theory invariant under the de Sitter group can only be built if $F^I_{\mu\nu} = 0$. The invariant de Sitter action is finally found to be

$$S_{dS} = \int d^4x \, \epsilon^{\mu\nu\rho\sigma} \epsilon_{IJKL4} \, F^{IJ}_{\mu\nu} F^{KL}_{\rho\sigma}. \tag{7.128}$$

---

[43] Geometrically, this is a fibre bundle with base space M_4 and fibre $\mathbf{dS_4}$. The de Sitter group acts on the matter fields in each fibre.

[44] Be reminded of conventions for the group generators and other related objects different from those elsewhere in this text.

Written in terms of the PGT objects this Lagrangian splits into three terms–being proportional to $\mathcal{R}^0, \mathcal{R}^{-2}, \mathcal{R}^{-4}$:

$$S^{(0)} = \frac{1}{4}\int d^4x\, \epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL}\hat{F}^{IJ}_{\mu\nu}\hat{F}^{KL}_{\rho\sigma} \tag{7.129a}$$

$$S^{(-2)} = \mathcal{R}^{-2}\int d^4x\, \epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL}A^I_\mu A^J_\nu \hat{F}^{KL}_{\rho\sigma} \tag{7.129b}$$

$$S^{(-4)} = \mathcal{R}^{-4}\int d^4x\, \epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL}A^I_\mu A^J_\nu A^K_\rho A^L_\sigma. \tag{7.129c}$$

Due to the structural similarities of the field strength (7.126) with the torsion and the curvature of a Riemann-Cartan spacetime it stands to reason to identify the gauge fields with the tetrads and the spin connections. If this is allowed, the action term (7.129a) corresponds to the Gauß-Bonnet topological invariant (7.105). The term (7.129b) is the Einstein-Cartan action and the part (7.129c) represents the term with a cosmological constant.

Still another approach towards a de Sitter gauge theory uses the mechanism of spontaneous symmetry breaking of the de Sitter symmetry, which justifies the foregoing identification of the translational gauge fields with the tetrads, see [484] and App. C in [49].

Other Gauge Groups

Other investigations dealt with e.g. the Weyl group, the conformal group or the affine group. For the Weyl group (or the conformal group, respectively) corresponding to one (or two) additional symmetry generators, there are, aside from the A^I_μ and A^{IJ}_μ, additional gauge fields. These new fields are not only add-ons; rather, their presence entails a change of the transformation properties of the translation and rotation gauge fields and field strength; for more details see e.g. [49]. Metric affine gravity (MAG) also received widespread attention. Here one dispenses from the metricity condition, and the non-metricity tensor $Q^{\lambda\mu\nu}$ (see (7.24)) serves as an additional field strength. It is associated to the metric, which now serves as a further gauge field, see e.g. [243], [265].

The decision as to which of these groups to employ as a gauge group is difficile by a semantic freedom of how to relate gauge connections to geometric entities in either metric or tetrad gravity. I mentioned this in the context of PGT and the de Sitter gauge theories.

The question of whether or not gravity is a gauge theory is still a disputed one. For further applications of gauge principles to gravitational interactions see ([85]), where also more recent approaches such as topological gravity in odd dimensions (in terms of so called Chern-Simons theories) are described.

7.6.4 Changing Structures and Modifying Principles

It was shown by V. Iyer and R.M. Wald [294] that any action for diffeomorphism invariant theories depending on the metric, the curvature tensor and its covariant derivatives (via D_μ), and on tensor matter fields Φ (and their covariant derivatives) necessarily has the form

$$S(g,\Phi) = \int d^D x \sqrt{-g}\mathcal{L}\Big(g_{\mu\nu}, R_{\mu\nu\rho\sigma}, D_{\mu_1} R_{\mu\nu\rho\sigma}, \dots, \\ D_{(\mu_1\dots\mu_m)} R_{\mu\nu\rho\sigma}, \Phi, D_{\mu_1}\Phi, \dots, D_{(\mu_1\dots\mu_n)}\Phi\Big)$$

Quite interestingly, many features known from GR are present in this large set of models independent of the specific form of the Lagrangian. This is true for instance for the generic structure of the Hamiltonian and the algebra of phase-space constraints (see Appendix C.3.3), and it also holds for the horizon entropy in these models. These structural similarities seem to be rooted in the structure of the Noether currents due to diffeomorphism covariance.

To simplify matters, consider the case in which $\mathcal{L}$ does not depend on the derivatives of the curvature tensor, but is otherwise arbitrary. Then one derives ([375], [403])

$$\delta(\sqrt{-g}\mathcal{L}(g_{\mu\nu}, R_{\mu\nu\rho\sigma})) = \sqrt{-g}\,\left(\mathcal{G}_{\mu\nu}\delta g^{\mu\nu} + D_\mu \delta\mathcal{W}^\mu\right)$$

where

$$\mathcal{G}_{\mu\nu} := (\mathcal{R}_{\mu\nu} - \frac{1}{2}\mathcal{L}g_{\mu\nu}) - 2D^\rho D^\sigma P_{\mu\rho\nu\sigma}$$

$$\mathcal{R}_{\mu\nu} := P_\mu{}^{\lambda\rho\sigma} R_{\nu\lambda\rho\sigma} \qquad\qquad P_{\mu\rho\nu\sigma} := \frac{\partial\mathcal{L}}{\partial R_{\mu\rho\nu\sigma}}.$$

For example, for GR with the Hilbert Lagrangian $\mathcal{L} = g^{\mu\nu}g^{\rho\sigma}R_{\rho\mu\sigma\nu}$, we have $P_{GR}^{\mu\nu\rho\sigma} = \frac{1}{2}(g^{\mu\rho}g^{\nu\sigma} - g^{\nu\rho}g^{\mu\sigma}), \mathcal{R}_{\mu\nu} = R_{\mu\nu},\ \mathcal{G}_{\mu\nu} = G_{\mu\nu}$.

Independent of the functional form of $\mathcal{L}$ it is found that, like the Einstein tensor $G_{\mu\nu}$, also the tensor $\mathcal{G}_{\mu\nu}$ obeys a Bianchi identity: $D^\mu\mathcal{G}_{\mu\nu} \equiv 0$.

For diffeomorphisms with $\delta_\xi g^{\mu\nu} = 2D^{(\mu}\xi^{\nu)}$ the invariance identity reads

$$\sqrt{-g}\,\left(2\mathcal{G}_{\mu\nu}D^{(\mu}\xi^{\nu)} + D_\mu\delta_\xi\mathcal{W}^\mu\right) = -\partial_\mu(\sqrt{-g}\mathcal{L}\xi^\mu)$$

or

$$2\mathcal{G}_{\mu\nu}D^\mu\xi^\nu + D_\mu\delta_\xi\mathcal{W}^\mu + D_\mu\mathcal{L}\xi^\mu = 0$$

which, due to the Bianchi identity, results in the existence of a covariantly conserved current $\bar{\mathcal{J}}_\xi^\mu$

$$\bar{\mathcal{J}}_\xi^\mu = 2\mathcal{G}^{\mu\nu}\xi_\nu + \delta_\xi\mathcal{W}_\mu + \mathcal{L}\xi_\mu = 2\mathcal{R}^{\mu\nu}\xi_\nu + \delta_\xi\mathcal{W}_\mu.$$

(This current was identified for GR in Sect. 7.5.4; see (7.92)). In expressing the boundary term $\delta_\xi \mathcal{W}^\mu$ through the variations of the metric and the connection, one then can write down the current $\bar{\mathcal{J}}^\mu_\xi$ for every Lagrangian $\mathcal{L}(g_{\mu\nu}, R_{\mu\nu\rho\sigma})$. Instead the current can also be expressed by a superpotential $\mathcal{U}^{\mu\nu}_\xi$:

$$\bar{\mathcal{J}}^\mu_\xi = D_\nu \mathcal{U}^{\mu\nu}_\xi \qquad\qquad \mathcal{J}^{\mu\nu}_\xi = 2P^{\mu\nu\rho\sigma} D_\rho \xi_\sigma - 4(D_\rho P^{\mu\nu\rho\sigma})\xi_\sigma.$$

In the case of GR, with the previously identified symbol $P^{\mu\nu\rho\sigma}_{GR}$, the superpotential is $\mathcal{U}^{\mu\nu}_\xi = 2D^{[\mu}\xi^{\nu]}$. It differs from the Komar superpotential $_K U^{\mu\nu}$ by a factor $\sqrt{-g}$ (a factor which is needed because the Noether current is on-shell conserved.)

Let me mention here a specific subset of Lagrangians, namely those with $\{DP = 0\}$. These are known as Lanczos-Lovelock[45] models of gravitation. They have covariantly conserved currents $\bar{\mathcal{J}}^\mu_\xi = 2P^{\mu\nu\rho\sigma} D_\nu D_\rho \xi_\sigma$. And of course the Hilbert-Lagrangian belongs to this subset.

Allowing Fourth Order Field Equations

One possibility to modify GR is to replace the Lagrange density $R - 2\Lambda$ by a nonlinear function $f(R)$. This leads to theories with Lagrangians of the generic form

$$S_f = \frac{1}{2\kappa} \int d^4x \sqrt{-g} f(R) + S_M[g, \Phi]. \tag{7.130}$$

In contrast to the linear case, where the second-order and the first-order variational formalism yield the same result, this is no longer the case for $f(R)$ models. In the second-order formalism the variation of the gravitational action is

$$\delta \int d^4x \sqrt{-g} f(R) = \int d^4x \sqrt{-g}\delta g^{\mu\nu} \Big\{ R_{\mu\nu} f'(R) - \frac{1}{2} g_{\mu\nu} f(R) + g_{\mu\nu} \nabla^2 f'(R) - \nabla_\mu \nabla_\nu f'(R) \Big\}.$$

Again care must be taken with boundary terms, which will be ignored here, however. The field equations

$$\{R_{\mu\nu} f'(R) - \frac{1}{2} g_{\mu\nu} f(R) + g_{\mu\nu} \nabla^2 f'(R) - \nabla_\mu \nabla_\mu f'(R)\} = -\kappa T_{\mu\nu}$$

are of fourth order in the metric–unless the function $f(R)$ is linear in R. However, interestingly enough the $f(R)$ theories circumvent the Ostrogradski theorem [567]. It has been known since quite some time that this model can be cast into the form of a scalar-tensor theory: Rewrite the action S_f by help of an auxiliary field χ as

$$S_f = \frac{1}{2\kappa} \int d^4x \sqrt{-g} \big[f(\chi) + f'(\chi)(R - \chi) \big] + S_M(g_{\mu\nu}, \Phi).$$

[45] Interestingly there are various ways to characterize these models, as also will be seen at other places in this book.

Indeed, variation with respect to χ leads to the equation

$$f''(R)\,(R-\chi)=0,$$

and therefore $\chi = R$ if $f''(\chi) \neq 0$. In this case $\phi = f'(\chi)$ is invertible and setting

$$V(\phi) \;=\; \chi(\phi)\,\phi \;-\; f(\chi(\phi)),$$

the action becomes

$$S_f = \frac{1}{2\kappa}\int d^4x\sqrt{-g}\,[\phi R - V(\phi)] + \; S_M[g,\Phi].$$

This is known as the Jordan frame representation of the action of a *Brans-Dicke* theory with a specific Brans-Dicke parameter; see e.g. [78]. There are $f(R)$ models that pass the solar system tests and reproduce all results from experimental cosmology. However, it was discovered that they cannot describe relativistic stars. The Brans-Dicke theory is a specific case of the class of scalar-tensor theories [203].

As for first-order variational formalism of $f(R)$ actions these circumvent the possible problems with higher than second order derivatives, but they get into trouble because they do not have a well-defined Cauchy problem. More details are to be found in [479].

Effective Field Theory Description of Gravity

If we would strictly observe the construction rules for actions in relativistic field theories as stated in Sect. 5.3.1., we never would have a chance to arrive at GR, because the Hilbert-Einstein action does not obey the renormalizability criterium. However, as described in Sect. 5.6, the original quest for renormalizability was reinterpreted with the notion of effective field theories. So why not interpret GR as an effective field theory? This point of view immediately leads to an action containing higher derivative terms:

$$\begin{aligned} S^g = \int d^4x\,\sqrt{-g}\Big[\alpha_0 M_{Pl}^4 + \frac{M_{Pl}^2}{2}R + \alpha_1 R^2 + \alpha_2 R_{\mu\nu}R^{\mu\nu} + \alpha_3 R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} \\ + \alpha_4 \Box R + \frac{\alpha_5}{m^2}R^3 + \cdots + \mathcal{L}_M\Big]. \end{aligned} \tag{7.131}$$

Written this way with the Planck mass M_{Pl}, all constants α_i are dimensionless. The first term with α_0 is of course related to the cosmological constant. The second term is the Hilbert-Einstein Lagrangian. The mass m is introduced for dimensional reasons and acquiring a meaning if (7.131) is taken as an effective theory, valid in a certain mass/energy range. It may be as large as the Planck mass but as small as the electron mass. Symmetry allows for further terms cubic in the curvature aside from the explicitly written one. For $D = 4$ we might set $\alpha_3 = 0$ and think of redefining α_1 and α_2 by help of the Gauß-Bonnet topological invariant (7.105). The term with α_4 may be dropped because $\sqrt{-g}\,\Box R = \partial_\mu(\sqrt{-g}\,D^\mu R)$ also amounts to a boundary term. In the case of vacuum gravity and of a vanishing cosmological constant, also the terms with α_1 and α_2 can be neglected since they are proportional to the lower order field equations: $G_{\mu\nu}[2\alpha_2 R^{\mu\nu}\text{-}(\alpha_2 + 2\alpha_1)Rg^{\mu\nu}] \propto [\alpha_1 R^2 + \alpha_2 R_{\mu\nu}R^{\mu\nu}]$. Therefore in this case the first term that might have physical effects is proportional to terms cubic in the Riemann tensor. In the presence of matter this is no longer true. Admittedly, terms proportional to $G_{\mu\nu}+\kappa T_{\mu\nu}$ can be neglected, but there are further terms with interactions of the form $T_{\mu\nu}R^{\mu\nu}$, $T_{\mu\nu}T^{\mu\nu}$ (and other contractions). Based on this effective Lagrangian, some quantum corrections for classical gravity were calculated like for instance the first quantum correction to Newton's law; for more details, see the review [63]. Recently there have been various hints that general relativity might be asymptotically safe.

Phantasies of Theoreticians

- Dispensing with curvature.
  We saw already two examples: CDJ only in terms of the spin connection, teleparallel gravity in terms only of torsion.
- Modifying the symmetry transformations
  - Bimetric Gravity: As its name implies, bimetric theories contain two metric tensors. Usually, one of them takes the role of the gravitational field to which matter is coupled. In the majority of bimetric theories, one forms the Christoffel connections with respect to the two metrics. This allows the formation of two curvatures and of a further invariant tensor from the difference of the two connections. Thus there is more freedom in building actions. Bimetric theories relate to massive gravity and to variable speed of light theories; see the review [93] on modified gravity.
  - Unimodular Gravity: This goes back to a perplexing proposal of Einstein from 1919. In [157] he starts to complain about his own field equations $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -\kappa T_{\mu\nu}$ these having no chance of being able to be coupled to an electromagnetic field. He argues that because the electromagnetic energy-momentum is traceless, this would enforce $R = 0$ which he dislikes. It seems as if he ignores that there are still non-trivial field equations $R_{\mu\nu} = -\kappa T_{\mu\nu}$. In any case he wanted the trace to vanish on both sides identically which led him to modify his field equation by slipping in another factor in front of the curvature scalar such that now

$$R_{\mu\nu} - \frac{1}{4}Rg_{\mu\nu} = -\kappa T_{\mu\nu}.$$

And now comes a stroke typical of a genius, namely at the same time to eliminate one of his other worries ("... *besonders schwerwiegender Schönheitsfehler der Theorie*"), namely the artificial insertion of the cosmological constant into the original field equations. Deviating now from the original, in which the modified field equations lead to a modified energy-momentum tensor, let me take another route. In his 1919 contribution to the Prussian Academy of Science Einstein did not talk about an action for his modified theory. It seems that despite of the existence of the Hilbert action and of the differential geometric foundation of GR, he ignores any symmetry aspects of his beautiful theory. His modifying factor (1/4) is indeed a consequence of a special case of general coordinate invariance: Start from the action

$$S = \frac{1}{2\kappa} \int d^4x \sqrt{-g}\, R + S_M\,[g, \Phi]$$

and vary this with respect to those transformations which leave the determinant of the metric invariant, that is

$$\delta^t \sqrt{-g} = 0 \;\leftrightarrow\; g_{\mu\nu}\delta^t g^{\mu\nu} = 0.$$

These volume-preserving transformations can be expressed by the unconstrained transformations as $\delta^t g^{\mu\nu} = \delta g^{\mu\nu} - \frac{1}{4} g^{\mu\nu} g_{\rho\sigma} \delta g^{\rho\sigma}$ and thus

$$\frac{\delta S}{\delta^t g^{\mu\nu}} = \frac{\delta S}{\delta g^{\mu\nu}} - \frac{1}{4} g_{\rho\sigma} \frac{\delta S}{\delta g^{\rho\sigma}} g_{\mu\nu}.$$

Inserting the previously derived result $\frac{\delta S}{\delta g^{\mu\nu}} = \frac{1}{2\kappa}(G^{\mu\nu} + \kappa T^{\mu\nu})$ into this expression one finds the field equations

$$\left(G_{\mu\nu} + \frac{1}{4} R g_{\mu\nu}\right) = \left(R_{\mu\nu} - \frac{1}{4} R g_{\mu\nu}\right) = -\kappa\left(T_{\mu\nu} - \frac{1}{4} g_{\mu\nu} T\right).$$

Using the contracted Bianchi identity $D^\mu G_{\mu\nu} = 0$ and assuming covariant conservation of the energy-momentum tensor one derives $\partial_\mu R = \kappa \partial_\mu T$. This integrates to $R = \kappa T + \lambda$ with an integration constant λ. Reinserting this into the field equations, we finally get

$$G_{\mu\nu} + \frac{1}{4} \lambda g_{\mu\nu} = -\kappa T_{\mu\nu}.$$

Thus, unimodular gravity (the name deriving from setting $\sqrt{-g} = 1$) allows to arrive at field equations including a cosmological constant. At the same time this is another example for which the symmetry group of the dynamical equations is larger than the symmetry group of the action. Since the determinant transforms under general coordinate transformations as $\delta g = -2g\xi^\lambda_{,\lambda}$ we immediately see that for unimodular gravity the action is invariant for those restricted coordinate transformations that obey $\xi^\lambda_{,\lambda} = 0$.
M. Henneaux and C. Teitelboim succeeded in reformulating unimodular gravity so that the full diffeomorphism group stays intact [272]. This is achieved by introducing an auxiliary field $A_{\lambda\mu\nu}$ which is totally antisymmetric in its indices.

– Hořava-Lifschitz gravity is one of the many attempts to reconcile the special role of time in quantum physics with the notion of coordinate time in GR, where it is nothing but a parameter on equal footing with the spatial coordinates. In a somewhat simplified approach, one might understand the action for Hořava-Lifschitz gravity as being constructed in such a way that it is invariant under space-independent time reparametrization. This is a restriction of the diffeomorphisms to $t \rightarrow \hat{t}(t), x^i \rightarrow \hat{x}^i(t, x^k)$. This seems to overthrow all we have learned from Einstein's special relativity and to be a step back to Newton's understanding of time. The philosophy of the approach is to recover Lorentz invariance from quantum gravity; see the status report [478].

- Reflecting about dimensions other than $D = 4$
The dimensionality $D = 4$ has very peculiar properties both mathematically as well as physically; see e.g. [30]. In $D = 2$ the Einstein tensor vanishes identically. For $D = 3$ GR becomes a genuine gauge theory of the Poincaré or the de Sitter groups. Gravitation physics in higher dimensions $D > 4$ became a fashionable subject of current research motivated mainly by string theory.

– Lanczos-Lovelock gravity [350] in D dimensions is defined by a gravitational Lagrangian

$$\mathcal{L}_L = \frac{c^4}{8\pi G_D} \sum_{0 \leqslant k \leqslant D/2} \alpha_k \lambda^{2(k-1)} \mathcal{L}_k \tag{7.132}$$

$$\mathcal{L}_k = \frac{1}{2^k} \delta^{A_1 \cdots A_{2k}}_{B_1 \cdots B_{2k}} R^{B_1 B_2}{}_{A_1 A_2} \cdots R^{B_{2k-1} B_{2k}}{}_{A_{2k-1} A_{2k}}.$$

Here, G_D is Newton's constant in a D-dimensional spacetime, λ is a length scale and α_k are real dimensionless parameters. The $\delta^{A_1 \cdots A_{2k}}_{B_1 \cdots B_{2k}}$ is the determinant tensor of order $2k$ (for example $\delta^{AB}_{CD} = \frac{1}{2}(\delta^A_C \delta^B_D - \delta^A_D \delta^B_C)$). The lowest order terms in the Lagrangian (7.132) are

$$\begin{aligned} \mathcal{L}_0 &\propto \sqrt{-g} \\ \mathcal{L}_1 &\propto \sqrt{-g} R \\ \mathcal{L}_2 &\propto \sqrt{-g}(R^2 - 4 R_{AB} R^{AB} + R_{ABCD} R^{ABCD}). \end{aligned}$$

The first term relates to the cosmological constant, the second to the Hilbert action, and the third is the Gauß-Bonnet term. For $D = 2$ and $D = 3$ the Lanczos-Lovelock action only contains the first two terms, for $D = 4$ and $D = 5$ the Gauß-Bonnet term is included.
The Euler derivative of the Lanczos-Lovelock Lagrangian with respect to the metric leads to the Lovelock tensor

$$L_{AB} = \sum_{0 \leqslant k \leqslant D/2} \alpha_k \lambda^{2(k-1)} L_{(k)AB}$$

$$L_{(k)}{}^A{}_B = -\frac{1}{2^{k+1}} \delta^{AC_1 \cdots C_{2k}}_{BD_1 \cdots D_{2k}} R^{D_1 D_2}{}_{C_1 C_2} \cdots R^{D_{2k-1} D_{2k}}{}_{C_{2k-1} C_{2k}}.$$

In the presence of matter, the field equations become $L_{AB} + \kappa^D T_{AB} = 0$. As proven by D. Lovelock, the tensor L_{AB} (1) is symmetric, (2) contains up to second derivatives of the metric and is linear in the second derivatives of the metric, (3) is covariantly conserved: $D_A L^{AB} = 0$, and is the only tensor having these properties. The explicit form of its first three terms is

$$\begin{aligned} L_{(0)AB} &\propto -\tfrac{1}{2} g_{AB} \\ L_{(1)AB} &\propto R_{AB} - \tfrac{1}{2} g_{AB} R \\ L_{(2)AB} &\propto R R_{AB} - 2 R_{AC} R^C_{\ B} - 2 R^{CD} R_{ACDB} \\ &\qquad + R_A^{\ CDE} R_{BCDE} - \tfrac{1}{4} g_{AB} (GB)^{AB} \end{aligned}$$

where (GB) is the Gauß-Bonnet tensor. In two and three dimensions the field equations are built from the first two terms. For $D = 2$, the Einstein tensor $R_{AB} - \frac{1}{2} g_{AB} R$ vanishes identically. This indicates that the Hilbert action in two dimensions is nothing but a boundary term. For $D = 4$ and $D = 5$, the $L_{(2)AB}$-term emerges. It vanishes identically in four dimensions, as the associated highest Lagrangian term is a divergence. This exhibits a general pattern of Lanczos-Lovelock gravity. The highest Lagrangian term in even dimensions is identically zero, and the highest Lovelock tensor is a boundary term in even dimensions [569]. It does not contribute to the field equations, but constitutes a topological invariant, namely the Euler invariant. As a matter of fact the Lanczos-Lovelock Lagrangian can be understood as the generic linear combination of the extension of the Euler invariants from all even dimensions below D to D dimensions [582].

Lanczos-Lovelock gravity in arbitrary dimension was extended to Riemann-Cartan theory in [361], where the construction with the Euler densities was carried over to include Pontryagin densities; see also App. F.3.2.

- Another class of generalized GR are the Kaluza-Klein models treated in the next chapter. Here, the basic idea is that extra dimensions are tiny and curled up, so that they escape observation for low energies.
- Perhaps influenced by the original Kaluza-Klein attempts, compactification initially meant a dimensional reduction to a Minkowski spacetime together with a small compact space. This changed however with the work of N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, who formulated a theory in which the extra dimensions are not of Planck size, but macroscopic. The basic idea, inspired by the D-branes in string theory, is the assumption that whereas the standard model interactions act in 4D-spacetime, gravity propagates into higher dimensions. Higher-dimensional gravity would influence Newton's law

$$F = \frac{m_1 m_2}{M_4^2} \frac{1}{r^2} \longrightarrow \frac{m_1 m_2}{M_{4+d}^2} \frac{1}{r^{2+d}}.$$

Here M_{4+d} is Planck's constant in $(4 + d)$-dimensions. If for instance the extra dimensions are assumed compact and of the order R we find the relation $M_4 = M_{4+d} R^{d/2}$. This offers a handle to solve the hierarchy problem: Assume

that the Planck mass in (4+d) is of the order of ten times the electroweak scale, then for $d = 2, 3, 4$ we get $R \sim (10^{-3} - 10^{-7})m$. This is well outside the range in which Newton' law is tested. The ADD model is a representative of models with "Large Extra Dimensions". L. Randall and R. Sundrum devised scenarios in which the extra dimensions are not flat but "warped". The different models of extra dimension make predictions which can be tested at the LHC [371]. Indeed the parameters of the models are by now included in the Particle Data Group's search for physics beyond the standard model.
– In odd dimensions D, and only in that case, gravity can be cast as a genuine gauge theory of either the Poincaré group $\mathbf{ISO(D-1,1)}$, the deSitter group $\mathbf{SO(D,1)}$ or the Anti-deSitter group $\mathbf{SO(D-1,2)}$. This is the domain of Chern-Simons gravity [577].

- … open end
nonsymmetric gravity [111]; Chern-Simons modified GR [296]; spontaneous breakdown of diffeomorphism symmetrie [505]; GR as the effective theory of GL(4,R) spontaneous symmetry breaking [507]; spinor gravity [260]

There is also a wealth of ideas by which general relativity is not fundamental and not necessarily described in terms of geometry but emergent from some other more basic entities. I will come back to these in the Conclusion, and specifically address the question how symmetries may emerge.

7.7 Concluding Remarks and Bibliographical Notes

This chapter on "General Relativity and Gravitation" has an "and" in its title: As a matter of fact, general relativity is our currently best model of gravitation. As explained, it overcomes conceptual weaknesses of Newton's formulation, and it makes predictions which were tested with an incredible amount of precision ([109, 562]). Hereby two aspects of GR must be distinguished: On the one hand, the equivalence principle which dictates how "matter" reacts to a given gravitational field, and on the other hand Einstein's field equations which describe the dynamics of the gravitational field in response to the energy-momentum of "matter". By now GR has passed "with flying colors" not only the so-called classical tests (those proposed by Einstein himself), but also further tests by ever more sophisticated experimental techniques enabled by satellite technology. As recently as 2011, there was the announcement that the Lense-Thirring predictions of frame-dragging had been confirmed by the Gravity-Probe B satellite experiment. Since–due to its weakness–the nature of gravitational attraction on small scales is not fully established, there are now and then (theoretical) considerations of deviations on the microscopic level[46].

Compared to internal symmetries, the invariance under general coordinate transformations is less restrictive with regard to the form of the action functional. I have

[46] A typical representative of this approach is the brane world.

described various modifications and alternatives to the Hilbert-Einstein action, but they are currently not at all required by experimental results.

There are of course a number of commendable (text-)books on GR, and which one to choose is a matter of taste. Although I grew up with [373], I preferred in my more recent lectures on general relativity, cosmology and black-hole physics the monographs [78], [75], [404], [489]. A recommendable review article on the essentials of classical GR is [149]. The relevant geometries and their interrelations are treated in detail in [263]. A compact presentation of the relation between the formulation of Riemann geometry either in terms of the metric/Levi-Civita connection or in the tetrad/spin connection is found in [146].

Even though GR has been very successful as a classical theory of gravity, its quantization is still an open issue. There is agreement in the scientific community that gravity must be quantized. But neither canonical nor path-integral quantization could be accomplished, either because of the field-dependent structure functions in the constraint algebra or the undefined diffeomorphism-group measure. Since the constant multiplying the gravitational part of the GR action has mass dimension $+2$, the theory is non-renormalizable. And indeed, a perturbative loop expansion results in ultraviolet-divergent Feynman diagrams. Striving to suppress these contributions, one needs to introduce infinitely many counter terms which become polynomial in the curvature and its covariant derivatives. Thus there is no predictive power at small distances. Actually, most of the attempts of modifying or extending GR are aimed at arriving at a better starting position in view of quantizing the theory.

Chapter 8
*Unified Field Theories

Further progress lies in the direction of making our equations invariant under wider and still wider transformations.

The motto of this chapter is the rather prophetic announcement by P.A.M. Dirac in 1930 in the preface to the first edition of his famous book on quantum mechanics [124]. From a present day stance we could in brief refer to "Symmetries of Everything" (abbreviated as SOE) which mimics the presumptuous TOE ("Theory of Everything").

8.1 Grand Unified Theories

8.1.1 Motivation and Basic Concepts

At the end of Chap. 6 various shortcomings of the Standard Model of particle physics were mentioned. One of them is the fact that three distinct coupling constants corresponding to the three factors in the gauge group $\mathbf{SU_C(3)} \times \mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$ need to be introduced. This came about because the three interactions differ in strength at "low" energies, that is at energies well below the W-mass. We also saw that at the scale of M_W the electromagnetic and the weak interactions become comparable in strength, a fact which after all made electroweak unification possible. What about a further unification with the strong interaction? Indeed a look at the running coupling constants makes one optimistic in believing that at a certain high-energy scale M_X, all three interactions converge in terms of strength. If we assume that

$$\alpha_C(M_X) = \alpha_I(M_X) = \alpha_Y(M_X) = \alpha_{GUT}$$

where $\alpha_i = g_i/4\pi \quad (i = C, I, Y \doteq 3, 2, 1)$, a qualitative assessment of (6.43)

$$\frac{1}{\alpha_i(\mu)} = \frac{1}{\alpha_{GUT}} + \frac{b_i}{2\pi} \ln \frac{\mu}{M_X}$$

demonstrates that the three straight lines $\alpha_i(\ln \mu)$ have a good chance of intersecting around $M_X \sim 10^{15}$ GeV; see Fig. 8.1a.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_8,

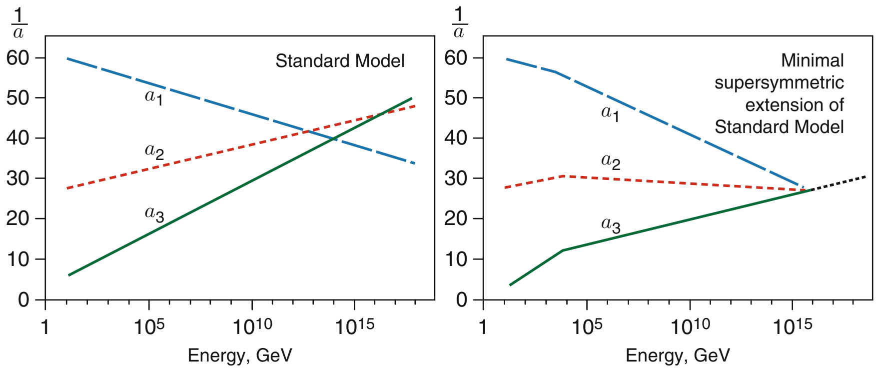


Fig. 8.1 Running coupling constants

Reflecting about an enhanced symmetry provokes the question of whether a larger gauge group $\mathbf{G}$ exists, which breaks down into the Standard Model group. The idea is to place the matter fields into appropriate representations of the group, to introduce as many gauge fields as requested by the dimension of the group, and then break the symmetry down to the Standard Model group by a Higgs mechanism. In order that $\mathbf{G}$ be a genuine grand unifying group, it must be simple.

The representations of the matter field must respect the structure of the Standard Model: They can be catalogued by their transformation behavior under the separate groups[1]; see also Table 6.1. For example, the left-handed up quark belongs to a color triplet, to a weak-isospin doublet $\begin{pmatrix} u \\ d \end{pmatrix}_L$, and it has the hypercharge $Y = 1/3$. Denote this by $(3, 2, 1/3)_L$. In this notation the right-handed up quark is $(3, 1, 1/3)_R$ and the right-handed down quark is represented by $(3, 1, -4/3)_R$. The leptons do not participate in strong interactions. Formally we give them a singlet "1" in the first entry, so they are registered as $(1, 2, -1)_L$ and $(1, 1, -2)_R$. For further purposes it is more convenient to work with only left-handed fields. The right-handed fields are thus written as the charge-conjugated left-handed fields. For the representation in the previous notation this means

$$\begin{aligned} u_L, d_L: &\quad \left(3, 2, \frac{1}{3}\right) \\ u_L^c: &\quad \left(\bar{3}, 1, -\frac{4}{3}\right) \\ d_L^c: &\quad \left(\bar{3}, 1, \frac{2}{3}\right) \\ \nu_L, e_L: &\quad (1, 2, -1) \\ e_L^c: &\quad (1, 1, 2). \end{aligned} \tag{8.1}$$

[1] In this section only one family is taken into account. The other two can simply be added later. Also minor complications–essentially in notation–due to quark mixing are ignored here.

The comment of A. Zee ([579]) to this catalogue is "This motley collection of representations practically cries for further unification. Who would have constructed the universe by throwing this strange looking list down?" Notice that each generation of leptons has 15 members (3 colored u- and d-quarks and their antiparticles, the electron and the positron, and a left-handed neutrino) if we expel the right-handed neutrinos.

The set of gauge bosons associated with the adjoint representation of **G** must of course comprise the eight gluons and the four electro-weakons of the Standard Model. The additional gauge bosons X are assumed to set the scale for symmetry breaking of this extended theory. And, whereas in the Standard Model only (lepton-lepton)- and (quark-quark)-transitions occurred, the X bosons mediate (quark-lepton)-transitions among those quark and leptons which are members of the same multiplet in the gauge group. Thus the new bosons carry both flavor and color properties (for this reason they are called *leptoquarks*). Because of the quark-lepton transitions both baryon number and lepton number conservation no longer holds in GUTs.

Since thus far, no reactions changing a quark into a lepton (or *vice versa*) have been observed, the leptoquarks X must be very heavy. A lower bound comes from experiments looking for proton decay $p \to e^+\pi^0 \to e^+\gamma\gamma$. The proton lifetime is of the order of

$$\tau_p \sim \frac{\hbar}{\alpha^2 c^2} \frac{M_X^4}{m_p^5} .$$

The experimental bound is $\tau_p \gtrsim 10^{32}$ years[2] from which $M_X \gtrsim 10^{15}$ GeV. This is compatible with the picture of the running coupling constants converging at $\sim M_X$.

All GUTs offer an explanation why electric charge is quantized: In a simple group all generators have discrete eigenvalues. And since the GUT amongst others describes electromagnetism, the charge operator Q must be among the set of generators. GUTs also relate the charges of those leptons and quarks which belong to the same multiplet. Another feature of GUTs is their prediction of the Weinberg angle simply by relating the normalization of the **U(1)** hypercharge operator with the operators in **G**; details for **SU(5)** below.

Let me list again the requirements on the GUT group **G** as demanded by particle physics. **G** must

- Cover the Standard Model group: $\mathbf{G} \supset \mathbf{SU(3)} \times \mathbf{SU(2)} \times \mathbf{U(1)}$.
- Be simple. Since the Standard Model group has four diagonal generators (strong isospin and hypercharge plus weak isospin and hypercharge), its rank is four. Therefore a GUT group **G** must be a group with rank $\geqslant 4$ in the Cartan classification[3].
- Allow for complex representations as we see them arising in the catalogue (8.1)

[2] This is much larger than 10^{10} years, the age of the universe.

[3] One may also contemplate product groups in the form $G = \tilde{G} \times \tilde{G}$. This is no longer a simple group, but one might envision that the two coupling constants are related by a discrete symmetry.

- Have representations which can host the fermions and the gauge bosons of the Standard Model, respecting their quantum numbers with regards to color, weak isospin and hypercharge.
- Be free of anomalies. It was mentioned already (see 6.37) that this requirement can be formulated in terms of the representation of the group generators as

$$\sum_{\text{repr}} \mathrm{Tr}[(T^a T^b + T^b T^a) T^c] = 0.$$

Of course even under these restrictions there are many groups that could serve as the GUT group. But we will see in a moment that already many of the smallest group (in terms of the dimension of the group) fall through the sieve.

The very first GUT, based on $\mathbf{G} = \mathbf{SU(5)}$, was investigated by H. Georgi and S. Glashow [214]. This model is minimal in that the gauge group is the smallest one fulfilling the foregoing requirements. The $\mathbf{SU(5)}$ GUT will in more detail be described below in Sect. 8.1.3. I only sketch another very attractive model that is based on the group $\mathbf{SO(10)}$, contrived by H. Fritzsch and P. Minkowski [200].

8.1.2 SU(5) *Grand Unification*

Within the Cartan classification of simple Lie groups, one finds the rank-4 groups $\mathbf{SU(5)}$, $\mathbf{SO(9)}$, $\mathbf{Sp(8)}$, $\mathbf{SO(8)}$, $\mathbf{F_4}$. Of these, from the requirement of complex representations, only $\mathbf{SU(5)}$ is acceptable.

In order to see whether and how the assignments (8.1) can be realized in representations of $\mathbf{SU(5)}$ we decompose the $\mathbf{SU(5)}$ representations into its $\mathbf{SU(3)}$, $\mathbf{SU(2)}$, and $\mathbf{U(1)}$ parts. The fundamental representation of $\mathbf{SU(5)}$ decomposes as

$$\mathbf{5} = \left(\mathbf{3}, \mathbf{1}, -\frac{2}{3}\right) \oplus (\mathbf{1}, \mathbf{2}, 1). \tag{8.2}$$

Higher-dimensional representations are obtained by multiplications $\mathbf{5} \otimes \mathbf{5}$ or $\mathbf{5} \otimes \bar{\mathbf{5}}$:

$$\begin{aligned}
\mathbf{5} \otimes \mathbf{5} &= \left[\left(\mathbf{3}, \mathbf{1}, -\tfrac{2}{3}\right) \oplus (\mathbf{1}, \mathbf{2}, 1)\right] \otimes \left[\left(\mathbf{3}, \mathbf{1}, -\tfrac{2}{3}\right) \oplus (\mathbf{1}, \mathbf{2}, 1)\right] \\
&= \left(\mathbf{3} \otimes \mathbf{3}, \mathbf{1}, -\tfrac{4}{3}\right) \oplus \left(\mathbf{3}, \mathbf{2}, \tfrac{1}{3}\right) \oplus \left(\mathbf{3}, \mathbf{2}, \tfrac{1}{3}\right) \oplus (\mathbf{1}, \mathbf{2} \otimes \mathbf{2}, 2) \\
&= \left[\left(\mathbf{6}, \mathbf{1}, -\tfrac{4}{3}\right) \oplus \left(\mathbf{3}, \mathbf{2}, \tfrac{1}{3}\right) \oplus (\mathbf{1}, \mathbf{3}, 2)\right]_S \oplus \left[\left(\mathbf{3}, \mathbf{2}, \tfrac{1}{3}\right) \oplus \left(\bar{\mathbf{3}}, \mathbf{1}, -\tfrac{4}{3}\right) \oplus (\mathbf{1}, \mathbf{1}, 2)\right]_A \\
&= \mathbf{15} \oplus \mathbf{10}.
\end{aligned}$$

In going from the first to the second line, we use that in the multiplication of two $\mathbf{U(1)}$'s we can simply add the hypercharges and in going from the second to the third we used the multiplication $\mathbf{3} \otimes \mathbf{3} = \mathbf{6}_S \oplus \bar{\mathbf{3}}_A$ in $\mathbf{SU(3)}$ and $\mathbf{2} \otimes \mathbf{2} = \mathbf{3}_S \oplus \mathbf{1}_A$ in $\mathbf{SU(2)}$. Similarly with

$$\bar{\mathbf{5}} = \left(\bar{\mathbf{3}}, \mathbf{1}, \frac{2}{3}\right) \oplus (\mathbf{1}, \mathbf{2}, -1)$$

and with $\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8}_S \oplus \mathbf{1}_A$, we can calculate $\mathbf{5} \otimes \bar{\mathbf{5}} = \mathbf{24}_S \oplus \mathbf{1}_A$ with the conclusion

$$\mathbf{24} = (\mathbf{8}, \mathbf{1}, 0) \oplus \left(\mathbf{3}, \mathbf{2}, \frac{5}{3}\right) \oplus \left(\bar{\mathbf{3}}, \mathbf{2}, -\frac{5}{3}\right) \oplus (\mathbf{1}, \mathbf{3}, 0) \oplus (\mathbf{1}, \mathbf{1}, 0).$$

These are simply results from multiplying group representations and decomposing them into other representations. Where can the fields assorted as (8.1) be put in place in the representations of $\mathbf{SU}(\mathbf{5})$? Remembering that each generation of fermions has 15 members, one might try to accommodate all of them in the **15**-representation. But this contains color-sextets and thus does not fit[4]. The **10**-representation contains already three of the species in (8.1). The **5** representation is not quite adequate, but $\bar{\mathbf{5}} = (\bar{\mathbf{3}}, \mathbf{1}, 2/3) \oplus (\mathbf{1}, \bar{\mathbf{2}}, 1)$ is. Therefore fermions most naturally fit into a $\bar{\mathbf{5}} \oplus \mathbf{10}$ representation of **SU(5)**. The leptons and quarks of each family can be gathered into a vector and an antisymmetric tensor

$$\bar{5}: \begin{pmatrix} d_r^c \\ d_g^c \\ d_b^c \\ e^- \\ \nu_e \end{pmatrix}_L \qquad \mathbf{10}: \begin{pmatrix} 0 & -u_b^c & u_g^c & u_r & d_r \\ . & 0 & -u_r^c & u_g & d_g \\ . & . & 0 & u_b & d_b \\ . & . & . & 0 & e^+ \\ . & . & . & . & 0 \end{pmatrix}_L . \tag{8.3}$$

Here we observe another interesting feature of a GUT. Because the charge operator Q needs to be a generator of the group, the sum of charges within a multiplet is zero. Consider in particular the $\bar{\mathbf{5}}$ representation for which

$$3Q(d^c) + Q(e^-) = 0.$$

Therefore the **SU(5)** GUT explains the fractional charge of the quarks as resulting from the number of colors. Together with $Q(u) = 1 + Q(d)$ we get for the charge of the proton because of its flavor composition $Q(p) = Q(uud) = 2 + 3Q(d) = -\, Q(e)$. Of course, the stability of the H-atom and the stability of other atoms hinges on $Q(p) + Q(e^-) = 0$. Experimentally, $|Q(p) + Q(e^-)| < 10^{-21} Q(p)$.

At and beyond the super-heavy scale there is only one coupling constant $g \simeq g_C \simeq g_I \simeq g_Y$ where g_C, g_I, and g_Y are the coupling constants of $\mathbf{SU(3)_C}$, $\mathbf{SU(2)_I}$, $\mathbf{U(1)_Y}$. In the Standard Model g_I and g_Y defined in the gauge boson couplings $g_I I_\mu^a + g_Y Y/2 B_\mu$ (see (6.28)) are related by $\tan\theta_W = g_Y/g_I$. The couplings could be rescaled by a rescaling of the gauge fields. We fixed, however, for the generators T^a of the gauge group $\mathrm{Tr}(T^a T^b) = 1/2\delta^{ab}$ in order to arrive at the usual F^2 kinetic term for the gauge fields in the Lagrangian (see the derivation of (5.74)). In the Standard Model, the two terms can be normalized independently of each other. But here, in the case of a simple gauge group, the relative

[4] There is also a 45-representation, which may give rise to hope for accommodating all three generations of quarks and leptons into this multiplet, but again its branchings into the Standard Model factor groups are inappropriate.

strength of the two couplings g_I and g_Y depends on the normalization of $(Y/2)$ relative to I_3. For the representation **5** we calculate

$$\mathrm{Tr}\left(\frac{Y}{2}\right)^2 = 3\left(\frac{1}{3}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{5}{6} \qquad \mathrm{Tr}\,(I_3)^2 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}.$$

Therefore, the generators I_3 and $\sqrt{\frac{3}{5}}(\frac{Y}{2})$ when properly normalized give the grand unification term including the electro-weakons is $g_I(I^a_\mu + \sqrt{\frac{3}{5}}(\frac{Y}{2})B_\mu)$. In other words we get

$$\tan\theta_W = \sqrt{\frac{3}{5}} \qquad \text{or} \qquad \sin^2\theta_W = \frac{3}{8} = 0.385.$$

This is far from the experimentally-observed value $\sin^2\theta_W = 0.23161 \pm 0.00018$. But bear in mind that the theoretical value holds at the grand unification scale. In order to compare it with the experimental value, one has to use the renormalization group flow down to energies corresponding to the W-boson mass, originally investigated in [215], see also [369]. This results in $\sin^2\theta_W(M_W) = 0.21$. Given the quite well-known experimental value this is still off by 10 %. Strikingly there is no longer a deviation if one makes the **SU(5)** model supersymmetric. In any case, notice that the Weinberg angle is determined by group theory.

As shown in Sect. 6.4.3, the Standard Model is free of anomalies, the reason being the pairing of leptons and quarks in each generation. Since in the **SU(5)** model a different representation has been chosen, one may wonder whether the model is anomaly-free too. It turns out that–remarkably–the $\mathbf{5}^*$ and the **10** have equal but opposite anomalies. Take for instance the anomaly-free condition (6.37) for $T^a = T^b = T^c = Q$. Then

$$\begin{aligned}\sum_{\bar{\mathbf{5}},\mathbf{10}} \mathrm{Tr}(Q^3) &= \left[3\left(\frac{1}{3}\right)^3 + (-1)^3\right] + \left[3\left(\frac{2}{3}\right)^3 + 3\left(-\frac{2}{3}\right)^3 + 3\left(-\frac{2}{3}\right)^3 + 1^3\right] \\ &= -\frac{8}{9} + \frac{8}{9} = 0.\end{aligned}$$

The algebra $\mathfrak{su}(5)$ is spanned by 24 generators, in physical terms corresponding to 24 gauge bosons. Since the Standard Model has 8 gluons and 4 electro-weakons, 12 further gauge bosons are to be introduced in the model. Indeed we need only two new basic types, called the X- and the Y-bosons, since they carry color and come pairwise as particles and anti-particles. The gauge bosons fit into the adjoined representation $\mathbf{24} = \overline{\mathbf{24}}$ representation by the assignment

$$\begin{aligned}(\mathbf{8},\mathbf{1},0) &\longleftrightarrow \text{gluons} \\ (\mathbf{1},\mathbf{3},0) &\longleftrightarrow W^1, W^2, W^3 \\ (\mathbf{1},\mathbf{1},0) &\longleftrightarrow B \\ \left(\mathbf{3},\mathbf{2},\frac{5}{3}\right) &\longleftrightarrow X^\alpha, Y^\alpha \\ \left(\bar{\mathbf{3}},\mathbf{2},-\frac{5}{3}\right) &\longleftrightarrow \overline{X}^\alpha, \overline{Y}^\alpha.\end{aligned}$$

We read from the multiplet structure that if the boson X^{α}_{μ} ($\alpha = 1, 2, 3$ the color index) is associated to $T_3 = {}^1\!/_2$ it has the charge ${}^4\!/_3$, and the Y^{α}_{μ} has charge ${}^1\!/_3$.

Given the allocations of fermions and gauge bosons to the representations of **SU(5)**, one next may construct the Lagrangian according to the Yang-Mills recipe. The Lagrangian contains the quadratic field strength part with the field strength built from the gauge fields F^{a}_{μ} of **SU(5)**. Furthermore, there is the interaction part with the coupling of the gauge bosons to the fermions in the form $F^{a}_{\mu} j^{\mu}_{a}$, where the currents are built as bilinear forms in the fermion fields respecting the previous assignment of quantum numbers. These bilinears also contain $\mathbf{SU(5)}$ representation matrices stemming from the covariant derivatives. All (boson-fermion) interaction terms are proportional to only one coupling constant g_5 (including the previously derived factor $3/5$ for the **U(1)** field B_{μ}). The interaction Lagrangian can be split into the part for the Standard Model and a further one including the terms with the X- and the Y- bosons. This new term mediates (quark-lepton)-transitions.

The extension of the spontaneous symmetry breaking mechanism in the GSW-model is more or less straightforward: Scalar fields with an appropriate potential need to be introduced. In a first step a 24-component real scalar field (with suitably chosen charges) breaks $\mathbf{SU(5)}$ down to the Standard Model group $\mathbf{SU_C(3)} \times \mathbf{SU_I(2)} \times \mathbf{U_Y(1)}$. Thereby the X- and Y-bosons become massive, while at this stage the other vector bosons remain massless. Next, the Standard Model symmetry breaking via a scalar Φ (according to 6.31)) must be mimicked within **SU(5)**. We recover this scalar in the $\Phi = (\mathbf{1}, \mathbf{2}, 1)$ part of the **5**-representation (8.2). The other part is a color triplet $\Psi = (\mathbf{3}, \mathbf{1}, -2/3)$ for which one might assume a non-vanishing vacuum expectation value. Technically everything seems to be feasible. There are, however, some pitfalls. For instance, the procedure of symmetry breaking in two steps, with each of the associated scalar bosons having its own minimum, is not renormalizable. One needs to add a gauge-invariant interaction term for the bosons. Interactions furthermore creep in due to radiative corrections. How can one prevent that the Higgs surviving the $\mathbf{SU(5)}$ breaking with its typical mass M_X and the electroweak Higgs keep apart? (This is one facet of the hierarchy problem; more about this below.)

As in the Standard Model case, fermion masses are generated by Yukawa interactions of the Higgs bosons with the fermions. The allowed interaction terms in the Lagrangian must respect the **SU(5)** symmetry. But from the very fact that the interaction couples two fermions to the scalars, we can infer structural results for the Yukawa terms. Since all fermions are in $(\bar{\mathbf{5}} \oplus \mathbf{10})$ we need to decompose $(\bar{\mathbf{5}} \oplus \mathbf{10}) \otimes (\bar{\mathbf{5}} \oplus \mathbf{10})$ into its constituents:

$$\bar{\mathbf{5}} \otimes \bar{\mathbf{5}} = \overline{\mathbf{10}} \oplus \overline{\mathbf{15}} \qquad \bar{\mathbf{5}} \otimes \mathbf{10} = \mathbf{5} \oplus \overline{\mathbf{45}} \qquad \mathbf{10} \otimes \mathbf{10} = \bar{\mathbf{5}} \otimes \mathbf{45} \oplus \mathbf{50}. \tag{8.4}$$

This shows that for fermion mass generation, the scalars are to be arranged in **5**, **10**, **15**, **45**, **50**. Two things are relevant here: Since the **24** does not arise in these decompositions, the scalar which is responsible for the GUT symmetry breaking does not participate in fermion mass generation. Furthermore, in the minimal version, the electroweak symmetry breaking scalar is in **5**. Only upon reflecting about non-minimal Higgs mechanisms are the other representations from (8.4) needed. The arrangement of quarks and leptons into common multiplets requires that the masses of the down quark and and electron are predicted to be the same (similarly $m_s = m_\mu$, $m_b = m_\tau$). Again this is definitely not true at our low energies, "low" with respect to 10^{15} GeV. But it is possible to rescale this by renormalization group techniques, and in fact one finds qualitative agreement with the experimental results.

In the **SU(5)** version, a proton may decay via the quark-lepton transitions $p \to d\bar{d}e^+$, $u\bar{u}e^+$, $u\bar{d}\bar{\nu}_e$ (with similar decays for the other two generations). Observe that $(B\text{-}L)$ is conserved. These semifinal states are of course not directly observable, but decays such as $p \to e^+\pi^0 \to e^+\gamma\gamma$ are. The life time of the proton is estimated to be of the order of 10^{28-30} years. Despite the uncertainties involved in the estimation, it is smaller than the experimental lower bound by at least two orders of magnitude. Although the **SU(5)** GUT does not agree with experiment, its concepts provide a framework for the investigation of other GUTs. At least it provides an existence proof for unified theories.

8.1.3 **SO(10)** *Grand Unification*

The group **SO(10)** contains **SU(5)** as a subgroup. Therefore, it is possible to express its representations in terms of **SU(5)**:

$$\mathbf{10} = \mathbf{5} \oplus \bar{\mathbf{5}} \qquad \mathbf{16} = \mathbf{10} \oplus \bar{\mathbf{5}} \oplus \mathbf{1} \qquad \mathbf{45} = \mathbf{24} \oplus \mathbf{10} \oplus \overline{\mathbf{10}} \oplus \mathbf{1}. \tag{8.5}$$

This looks quite intriguing: The **16**-representation displays the $\mathbf{10} \oplus \bar{\mathbf{5}}$ of the **SU(5)** GUT. The additional singlet may be referred to a right-handed neutrino. The (adjoint) **45** representation hosts the gauge bosons, and we find among these the **24** from the **SU(5)**-GUT. The singlet in here can be associated with the $(B\text{-}L)$ quantum number.

There is a simple and elegant argument to see that **SO(10)** is free of anomalies: The condition (6.37) reads in this case $D^{ijklmn} = \mathrm{Tr}(M^{ij}\{M^{kl}, M^{mn}\}) = 0$. The tensor D^{ijklmn} is invariant under a **SO(10)** transformation $M^{ij} \to O^T M^{ij} O$. Since one cannot build an invariant six-index tensor with the symmetry properties of the tensor D, like $D^{ijklmn} = -D^{jiklmn}$ etc., necessarily $D \equiv 0$.

There are various modes through which **SO(10)** can be broken, for instance

$$\begin{array}{ccccc} & & \mathbf{SU(4)}\times\mathbf{SU_L(2)}\times\mathbf{SU_R(2)}\searrow & & \\ \mathbf{SO(10)}{\nearrow \atop \searrow} & \longrightarrow & \mathbf{SU(5)} & \longrightarrow \mathbf{SU(3)}\times\mathbf{SU(2)}\times\mathbf{U(1)} & \to \mathbf{SU(3)}\times\mathbf{U(1)} \\ & & \mathbf{SU(4)}\times\mathbf{SU(2)}\times\mathbf{U(1)}\nearrow. & & \end{array}$$

According to these different modes different assignments of isospin, charge, and hypercharge to the symmetry breaking scalar fields become mandatory.

Let me mention further interesting features of the **SO(10)** GUT; details to be found for instance in [369]. It has a local $(B\text{-}L)$-symmetry. It conserves parity–interestingly enough, parity becomes a kind of an internal continuous symmetry. From a mathematical point of view, the most esthetic formulation makes use of **SO(10)** spinors. Thus a fermion which is understood as a **SO(3, 1)** spinor in four dimensional spacetime becomes a spinor in a $4 + 10$ dimensional "space-time".

8.1.4 Instead of a Conclusion

Here only the minimal **SU(5)** and the **SO(10)** GUTs were sketched. Other grand unifying groups have of course been investigated. I mentioned restrictions on these groups. These indeed only leave $\mathbf{SU(N)}$ for $N \geqslant 5$, $\mathbf{SO(4\,N+2)}$ for $N \geqslant 2$, and the exceptional group $\mathbf{E_6}$. Again, group theory (hence symmetries) and physics go hand in hand: The $\mathbf{E_6}$ symmetries can be split into a **SO(10)** and a **U(1)** part as

$$\mathbf{27} = \mathbf{16} \oplus \mathbf{10} \oplus \mathbf{1} \qquad\qquad \mathbf{78} = \mathbf{45} \oplus \mathbf{16} \oplus \overline{\mathbf{16}} \oplus \mathbf{1}$$

and quite naturally encompass the **SO(10)** representations **16** and **45** that served to host the fermions of one generation and the gauge bosons. Thus, you find in the literature investigations about $\mathbf{E_6}$ gauge theories and the chain of spontaneous symmetry breakings along

$$\mathbf{E_6} \overset{E}{\to} \mathbf{SO(10)} \overset{S}{\to} \mathbf{SO(5)} \overset{X,Y}{\to} \mathbf{SU(3)}\times\mathbf{SU(2)}\times\mathbf{U(1)} \overset{W,Z}{\to} \mathbf{SU(3)}\times\mathbf{U(1)}$$

where I indicated that in each of the steps, new gauge bosons become massive. You see that the theory of these large groups becomes rather involved. The group theory for building grand unified theories is in detail worked out in [475].

In any case, none of the GUTs is able to predict the number of generations to be three. Also the number of free parameters compared to the Standard Model is not drastically reduced: Although there is only one coupling constant and there are fewer masses, one needs an increased number of parameters in the Higgs sector.

The most serious problem is the so-called "hierarchy problem". Take for example the **SU(5)** GUT. There are roughly a dozen orders of magnitude between the electroweak and the leptoquark mass scales, given by $M_W/M_X = 10^{-13}$. It needs an incredible fine-tuning to keep these masses apart in each order of perturbation theory. And if they are kept apart, does this mean that there is no new physics in the "desert" between M_W and M_X? Also observe that on this scale the Planck mass $M_P \approx 10^{19}$ GeV on which gravity effects should become relevant is not "far away". And it is even nearer in the GUTs with larger symmetry groups.

8.2 Kaluza-Klein Theory

8.2.1 Kaluza's and Klein's Contributions to the KK Theory

The names of Theodor Kaluza and Oskar Klein are used to merge the contributions of these two persons into something called "Kaluza-Klein theory" or–given its not established status–"KK models".

Th. Kaluza

Kaluza's basic idea was to consider Einstein's theory of gravity in a five dimensional (pseudo-) Riemannian spacetime with the line element $d\sigma^2 = \hat{g}_{AB} d\hat{x}^A d\hat{x}^B$

where $A, B = (0, \dots, 4)$, or with the 4+1 split

$$(\hat{g}_{AB}) = \begin{pmatrix} g_{\mu\nu} & A_\mu \\ A_\nu & V \end{pmatrix}, \tag{8.6}$$

the greek indices running from $0, \dots, 3$. Kaluza made the assumption that the components of the fields $g_{\mu\nu}$, A_μ, and V do not depend on the fifth coordinate,

$$\partial_4 \hat{g}_{AB} = 0, \tag{8.7}$$

called the *cylinder condition*. It seems that he introduced this heuristic condition in order to account for the non-observability of the fifth dimension. Kaluza then showed that the fifteen linearized vacuum field equations $\hat{R}_{AB} = 0$ after making the 4+1 split, become the ten Einstein vacuum equations for the metric $g_{\mu\nu}$, the four Maxwell equations for A_μ, and a field equation for V. Furthermore, he showed that–again in the linear approximation–the geodesic equations for a free particle in five dimensions split into the equations of motion for a charged particle in the four dimensions and into a relation of proportionality between the fifth component of the particle's momentum and the electric charge. Because of the cylinder condition this momentum is conserved and thus charge conservation originates from the fifth dimension. These intriguing observations, however, had their price. Perhaps the hardest to pay is that spacetime has five dimensions. And why the cylinder condition, anyway? Furthermore, there is a scalar field that does not manifest itself in nature. Another peculiarity is that the matter fields are part of the geometry but do not serve as a source for the gravitational field.

When A. Einstein received the manuscript from T. Kaluza, a mathematician at the University of Königsberg (now Kaliningrad, Russia), he was quite amazed as can be seen from a letter to Kaluza from April 21st, 1919, in which he wrote "The idea of achieving [a unified field theory] by means of a five-dimensional cylinder world never dawned on me. [...] At first glance I like your idea enormously." Kaluza had found an elegant way to combine gravity and electromagnetism in a single (classically) coherent "unified field theory". In fact, Einstein allowed himself two years before he accepted Kaluza's paper in December 1921 for the Proceedings of the Prussian Academy of Science [309]. Although Kaluza was designated in a German book as the "man who invented the fifth dimension", this is actually not fair, since already in 1914 the Finnish theoretician G. Nordström had this idea–including the cylinder condition; for the early history see [228], [430], and the introduction to [11].

O. Klein

In 1926 Oscar Klein[5] reconsidered the work of Kaluza, and extended it in various ways. Altogether, Klein wrote four publications on this topic between 1926–1928.

[5] The same person as in the Klein-Gordon equation and in various other Klein-XX, but unrelated to F. Klein, the supervisor of Emmy Noether.

In [321] he first argued for a parametrization of the five metric different from (8.6). Another parametrization seems only to be a matter of taste, but different parametrization lead to field equations which may differ drastically in their appearance. In Kaluza's parametrization, the field equations become rather nasty–and this seems to be the reason why Kaluza showed the equivalence of 5D General Relativity with the 4D Einstein-Maxwell theory only in the linear approximation. For finding a more appropriate parametrization, Klein utilized symmetry arguments (what else?). He reflected upon the symmetries that remain after imposing the cylinder condition. This condition is of course not generally covariant. In denoting the coordinates as $x^A = (x^\mu, y)$ one finds that the condition (8.7) remains valid under coordinate transformations only if they are of a restricted form

$$x'^\mu = f^\mu(x) \qquad\qquad y' = \lambda y + f(x), \tag{8.8}$$

where λ is a constant. Here, let us set $\lambda = 1$; the case $\lambda \neq 1$ is related to Weyl rescalings and will be discussed separately. The change in the metric under these transformations can be calculated from

$$\hat{g}'_{AB} = \frac{\partial x^C}{\partial x'^A}\frac{\partial x^D}{\partial x'^B}\hat{g}_{AB}.$$

At first, one establishes that $\hat{g}'_{44} = \hat{g}_{44}$. Thus it transforms as a scalar with respect to (8.8). In the sequel this scalar is denoted $\phi = \hat{g}_{44}$. Further

$$\hat{g}'_{\mu 4} = f^\nu_{,\mu}\hat{g}_{\nu 4} - f_{,\mu}\phi. \tag{8.9}$$

Therefore the field

$$B_\mu := \frac{\hat{g}_{\mu 4}}{\phi}$$

transforms as $B'_\mu = f^\nu_{,\mu}B_\nu - f_{,\mu}$. Hence B_μ behaves as a covariant vector under the subset of transformations out of (8.8) for which the y-coordinate is unchanged ($f = 0$), and like a gauge potential for transformations with $x'^\mu = x^\mu$. From (8.9), Klein further derived the invariance of the differential $d\theta = dy + \phi^{-1}\hat{g}_{\mu 4}dx^\mu$. Similarly $ds^2 = \left(\hat{g}_{\mu\nu} - \phi^{-1}\hat{g}_{\mu 4}\hat{g}_{\nu 4}\right)dx^\mu dx^\nu$ stays invariant. Therefore the full line element can be written as

$$d\sigma^2 = ds^2 + \phi d\theta^2 = g_{\mu\nu}dx^\mu dx^\nu + \phi(dy + B_\lambda dx^\lambda)^2 \tag{8.10}$$

where ds^2 is identified with the line element of the 4-geometry. Thus the five dimensional metric becomes

$$(\hat{g}_{AB}) = \begin{pmatrix} g_{\mu\nu} + \phi B_\mu B_\nu & \phi B_\mu \\ \phi B_\nu & \phi \end{pmatrix} \tag{8.11a}$$

$$(\hat{g}^{AB}) = \begin{pmatrix} g^{\nu\lambda} & -B_\rho g^{\nu\rho} \\ -B_\rho g^{\rho\lambda} & \phi^{-1} + B_\rho B_\sigma g^{\rho\sigma} \end{pmatrix}. \tag{8.11b}$$

This parametrization of the metric was independently derived by Einstein [159]; the work of Klein was brought to his attention by H. Mandel, who himself reconsidered Kaluza's work. In his first paper [321] Klein also attempted to link the five-dimensional theory to quantum mechanics. Being rather ingenious in argumentation, it seems a little hard to digest from today's viewpoint. Nevertheless he implicitly assumed that the fifth dimension is closed with $0 \leqslant y/r \leqslant 2\pi$, where r is the radius of the circle S^1. In another 1926 article [322] he made this explicit and derived a relation $r = \hbar c\sqrt{2\kappa}/e$ showing the dependency of the radius of the circle on Planck's constant, the vacuum velocity of light, the gravitational constant, and on the elementary electric charge (this will be derived later in this section; see (8.23)). This gave a hint that the extra dimension is small and of the size of the Planck length. Thus it is the second 1926 article by Klein that laid the origin of what is now called "compactification". Instead of using the original argumentation of Klein, let us follow a modern approach.

8.2.2 The 5D Model

Field Equations

The modern approach–and the one pursued in this book–is to start with an action in a five-dimensional Riemann space V_5:

$$S_{KK} = \frac{1}{2\hat{\kappa}} \int d^5z \sqrt{-\hat{g}}\hat{R}. \tag{8.12}$$

Here $\hat{\kappa}$ is the gravitational Einstein constant in five dimensions, and $d^5z = dy\, d^4x$. The five-dimensional curvature scalar is defined in terms of the five-dimensional metric and Ricci tensor as $\hat{R} = \hat{g}_{AB}\hat{R}^{AB}$ where

$$\hat{R}_{AB} = \partial_C \hat{\Gamma}^C{}_{AB} - \partial_B \hat{\Gamma}^C{}_{AC} + \hat{\Gamma}^C{}_{AB}\hat{\Gamma}^D{}_{CD} - \hat{\Gamma}^C{}_{AD}\hat{\Gamma}^D{}_{BC},$$
$$\hat{\Gamma}^C{}_{AB} = \frac{1}{2}\hat{g}^{CD}\left(\partial_A \hat{g}_{DB} + \partial_B \hat{g}_{DA} - \partial_D \hat{g}_{AB}\right).$$

All geometric objects (connections, Ricci tensor, ...) can–like the metric–be split into a (4+1) form, and be explicitly computed. This is a little tedious, see e.g. [49], [141]. The field equations resulting from the action (8.12) are–surface terms neglected–obviously

$$\hat{G}_{AB} = \hat{R}_{AB} - \frac{1}{2}\hat{g}_{AB}\hat{R} = 0. \tag{8.13}$$

Given the cylinder condition $\partial_y \hat{g}_{AB} = 0$ the (4+1)-split of (8.13) becomes equivalent to the set of field equations

$$G_{\mu\nu} = \frac{\gamma^2\phi}{2}T^{EM}_{\mu\nu} - \phi^{-1/2}\left(D_\mu D_\nu\sqrt{\phi} - g_{\mu\nu}\Box_x\sqrt{\phi}\right) \tag{8.14a}$$

$$D_\mu(\phi^{3/2}F^{\mu\nu}) = 0 \tag{8.14b}$$

$$\Box_x\sqrt{\phi} = \frac{\gamma^2\phi^{3/2}}{4}F_{\mu\nu}F^{\mu\nu}. \tag{8.14c}$$

Here $G_{\mu\nu}$ is the (four-dimensional) Einstein tensor, and $\Box_x = \eta^{\mu\nu}\partial_\mu\partial_\nu$. Further I've set

$$B_\mu = \gamma A_\mu$$

with a (coupling) constant γ, to be determined, and identify $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ with the field strength, and $T^{EM}_{\mu\nu}$ with the symmetrized energy-momentum tensor (3.64) of electrodynamics. D_μ is the covariant derivative with respect to the connection $\Gamma^\lambda_{\mu\nu}$. The equations (8.14) are "almost" the field equations of a Maxwell field coupled to gravity. Only if the scalar field ϕ is constant (and positive) they are equivalent to the Einstein and Maxwell field equations

$$G_{\mu\nu} = \frac{\gamma^2\phi}{2} T^{EM}_{\mu\nu} \qquad\qquad \partial_\mu F^{\mu\nu} = 0$$

provided that $\gamma = \sqrt{\frac{2\kappa}{\phi}}$. Contrary to Kaluza's formulation (of the same theory, albeit with another interpretation), here the electromagnetic field serves as a source for the gravitational field. Also, observe that the formation of the Einstein-Maxwell theory is exact within this frame and is not only valid approximately in a linearized treatment.

Instead of disassembling the five-dimensional field equations into their components, we could immediately take the original action (8.12) and express the Lagrangian density in its $(4+1)$ form. At first, we observe that for the metric (8.11a), we have $\sqrt{-\hat{g}} = \sqrt{-g}\sqrt{\phi}$, where g is the determinant of the 4-metric. Further

$$\hat{R} = R_4 - \frac{\gamma^2}{4}\phi F_{\mu\nu}F^{\mu\nu} + \frac{2}{\sqrt{\phi}}\Box_x\sqrt{\phi}$$

where R_4 is the curvature scalar in V_4. The action (8.12) can then be written in a form such that it reveals the four-dimensional Einstein term, the Maxwell term and the scalar term as its ingredients. Given the cylinder condition, none of the fields depends on the fifth coordinate. Thus one can simply integrate with respect to y and arrives at

$$S_{KK} = \frac{1}{2\hat{\kappa}}\int dy \int d^4x\sqrt{-g}\sqrt{\phi}\Big(R_4 - \frac{\gamma^2}{4}\phi F_{\mu\nu}F^{\mu\nu} + \frac{2}{\sqrt{\phi}}\Box_x\sqrt{\phi}\Big).$$

Only if $\phi = \phi_c$ is a constant is the first term identical to the Hilbert-Einstein action, while the second is the Maxwell action if we identify $\hat{\kappa} = \kappa\sqrt{\phi_c}\int dy$ and, again, $\gamma^2\phi = 2\kappa$. This is a significant hint: In order that the y integral becomes finite, the extra dimension must be closed with a radius r: $\hat{\kappa} = \kappa(2\pi r\sqrt{\phi_c})$. By introducing $\tilde{\phi} = \phi/\phi_c$ we can write

$$S_{KK} = -\int d^4x\sqrt{-g}\sqrt{\tilde{\phi}}\left(\frac{1}{2\kappa}R_4 + \frac{1}{4}\tilde{\phi}F_{\mu\nu}F^{\mu\nu} + (\text{terms in } \tilde{\phi})\right). \qquad (8.15)$$

The emergence of 4D general relativity coupled to electromagnetic fields from a five-dimensional GR formulation is technically the *dimensional reduction* from $V_4 \times S^1$

to V_4. In the literature, this is often referred to as the "Kaluza-Klein-miracle". But be aware that the "miracle" only occurs if the 44-component of the metric is assumed to be constant. And this stance was taken in most of the old literature on the KK-model. But why assume that $\hat{g}_{44} = \text{const}$? It is not even reasonable, since due to the field equation (8.14c) it has as a consequence $F_{\mu\nu}F^{\mu\nu} = 0$, and this is a severe restriction on the Maxwell field. This was first pointed out by P. Jordan and by Y. R. Thiry in the late 1940's who then tried to make use of the scalar in a theory in which the gravitational constant is replaced by a dynamical field. Their work also lead to a revival of Brans-Dicke theories for gravitation.

Conformal Rescaling

The action (8.15) has the factor $\sqrt{\phi}$ before the gravitational part. This annoying feature can be avoided by a conformal rescaling

$$\hat{g}_{AB} \longrightarrow \bar{g}_{AB} = \Omega^2 \hat{g}_{AB}.$$

A convenient choice is $\Omega^2 = \phi^{-1/3}$. This amounts to the metric

$$(\bar{g}_{AB}) = \phi^{-1/3} \begin{pmatrix} g_{\mu\nu} + \gamma^2 \phi A_\mu A_\nu & \gamma\phi A_\mu \\ \gamma\phi A_\nu & \phi \end{pmatrix} \tag{8.16}$$

and to the conformally rescaled action

$$S = \int d^4x \sqrt{-\bar{g}} \left(\frac{\bar{R}}{2\kappa} - \frac{1}{4}\phi \bar{F}_{\mu\nu}\bar{F}^{\mu\nu} - \frac{1}{6\gamma^2} \frac{\bar{\partial}_\mu \phi \bar{\partial}^\mu \phi}{\phi^2} \right).$$

Here all quantities with an "overbar" refer to the rescaled metric, and $\bar{\partial}_\mu \phi = \bar{g}^{\mu\nu}\partial_\mu \phi$. Now indeed after this rescaling the gravitational part has the "conventional" form. We also see that the correct ratio of the Einstein and the Maxwell term requires $\phi \geqslant 0$, and thus the extra dimension must be taken to be space-like. But here, the problem of conformal ambiguity pops up, namely which field to identify with the gravitational field: either $g_{\mu\nu}$ as in the Klein metric (8.11a) or $\bar{g}_{\mu\nu}$ as in (8.16). Various investigations tend to regard the rescaled metric as the "true" one; see [400] for a brief review with hints to further references. Indeed you often will find the parametrization (8.16) in the literature.

Geodesics

As mentioned before, both Kaluza and Klein investigated the geodesic equation in $D = 5$. The Lagrangian for a structureless particle with mass m can simply be read off from (8.10) as:

$$L = \frac{m}{2}[g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu + \phi(\dot{y} + \gamma A_\lambda \dot{x}^\lambda)^2].$$

Because of the cylinder condition the fifth coordinate is a cyclic coordinate (i.e. $\partial L/\partial y = 0$), and therefore, from its equation of motion, we get the conservation law for the momentum in the fifth dimension:

$$p := \frac{\partial L}{\partial \dot{y}} = m\phi(\dot{y} + \gamma A_\lambda \dot{x}^\lambda) = \text{const.}$$

The other equations (those for the x^μ) are, by using the expressions for the connections,

$$\Gamma^\mu{}_{\nu 4} = -\frac{\gamma}{2}(F^\mu{}_\nu \phi + A_\nu \partial^\mu \phi) \qquad\qquad \Gamma^\mu{}_{44} = -\frac{1}{2}\partial^\mu \phi,$$

exposed as

$$\ddot{x}^\mu + \Gamma^\mu{}_{\rho\sigma}\dot{x}^\rho \dot{x}^\sigma = -2\Gamma^\mu{}_{\nu 4}\dot{x}^\mu \dot{y} - \Gamma^\mu{}_{44}\dot{y}^2 = \frac{p\gamma}{m}F^\mu{}_\nu \dot{x}^\nu + \frac{p^2}{2\,m^2}\phi^{-2}\partial^\mu\phi + \mathcal{O}(\gamma^2). \tag{8.17}$$

The first term on the right-hand side can be identified with the Lorentz force if

$$\frac{p\gamma}{m} = \frac{q}{c},$$

where q is the charge of the particle. This is the observation first made by Kaluza, that charge is conserved because of the fifth dimension and the cylinder condition. The second term in (8.17) drops out again for $\phi = \text{const}$. Otherwise its role in the geodesic equation remains obscure at this stage. In any case, electrically neutral point particles moving along geodesics in V_5 mimic the motion of charged particles in V_4. This is remarkable: What is called a charge depends on the dimension.

Harmonic Expansion

The cylinder condition of Kaluza and the compactifcation idea of Klein amount to assuming that the five-dimensional manifold is the product of a four-dimensional (pseudo-)Riemannian spacetime and a circle: $V_5 = V_4 \times S^1$. The vacuum (or ground) state is therefore presumed to be no longer the five-dimensional Minkowski space M_5, but $M_4 \times S^1$. All fields can be expanded around this ground state as periodic functions of the coordinate y on the circle. If r is the radius of the circle and θ the angular coordinate we have $y = r\theta$. Therefore all field quantities $\mathcal{F} = \{g_{\mu\nu}, A_\mu, \phi\}$ admit a harmonic expansion

$$\mathcal{F}(x, y) = \sum_{n=-\infty}^{+\infty} \mathcal{F}_n(x) Y_n(y) \qquad\qquad \mathcal{F}_n^* = \mathcal{F}_{-n} \tag{8.18}$$

which in one dimension is nothing but the Fourier expansion with

$$Y_n(y) = \frac{1}{\sqrt{2\pi r}} e^{iny/r} \qquad\qquad \int dy\, Y_n Y_m = \delta_{n,-m} \quad . \tag{8.19}$$

We now may expand the metric $\hat{g}_{AB}$ around the ground state metric

$$< \hat{g}_{AB} >= \hat{\eta}_{AB} = \begin{pmatrix} \eta_{\mu\nu} & 0 \\ 0 & 1 \end{pmatrix} \tag{8.20}$$

as $\hat{g}_{AB} = \hat{\eta}_{AB} + \hat{h}_{AB}$, where $\hat{h}$ is a small fluctuation. Now choosing for instance the DeDonder coordinate conditions (7.53a) in 5D to arrive at[6]

$$\hat{\Box}\hat{h}_{AB} = (\hat{\eta}^{AB}\partial_A\partial_B)\,\hat{h}_{AB} = 0,$$

reveals that the $\hat{h}_{AB}$ are massless fields in 5D. Expanding these according to (8.18) we get

$$\hat{\Box}\hat{h}_{AB} = (\Box_x + \partial_y^2)\sum_{n=-\infty}^{+\infty} h^n_{AB}(x)Y_n(y) = 0$$

or–with $\partial_y^2 Y_n = -(n^2/r^2)Y_n$

$$(\Box_x - \frac{n^2}{r^2})h^n_{AB} = 0.$$

This is an infinite number of fields (called by some the *KK tower*) with masses $m_n = (n\hbar c)/r$. Only the lowest mode $(n = 0)$ is massless. We will see in a moment that the $(n \neq 0)$ masses are of the order of the Planck length. For sufficiently small energies these modes cannot be excited and it becomes reasonable to retain only the lowest one. Thus the previously imposed cylinder condition amounts to keeping only the $n = 0$ modes of the fields.

Charge Quantization

In order to derive relations among the masses, the charges and the compactified dimensions consider the dynamics of a real scalar field Ψ with mass M in five dimensions, defined by the action

$$S_\Psi = \frac{1}{2}\int d^4x\,dy\sqrt{\phi}\left(\hat{g}^{AB}\partial_A\Psi\partial_B\Psi - M^2\Psi^2\right).$$

Investigate this in "flat" space defined by

$$\check{g}_{AB} = \begin{pmatrix} \eta_{\mu\nu} + B_\mu B_\nu & B_\mu \\ B_\mu & 1 \end{pmatrix}.$$

With the inverse metric (8.11b) we get for the first term in the integrand

$$\check{g}^{AB}\partial_A\Psi\partial_B\Psi = \eta^{\mu\nu}\big[(\partial_\mu - B_\mu\partial_y)\Psi\big]\big[(\partial_\nu - B_\nu\partial_y)\Psi\big] + (\partial_y\Psi)^2.$$

[6] This calculation parallels the "derivation" of gravitational waves in 4D.

On expanding $\Psi(x, y)$ into harmonics according to (8.18), the action becomes

$$S_\Psi = \int d^4x\, dy \sum_{m,n} Y_n Y_m \Big\{ \frac{1}{2}\big[(\partial_\mu - \frac{im}{r} B_\mu)\psi_m\big]\big[(\partial^\mu - \frac{in}{r} B^\mu)\psi_n\big] + \frac{1}{2}\big[\frac{mn}{r^2} - M^2\big]\psi_m\psi_n \Big\}.$$

We can carry out the y-integration by making use of the orthogonality relation (8.19). On further replacing $B_\mu = \gamma A_\mu$, we finally arrive at

$$S_\Psi = \int d^4x \sum_{n=0}^{\infty} \Big\{ \frac{1}{2}\big[(\partial_\mu - iq_n A_\mu)\psi_n\big]\big[(\partial^\mu + iq_n A^\mu)\psi_n^*\big] - \frac{1}{2} m_n^2 \psi_n \psi_n^* \Big\}. \quad (8.21)$$

This describes a tower of Kaluza-Klein scalars ψ_n with charges

$$q_n = n\frac{\gamma}{r} = n\frac{\sqrt{2\kappa}}{r}. \quad (8.22)$$

The lowest one $(n = 0)$ is a neutral scalar. Since the charge is a multiple of the elementary charge e, and if we replace the suppressed power of $\hbar$:

$$e^2 = \frac{2\kappa}{r^2}\hbar^2. \quad (8.23)$$

On the other hand, this relation can be used to determine the radius of compactification:

$$\alpha = \frac{e^2}{4\pi\hbar c} \approx \frac{1}{137} \qquad \curvearrowright \qquad r = \frac{2}{\sqrt{\alpha}}\sqrt{\frac{G\hbar}{c}} \approx 0,17 L_{Pl}$$

where L_{Pl} is the Planck length. The KK-tower scalar fields have masses

$$m_n^2 = M^2 + n^2 \frac{\hbar^2 c^2}{r^2} \approx M^2 + n^2 M_{Pl}^2. \quad (8.24)$$

Since r is of the order of the Planck length, all modes aside from the zero mode become very heavy, and can be neglected at sufficiently low energies $(E \ll E_{Pl} = 1.2 \times 10^{16}$ TeV). Thus even the most powerful accelerators operate on this scale in the “low-energy” range. Things are quite different in the very early universe. Thus it is no surprise that much work has been done on Kaluza-Klein cosmology. Current experiments limit the size of the fifth dimension to $r \leqslant 10^{-19}$ m.

Charge quantization due to an extra dimension is intriguing, but it has its flaws. For instance from (8.22) we realize that charged particles have $n \neq 0$. But then according to (8.24) their masses are of the order of the Planck mass.

Invariances and Symmetries

After all of the previous manipulations on the action, the field equations and the ground state of the KK model, it is rather beneficial to reflect on the symmetries of the model in its various formulations.

The original action (8.12), namely Hilbert-Einstein gravity in 5D, is invariant under general coordinate transformations $z' = z'(z)$. The field Eqs. (8.13) have these symmetries as well. If the ground state of this theory is assumed to be 5D Minkowski spacetime M_5, also the ground state is invariant. Now the central assumption of the KK models is to consider as ground state the manifold $M_4 \times S^1$. So far, we did not care about the justification of this assumption. In non-gravitational physics a "ground state" of a physical system is defined as the stable solution of the dynamical equations with the lowest energy. For gravitational theories, however, this definition is of no use, because as discussed in length in Sect. 7.5.4, in these theories the concept of energy is ambiguous. What seems to be clear is that the definition of energy hinges on the boundary condition, and only configurations with the same boundary conditions can be compared. And also the quest for stability cannot discriminate between M_5 and $M_4 \times S^1$; for more about these topics concerning "ground state and stability" see [49].

We realized that the cylinder condition is an approximation of the full theory–higher modes in the harmonic expansion being neglected. In this case, the coordinate transformations are restricted to the form (8.8), or infinitesimally

$$x'^{\mu} = x^{\mu} + \xi^{\mu}(x) \qquad\qquad y' = y + \epsilon y + \zeta(x).$$

Let these be divided into three distinct classes

- $\epsilon = 0,\ \zeta(x) = 0$: The remaining symmetry transformations are coordinate transformations in V_4 under which the fields $g_{\mu\nu}$, B_μ and ϕ transform as a second order tensor, vector and scalar, respectively. Both the field equations and the action (8.15) are invariant.
- $\epsilon = 0,\ \xi^{\mu}(x) = 0$: As derived before, this results in $\bar{\delta}\phi = 0$, $\bar{\delta}B_\mu = -\partial_\mu\zeta$, $\bar{\delta}g_{\mu\nu} = 0$. This was the observation that the B_μ transform like gauge potentials. But be aware that the transformation does not distinguish a $\mathbf{U(1)}$ symmetry from a $\mathbf{T_1}$ symmetry in the action (8.15) because the y-dependence is integrated out.
- $\xi^{\mu}(x) = 0,\ \zeta(x) = 0$: This case $(\lambda \neq 1)$ was postponed before.

  $$\delta\phi = -2\epsilon\phi \qquad\qquad \delta B_\mu = \epsilon B_\mu \qquad\qquad \delta g_{\mu\nu} = 0.$$

  The field equations (8.14) are invariant under these global scale transformations. At first sight the action (8.15) is not. But if one remembers that $\hat{\kappa}$ and κ are related by a factor $\int dy = \epsilon$, the invariance of the action can be achieved by rescaling the coupling constants.

Strictly speaking the action (8.15) is valid for the zero modes in the harmonic expansion according to (8.18), and I should have this indicated by a further index (0) on the fields. Thus $g^{(0)}_{\mu\nu}$ is massless due to general covariance, and $B^{(0)}_\mu$ is massless due to gauge invariance. Since the scale transformations are only global, $\phi^{(0)}$ is massless because it is the Goldstone boson associated with the spontaneous breakdown of scale invariance.

The full theory includes the infinite tower of KK states. Its symmetries were investigated by L. Dolan and M.J.Duff [132]. They expand the descriptors ξ^μ, ζ in Fourier components

$$\xi^\mu(x,\theta) = \sum_{n=-\infty}^{+\infty} (a_n^\mu + \omega^\mu_{\nu n} x^\nu) Y_n(\theta) \qquad \zeta^4(x,\theta) = \sum_{n=-\infty}^{+\infty} c_n Y_n(\theta).$$

The generators/charges of these symmetry transformations are

$$P_\mu^n = i e^{in\theta} \partial_\mu \qquad M_{\mu\nu}^n = i e^{in\theta}(x_\mu \partial_\nu - x_\nu \partial_\mu) \qquad Q^n = i e^{in\theta} \partial_\theta.$$

These form an infinitesimal algebra that generalizes the Poincaré algebra. It is a Kac-Moody algebra (see A.2.6.) characterizing the symmetry of the full four-dimensional action. However the vacuum

$$< g_{\mu\nu} >= \eta_{\mu\nu} \qquad < B_\mu >= 0 \qquad < \phi >= 1$$

is only invariant under $\mathbf{T_4} \times \mathbf{U(1)}$, so the full Kac-Moody symmetry is spontaneously broken. Thus one may expect Goldstone bosons, and in the full analysis (see e.g. [20]) one discovers that indeed the fields A_n^μ and ϕ_n for $n \neq 0$ become Goldstone bosons.

8.2.3 Beyond Five Dimensions: Einstein-Yang-Mills Theory

T. Kaluza aimed to unify the gravitational and electromagnetic interactions within a higher dimensional gravitational theory. These were the only known fundamental interactions known in his time. Whats about including the weak and strong interactions? As will be demonstrated next, the Klein-Kaluza approach can be extended beyond five dimensions and to symmetry groups other than $\mathbf{U(1)}$. This was pioneered by O. Klein in 1937, investigated for a "theory of meson-nucleon interactions" by W. Pauli in 1953 (see [486], [429], [430])[7] and posed as an exercise in the 1963 Les Houches lectures by B. DeWitt [120]. Surely it is an intriguing endeavor to "derive" Yang-Mills theory from higher-dimensional gravity. A first complete derivation yielding the coupling of four-dimensional gravitational field with an arbitrary Yang-Mills and scalar fields was given in 1975 by Y. M. Cho and P. G. O. Freund [92].

---

[7] Indeed, Pauli deduced in this dimensional reduction the transformation of an $\mathbf{SU(2)}$ vector field, later named after Yang and Mills.

(4 + *d*) Geometry

Playing with higher dimensions and mimicking the compactification of the 5D case, one may assume that the relevant D-dimensional manifold is the product of a space-time V_4 and a (compact space) B^d of dimension d. This is a rather stringent generalization from the five dimensional Kaluza-Klein model where $B^1 = S^1$. Write the metric in $D = 4 + d$ dimensions in analogy with (8.11a) in the form

$$\hat{g}_{AB} = \begin{pmatrix} g_{\mu\nu} + \phi_{ij} B^i_\mu B^j_\nu & \phi_{ij} B^i_\mu \\ \\ \phi_{ij} B^j_\nu & \phi_{ij} \end{pmatrix}. \tag{8.25}$$

Here the indices A, B are in the range $(0, \ldots, D-1)$. The indices $(\mu, \nu = 0, \ldots, 3)$ refer to the x^μ coordinates in M_4, and $(i, j = 4, \ldots D)$ refer to the y^i coordinates in B^d. Furthermore, $g_{\mu\nu}(x)$ is the metric tensor of M_4 and $\phi_{ij}(y)$ is the metric tensor of B^d. The assumption that the $g_{\mu\nu}$ do not depend on the additional coordinates y^i is a generalization of Kaluza's cylinder condition, and the restriction $\phi_{ij} = \phi_{ij}(y)$ mimics the positing of the scalar field to a constant in the 5D case. The inverse of (8.25) is

$$\hat{g}^{AB} = \begin{pmatrix} g^{\nu\lambda} & -B^k_\rho g^{\nu\rho} \\ \\ -B^j_\rho g^{\rho\lambda} & \phi^{jk} + B^j_\rho B^k_\sigma g^{\rho\sigma} \end{pmatrix},$$

where $g^{\nu\lambda}$ is the inverse of $g_{\mu\nu}$ and ϕ^{jk} is the inverse of ϕ_{ij}. Furthermore, $\det \hat{g} = \det g \times \det \phi$. The $(4+d)$ curvature scalar for this metric is calculated to be

$$\hat{R} = \hat{R}^{AB}\,\hat{g}_{AB} = R_4(x) + R_{B^d}(y) - \frac{1}{4}\phi_{ij}(y) B^i_{\mu\nu} B^j_{\rho\sigma} g^{\mu\rho}(x) g^{\nu\sigma}(x). \tag{8.26}$$

Here R_4 and R_{B^d} are the four- and d-dimensional curvature scalars, respectively, and $B^i_{\mu\nu}$ abbreviates

$$B^i_{\mu\nu}(x, y) := \partial_\mu B^i_\nu - B^j_\nu \partial_j B^i_\mu - (\mu \to \nu). \tag{8.27}$$

This is not the Yang-Mills field strength tensor, and we so far do not see how group structure constants could blossom from this obviously naive approach. But we may remember that O. Klein was led to the parametrization (8.11a) by symmetry arguments. Only in this specific form the B_μ transform as covariant vectors with respect to the Minkowski coordinates and as gauge potentials with respect to the coordinates in S^1. A similar line of reasoning must be pursued in case of more than one extra dimension. At first we seek gauge transformations, and start with transformations of the form

$$x' = x'(x) \qquad\qquad y' = y'(x, y),$$

for which one finds

$$\phi'_{ij} = \frac{\partial y^k}{\partial y'^i}\frac{\partial y^l}{\partial y'^j}\phi_{kl}, \quad B'^i_{\ \mu} = \frac{\partial x^\nu}{\partial x'^\mu}\left(\frac{\partial y'^i}{\partial y^k}B^k_\nu - \frac{\partial y'^i}{\partial x^\nu}\right), \quad g'_{\ \mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu}g_{\rho\sigma}. \tag{8.28}$$

Thus $g_{\mu\nu}$ in fact transforms as a V_4-tensor and as a B^d-scalar. Furthermore, ϕ_{ij} transforms as a B^d-tensor and as a V_4-scalar. In general, the B^i transform inhomogeneously. Only for transformations with $y' = y$ does each B^i transform as a covariant vector. Gauge transformations should be concealed in the extra dimensions. But if the transformations are restricted to $x' = x$, which means

$$B'^i_{\ \mu} = \frac{\partial y'^i}{\partial y^k}B^k_\mu - \frac{\partial y'^i}{\partial x^\mu}, \tag{8.29}$$

then the fields B^i_μ still do not exhibit a Yang-Mills field behavior. Thus, something is missing here again.

Isometries as Origin of Yang-Mills Fields

This missing link can be obtained from the observation that the field ϕ in the $5D$ model is a scalar both under M_4 and S^1 transformations. If we carry this over to the higher-dimensional case, we require $\bar{\delta}\phi_{ij} = 0$, and this restricts the admissible transformations in the space B^d. Actually this condition reveals that the infinitesimal descriptors ξ^i in $y'^i = y^i + \xi^i$ are (linear combinations of) Killing vectors k_a of B^d, that is

$$y'^i = y^i + \epsilon^a(x)k^i_a(y). \tag{8.30}$$

The isometry group in the manifold B^d defined by the Killing vectors k_a gives rise to the Yang-Mills theory part by defining

$$B^i_\mu(x, y) = \gamma k^i_a(y)A^a_\mu(x) \tag{8.31}$$

with a coupling constant γ (which as will be seen later can meaningfully be absorbed in the Killing vector fields). We may now determine the transformation of the vector fields A^a_μ with respect to the transformations (8.30) starting from (8.29):

$$\gamma k^i_a A'^a_{\ \mu} = (\delta^i_j + \epsilon^a\partial_j k^i_a)\gamma k^j_b A^b_\mu - k^i_a\partial_\mu\epsilon^a.$$

Since k_a are Killing vectors they obey a relation (E.22), that is $k^i_a\,\partial_i k^j_b - k^i_b\,\partial_i k^j_a = f_{abc}k^j_c$, and using this we finally arrive at

$$A'^a_{\ \mu} = A^a_\mu - \frac{1}{\gamma}\partial_\mu\epsilon^a + \frac{1}{\gamma}f^{bca}A^b_\mu\epsilon^c = A^a_\mu - \frac{1}{\gamma}D_\mu\epsilon^a$$

where f^{bca} are the structure constants of the isometry group. Thus A^a_μ has the transformation property of a gauge potential in a Yang-Mills theory which we were seeking. One further readily realizes that the "pseudo"-field strengths $B^i_{\mu\nu}$ from (8.27) are related to the Yang-Mills field strengths by $B^i_{\mu\nu}(x, y) = \gamma k^i_a(y) F^a_{\mu\nu}(x)$, where

$$F^a_{\mu\nu}(x) = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - \gamma A^b_\mu A^c_\nu f^{bca}.$$

By this the D-dimensional curvature (8.26) becomes

$$\hat{R} = R_4 + R_{B^d} - \frac{\gamma^2}{4}\phi_{ij} k^i_a k^j_b F^a_{\mu\nu} F^b_{\rho\sigma} g^{\mu\rho} g^{\nu\sigma}.$$

The D-dimensional Hilbert-Einstein action

$$S_{DHE} = \frac{1}{2\hat{\kappa}} \int d^4x d^d y \sqrt{-\hat{g}}\, \hat{R}$$

has three terms

$$\begin{aligned}
S_{DHE} =& \frac{1}{2\hat{\kappa}} \int d^4x \sqrt{-\det g}\, R_4 \int d^d y \sqrt{\det \phi} \\
&+ \frac{1}{2\hat{\kappa}} \int d^4x \sqrt{-\det g} \int d^d y \sqrt{\det \phi}\, R_{B^d} \\
&- \frac{1}{2\hat{\kappa}} \frac{\gamma^2}{4} \int d^4x \sqrt{-\det g} F^a_{\mu\nu} F^b_{\rho\sigma} g^{\mu\rho} g^{\nu\sigma} \int d^d y \sqrt{\det \phi}\, \phi_{ij} k^i_a k^j_b .
\end{aligned}$$

In order to recover from the first and third term the coupled Einstein-Yang-Mills theory in four dimensions we require

$$\hat{\kappa} = \kappa \int d^d y \sqrt{\det \phi} \tag{8.32a}$$

$$\delta_{ab} = \frac{\gamma^2}{2\hat{\kappa}} \int d^d y \sqrt{\det \phi}\, \phi_{ij} k^i_a k^j_b . \tag{8.32b}$$

Next we observe that the second term is completely determined by the isometry group, and we introduce the constant

$$\Lambda = -\frac{\int d^d y \sqrt{\det \phi}\, R_{B^d}}{2 \int d^d y \sqrt{\det \phi}} . \tag{8.33}$$

Equation (8.32a) relates the gravitational constants in $(D+4)$ and in four dimensions to the group volume, and (8.32b) is simply an (ortho)-normalization of the Killing vectors. Thus

$$S_{DHE} = \int d^4x \sqrt{-\det g} \left(\frac{1}{2\kappa}(R_4 - 2\Lambda) - \frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} \right)$$

is indeed the Einstein-Yang-Mills theory in four dimensions (together with a cosmological term) as derived from a vacuum Hilbert-Einstein theory in $4+D$ dimensions with a gauge group characterized by the structure constants of the isometry algebra of B^d. This is a direct generalization of the Abelian case where the gauge symmetry in four dimensions originates in the isometries of the circle in the fifth dimension. The mathematics behind this can be understood in terms of coset spaces (see Appendix E.4.4): the compactified space is a coset space $B^d = \mathbf{G}/\mathbf{H}$. Group transformations in $\mathbf{G}$ can be expressed as coordinate transformations in $\mathbf{G}/\mathbf{H}$. Group transformations leave the metric tensor in the compactified space invariant. One "obtains" a Yang-Mills theory based on the group $\mathbf{G}$, and "derives"[8] the theory from GR in $4{+}d$ dimensions where $d = \dim \mathbf{G}$ -dim $\mathbf{H}$. (In the literature, you may find this coset reduction also under the name of Pauli reduction in order to distinguish it from the DeWitt reduction in which the higher-dimensional theory is reduced on a compact group manifold $\mathbf{G}$. Here the isometry group is $\mathbf{G} \times \mathbf{G}$. A consistent interpretation is possible if only the gauge bosons of one copy of $\mathbf{G}$ are retained in the truncation to zero modes; for the different aspects of group and coset reductions see [106].)

Let us work out what this means in an example slightly more complex than the five-dimensional $V_4 \times S^1$ manifold. Replace S^1 by S^2, that is the surface of a sphere with radius r. Choose coordinates (θ, φ) for the sphere. Then $(\phi_{ij}) = \mathrm{diag}(r^2, r^2 \sin^2 \varphi)$ and therefore $\sqrt{\det \phi} = r^2 \sin \varphi$. Thus according to (8.32a)

$$\hat{\kappa} = \kappa \int d^2 y \sqrt{\det \phi} = \kappa r^2 \int d\theta d\varphi \sin \varphi = 4\pi r^2 \kappa.$$

The curvature scalar of S^2 is determined to be $R_{S^2} = -2/r^2$. It is a constant–as it must be since S^2 is maximally symmetric; see Appendix E.4.4. For the cosmological constant (8.33) we simply find $\Lambda = 1/r^2$. The Killing vectors of S^2 are determined as (E.23). Since these are not yet normalized, we write them

$$K_1 = \begin{pmatrix} \sin \varphi \\ \cos \varphi \cot \theta \end{pmatrix} \qquad K_2 = \begin{pmatrix} \cos \varphi \\ -\sin \varphi \cot \theta \end{pmatrix} \qquad K_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The associated generators $X_a = K_a^i \partial_i$ obey the $\mathfrak{so}(3)$-algebra $[X_a, X_b] = \epsilon_{abc} X_c$. Since

$$\int d\theta d\varphi \sqrt{\det \phi}\, \phi_{ij} K_a^i K_b^j = \frac{8\pi r^4}{3} \delta_{ab},$$

consistency with (8.32b) requires that $k_a = \gamma' K_a$ with $\gamma' = \sqrt{3\kappa}/(r\gamma)$. At the same time, the structure constants of the gauge algebra need to be rescaled as $\epsilon_{ijk} \to \gamma' \epsilon_{ijk}$. We see again that the coupling constant in the gauge theory is related to the typical dimension (diameter or circumference) of the internal space. Strictly speaking, everything derived so far holds only for the zero modes in a harmonic expansion of the fields

$$\mathcal{F}(x, y) = \sum_{l,m} \mathcal{F}_{lm}(x) Y_{lm}(\theta, \varphi)$$

[8] The use of the quotes to be clarified below.

where the Y_{lm} are the spherical harmonics, that is the eigenfunctions of the d'Alembertian $\Box_y$:

$$\Box_y Y_{lm} = \frac{l(l+1)}{r^2} Y_{lm} \qquad (l = 0, 1, \ldots\infty; m = -l, -l+1, \ldots, +l).$$

The resulting KK tower is characterized by masses $m_l^2 = l(l+1)/r^2$.

What would happen if one applies (8.32) and (8.33) to a torus $T^2 = S^1 \otimes S^1$ instead of a two-sphere? The line element of a torus is $ds^2 = r_1^2 d\psi^2 + r_2^2 d\varphi^2$ with two constant radii (r_1, r_2). From this we get $\sqrt{\det\phi} = r_1 r_2$ and $R_{T^2} = 0$ (indeed, a doughnut is flat!). There are only two independent Killing vectors: $K_1 = (1, 0)$, $K_2 = (0, 1)$. The two generators $X_a = K_a^i \partial_i$ commute–in fact they are generators for the two underlying $\mathbf{U(1)}$ groups. Due to

$$\int d\theta d\varphi \sqrt{\det\phi}\, \phi_{ij} K_a^i K_b^j = (2\pi\gamma^2) r_1 r_2\, \phi_{ij} K_a^i K_b^j$$

the K_a are orthogonal but each of them needs its own normalization ($k_a = \gamma_a' K_a$) in order to fulfill (8.32b). Here, we recover the two coupling constants of an $\mathbf{U(1)} \otimes \mathbf{U(1)}$ gauge theory.

These results can be generalized to d-dimensional spheres, for which the isometry group is $\mathbf{SO(d+1)}$, and obviously to torii as well. S. Weinberg [535] derived relations between the "circumference" of general B^d and the coupling constants of the associated Yang-Mills theory.

Although the previous findings sound rather promising and seem to suggest that the field equations of general relativity and Yang-Mills theories in four dimensions can be derived from a GR action in higher dimensions, this is not the full story. More care is needed to justify this claim. For instance there is still the trick with integrating out the coordinates in B^d in the form (8.32), a trick applicable only because of the assumption that ϕ_{ij} does not depend on the x-coordinates. It was mentioned that this is a generalization of the scalar field being constant in the $5D$ case. But we saw already that this assumption is not consistent with the full set of field equations in the sense that it enforces $F_{\mu\nu}F^{\mu\nu} = 0$. Thus it is no surprise that inconsistencies of this kind arise in the higher-dimensional case, also. And, as as will be seen, there are other inconsistencies as well.

Since the reduction of a higher-dimensional gravity theory to four dimensions became a widespread procedure in supergravity and for string models a lot of effort has been spent on "consistent dimensional reduction". Furthermore, there is the question of whether and how a higher-dimensional theory has an internal (dynamical) mechanism by which the reduction process takes place, this is called "spontaneous compactification".

Consistent Dimensional Reduction

Usually, when one speaks of Kaluza-Klein reduction, one has in mind a compactification together with a restriction to the zero modes. In this sense, the original

S^1 version is said to fulfill the criterion of consistent dimensional reduction, or consistent truncation. And it can be shown that consistency is guaranteed in torus compactifications $(S^1 \times \cdots \times S^1)$. The quest for consistency is to show that the lower dimensional field equations equations truncated to the zero modes provide a solution of the higher-dimensional equations.

But, as pointed out clearly in [142] (although W. Pauli already came to this point for the S^2 case), for any sphere reduction an inconsistency becomes visible in the gravitational field equations

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{1}{2} k_{ia} k^i_b \left[F^a_{\mu\rho} F^{b\ \rho}_{\nu} - \frac{1}{4} F^a_{\rho\sigma} F^{b\rho\sigma} g_{\mu\nu} \right].$$

The inconsistency is there because the left-hand side is a function of the coordinates x^μ, whereas the right hand side has $Y_{ab} := k_{ia} k^i_b$ depending on the coordinates y^n of the internal manifold. Off course the inconsistency would disappear if $Y_{ab} = \lambda \delta_{ab}$. But one can easily convince oneself that this is hard to achieve: The matrix Y is defined in terms of d-vectors, where d is the dimension of the internal space B^d. On the other hand the (a, b)-indices range over the dimension (dim $\mathbf{G}$) of the symmetry group. For the coset reduction $B^d = \mathbf{G}/\mathbf{H}$ it holds that $\dim \mathbf{G} > d$ and therefore Y has zero eigenvalues and cannot be proportional to the unit matrix. The cause of this inconsistency is the fact that the gauge boson fields act as sources not only for the massless graviton, but also for massive gravitons that one wants to truncate away.

This is quite general, and only in very exceptional cases a consistent reduction is possible at all; the reasons why this is the case are still not understood completely. Most of the exceptions do come from supergravity, the most famous one being the S^7 compactifications of $D = 11$ supergravity to four-dimensional SO(8) gauged N=8 supergravity [119]. (This theory is interesting also because it is an example of spontaneous compactification, see the next subsection.) However, supersymmetry is not a necessary ingredient. Thus for instance consistent truncations are known for the bosonic string and for brane-world scenarios.

Spontaneous Compactification

Even if a consistent truncation of a higher-dimensional theory is possible, there is still the question about the origin of the reduction: Are there mechanism inherent to a theory, why a specific compactification is preferred compared over another one? Why for instance should a compactification of a D-dimensional manifold as $M^D = V^4 \times B^{D-4}$ (with V^4 being a pseudo-Riemann space and B^{D-4} a compact space) be favored?

Let me start with a word of caution: Already in the $5D$ case, one certainly must ask why the extra dimension should compactify to a circle. But at least it is true that the ground state given by the metric

$$< \hat{g}_{AB} >= \begin{pmatrix} \eta_{\mu\nu} & 0 \\ 0 & 1 \end{pmatrix}$$

is a solution of the vacuum field equations. In fact the metrics of the Minkowski space M_5 and of $M_4 \times S^1$ are the same because the two manifolds are locally isomorphic. In higher dimensions, one meets the obstacle that the ground state metric in general is not a solution of the vacuum field equations. Or, expressed in another way, that four-space cannot be flat. This can be seen as follows: take the ground state metric to be

$$< \hat{g}_{AB} >= \begin{pmatrix} \eta_{\mu\nu} & 0 \\ 0 & < \phi_{ij} > \end{pmatrix}.$$

The Einstein field equations are (allowing for a cosmological constant $\hat{\Lambda}$)

$$\hat{R}_{AB} - \frac{1}{2} < \hat{g}_{AB} > (\hat{R} + \hat{\Lambda}) = 0. \tag{8.34}$$

Assume that 4-space is flat, then $\hat{R}_{\mu\nu} = R_{\mu\nu} = 0$ and therefore $\hat{R} + \hat{\Lambda} = 0$. But then $\hat{R}_{ij} = 0$ as well, which of course is not true if the internal space is curved. Here it is the case that M_D and $M_4 \times B^d$ are not locally isomorphic unless, of course B^d is flat. But then its isometry group would be trivially a product of $\mathbf{U(1)}$ groups. We saw this in the example of the 2-torus as an internal space.

Various compactification mechanisms have been investigated, amongst others the inclusion of further sources (conceivably also extra vector and gauge fields), a change of the geometry (e.g. introducing torsion) or modified actions (e.g. including terms quadratic in the curvature tensor).

Adding source terms to the theory means to derive the field equations from

$$S = \int d^D x \sqrt{|\det \hat{g}|}\, \frac{1}{2\hat{\kappa}} (\hat{R} - 2\hat{\Lambda}) + S_M \tag{8.35}$$

where S_M is some matter action part giving rise to the energy-momentum term $\hat{T}_{AB}$ defined by

$$\delta S_M = \frac{1}{2} \int d^D x \sqrt{|\det g|}\, \hat{T}_{AB} \delta \hat{g}^{AB}.$$

Now the field equations are–instead of (8.34) -

$$\hat{R}_{AB} - \frac{1}{2} \hat{g}_{AB} (\hat{R} + \hat{\Lambda}) = -\hat{\kappa} \hat{T}_{AB}. \tag{8.36}$$

If 4-space is flat, that is, if $\hat{R}_{\mu\nu} = 0 = \hat{R}_4$ we derive from the previous expression

$$\hat{T}_{\mu\nu} = \frac{C_4}{\hat{\kappa}} \eta_{\mu\nu} \qquad \text{with} \qquad C_4 = \frac{1}{2}(R_{B^d} + \hat{\Lambda}).$$

If the manifold B^d is an Einstein space

$$\hat{R}_{ij} = -\frac{K}{2} \phi_{ij}$$

with a constant K, related to the curvature scalar in B^d by $R_{B^d} = \frac{1}{2}d(1-d)K$, we derive from (8.36) that

$$\hat{T}_{ij} = \frac{c_d}{\hat{\kappa}}\phi_{ij} \qquad \text{with} \qquad c_d = \frac{1}{2}(R_{B_d} + \hat{\Lambda} + K).$$

Thus we may write the field equations for the "internal" components as

$$\hat{R}_{ij} - \phi_{ij} C_4 = -C_{B^d}\phi_{ij}$$

or $\hat{R}_{ij} = (C_4 - C_{B^d})\phi_{ij}$. For a consistent compactification of the manifold one must therefore require

$$(C_4 - C_{B^d}) < 0. \tag{8.37}$$

Since also $R_{B^d} = d(C_4 - C_{B^d})$, one derives a relation for the cosmological term:

$$\hat{\Lambda} = (2-d)C_4 + 2dC_{B^d}.$$

E. Cremmer and J. Scherk made the proposal to add further specific gauge fields to a KK-model in order to achieve "spontaneous compactification". They showed a mechanism for $B^d = S^d$. Their procedure–although somehow at odds with the original Kaluza idea–was elaborated by J. F. Luciani [351] for arbitrary "internal" spaces B^d. Luciani investigated which conditions must hold for the internal space if one couples a Yang-Mills theory with gauge group $\mathbf{G}$ to a Kaluza-Klein theory and looks for solutions of the field equations where the additional gauge fields obey

$$A^a_\mu = 0 \qquad A^a_i = \lambda k^a_i.$$

Here the k^a_i are the Killing vectors of B^d and λ is a constant. In the case that $\mathbf{G}$ is the same as the isometry group of B^d, there is a solution of the full set of field equations, if the internal metric is $\phi_{ij} = (1/2\lambda)k^a_i k^b_j g_{ab}$ with g_{ab} being the Cartan-Killing metric for the group $\mathbf{G}$. In this case the energy-momentum tensor on the right-hand side of (8.36) is the Belinfante tensor in D dimensions

$$\hat{T}_{AB} = F^a_{AC} F_{aB}{}^C - \frac{1}{4}\hat{g}_{AB} F^a_{CD} F^{CD}_a$$

built from the field-strength tensor belonging to the supplementary Yang-Mills theory. Now one derives

$$T_{\mu\nu} = -\frac{1}{4} g_{\mu\nu} F^a_{kl} F^{kl}_a \qquad T_{ij} = \frac{\lambda^2}{4}\phi_{ij} F^a_{kl} F^{kl}_a.$$

Relating this to the previous consideration we identify

$$C_4 = -\frac{\hat{\kappa}}{4} F^a_{kl} F^{kl}_a \qquad C_{B^d} = \frac{\lambda^2 \hat{\kappa}}{4} F^a_{kl} F^{kl}_a$$

and thus indeed (8.37) is fulfilled.

Another interesting procedure of spontaneous dimensional reduction goes under the name *Freund-Rubin compactification*, named after P.G.O. Freund and M.A. Rubin and developed in [197]. Inspired by the prominent role of an antisymmetric rank-3 tensor field in $11D$-supergravity (see Sect. 8.3.3), they investigated how tensor fields of arbitrary rank influence compactification. The main points are illustrated already on the example with a rank-1 tensor field, that is simply a vector field: Consider Einstein-Maxwell field theory in D dimensions

$$R_{AB} - \frac{1}{2} g_{AB} R = -\kappa T_{AB}, \qquad \frac{1}{\sqrt{|g|}} \partial_A \left(\sqrt{|g|} F^{AB}\right) = 0$$
$$T_{AB} = F_{AC} F_B{}^C - \frac{1}{4} g_{AB} F_{CD} F^{CD}.$$

Now look for solutions of these equations such that the D-dimensional spacetime becomes a product of a two-dimensional and a $(D-2)$-dimensional Riemann manifold: $V_D = V_2 \times V_{D-2}$, which means that

$$(g_{AB}) := \begin{pmatrix} g_{mn}(x^p) & 0 \\ 0 & g_{\bar{m}\bar{n}}(x^{\bar{p}}) \end{pmatrix}. \tag{8.38}$$

where (m, n, p) take on values in V_2 and $(\bar{m}, \bar{n}, \bar{p})$ take on values in V_{D-2}. The Maxwell field equations admit solutions

$$F^{AB} = \frac{\epsilon^{AB}}{\sqrt{|g_2|}} f$$

with a constant f, with $|g_2| = |\det(g_{mn})|$, and with

$$\epsilon^{AB} = \begin{cases} \epsilon^{mn} & \text{for A = m, ..., B = n,} \\ 0 & \text{otherwise.} \end{cases}$$

where ϵ^{mn} is the totally antisymmetric tensor in V_2. For this solution we now calculate the energy-momentum tensor. At first

$$F^2 := F_{CD} F^{CD} = g_{km} g_{jn} F^{kj} F^{mn} = \frac{f^2}{|g_2|} g_{km} g_{jn} \epsilon^{kj} \epsilon^{mn} = \frac{f^2}{|g_2|} 2 g_2 = 2 f^2 \operatorname{sgn}(g_2).$$

Furthermore,

$$K_{AB} := F_{AC} F_B{}^C = g_{Ai} g_{jk} g_{Bk} F^{ik} F^{kj}$$

so that $K_{A\bar{n}} = 0$ and

$$K_{mn} = \frac{f^2}{|g_2|} g_{mi} g_{jk} g_{nk} \epsilon^{ik} \epsilon^{kj} = \frac{f^2}{|g_2|} g_{mn}\, g_2 = f^2 \operatorname{sgn}(g_2)\, g_{mn}.$$

Since $T_{AB} = K_{AB} - 1/4 g_{AB} F^2$, we finally arrive at

$$T_{mn} = \frac{1}{2} f^2 \operatorname{sgn}(g_2)\, g_{mn} = -T_{\bar{m}\bar{n}}.$$

The condition (8.37), now read as $(C_2 - C_{D-2}) < 0$, is fulfilled if the time dimension is in V_2. Then V_{D-2} is compact. This holds also the other way: If the time dimension is in V_{D-2}, the space V_2 is compact. This can be seen by contracting the two field equations

$$\begin{aligned} R_{mn} - \frac{1}{2} g_{mn}(R_2 + R_{D-2}) &= -\hat{\kappa}\, T_{mn} = -\frac{\lambda}{2} g_{mn} \\ R_{\bar{m}\bar{n}} - \frac{1}{2} g_{\bar{m}\bar{n}}(R_2 + R_{D-2}) &= -\hat{\kappa}\, T_{\bar{m}\bar{n}} = +\frac{\lambda}{2} g_{\bar{m}\bar{n}} \end{aligned}$$

with g^{mn} and $g^{\bar{m}\bar{n}}$, and by solving the resulting algebraic equations for the curvature scalars:

$$R_{D-2} = \lambda \qquad R_2 = -2\frac{D-3}{D-2}\lambda \qquad \text{with} \qquad \lambda := \kappa f^2 \mathrm{sgn}(g_2). \tag{8.39}$$

Keep in mind that $D \geqslant 2$. For $D = 3$, the spacetime curvature vanishes and $R_1 = \lambda$. And for $D \geqslant 4$ the scalar curvatures do have opposite signs. From the previous expressions for the Ricci tensors, one infers that λ can be interpreted as a cosmological constant. If it is positive, it makes V_2 an anti-de Sitter space and V_{D-2} a de Sitter space.

Freund and Rubin generalized this line of reasoning to antisymmetric tensor field $A_{A_1 A_2 \ldots A_{s-1}}$ of rank $(s-1)$. The corresponding field strength built according to

$$F_{A_1 A_2 \ldots A_s} = \partial_{A_1} A_{A_2 A_3 \ldots A_s} - \partial_{A_2} A_{A_1 A_3 \ldots A_s} + \ldots$$

is then a completely antisymmetric tensor of rank s. Add a matter term proportional to

$$-\int d^D x \sqrt{|g|} F_{A_1 A_2 \ldots A_s} F^{A_1 A_2 \ldots A_s}$$

to the D-dimensional Hilbert-Einstein action and look for solutions of the field equations such that D-dimensional spacetime has the form of a product space $V_s \times V_{D-s}$:

$$(g_{AB}) := \begin{pmatrix} g_{mn}(x^p) & 0 \\ 0 & g_{\bar{m}\bar{n}}(x^{\bar{p}}) \end{pmatrix}$$

where (m, n, p) take on values in V_s and $(\bar{m}, \bar{n}, \bar{p})$ take on values in V_{D-s}. Quite similar to the case $s = 2$, the matter field equations admit solutions

$$F^{C_1 \ldots C_s} = \frac{\epsilon^{C_1 \ldots C_s}}{\sqrt{|g_s|}} f.$$

Finally the generalization of (8.39) is

$$R_{D-s} = \frac{(s-1)(D-s)}{D-2}\lambda \quad R_s = -s\frac{D-s-1}{D-2}\lambda \qquad \text{with} \qquad \lambda := \kappa f^2\, \mathrm{sgn}(g_s).$$

For $D > s + 1$ again either s or $(D - s)$ space-like dimensions become compact. If we want to deal with a $(4 + d)$-theory, obviously $s = 4$ requires the introduction of a completely antisymmetric tensor field A_{CBD}.

Without further restrictions, antisymmetric tensor fields of arbitrary rank $(1 \leqslant s \leqslant D)$ can be introduced, and the number of compactified dimensions is arbitrary as well. This changes in the case of supergravity, where the type of admissible fields is largely determined from the representations of the superalgebra. As it turns out, $D = 11$ supergravity needs fields with $s = 4$ and this leads to a preferential compactification of seven or four dimensions.

Standard Model in a Klein-Kaluza Context

One of the previous examples dealt with S^2 as the internal space with $\mathbf{SO(3)}$ as its isometry group. Now we ask with E. Witten [565] "What is the minimum dimension of a manifold which can have $\mathbf{SU(3)xSU(2)xU(1)}$ as a symmetry?" That is, we are looking for the manifold that has the group of the Standard Model as its isometry group. We know already that $\mathbf{U(1)}$ and $\mathbf{SO(3)} \cong \mathbf{SU(2)}$ are isometry groups for the circle S^1 and the surface of a sphere S^2, respectively. The manifold with isometry group $\mathbf{SU(3)}$ can be shown to be the complex projective space CP^2. This is defined as the space spanned by three complex coordinates z_i for which the point $\{z_1, z_2, z_3\}$ is identified with the point $\{\lambda z_1, \lambda z_2, \lambda z_3\}$ for non-zero complex λ. Thus one reasonable choice for the internal space is

$$B^7 = CP^2 \otimes S^2 \otimes S^1 .$$

This is not the only $7D$ space with $\mathbf{SU(3)xSU(2)xU(1)}$ as its isometry group, but no space with lower dimensionality has this symmetry. In order to recover the Standard Model we need a KK-model for which the dimension D is at least $4 + 7 = 11$. In [565] Witten showed that all seven-dimensional manifolds M^{pqr} that have the symmetry group of the Standard Model as isometry group can be classified and analyzed by three integer parameters (p, q, r). (In his notation $B^7 = M^{001}$.)

As will be explained in the next section, $D = 11$ is the maximal dimension in which supergravity can be formulated (provided massless fields of spin larger than two are excluded). Witten in [565]: "It is certainly a very intriguing numerical coincidence that eleven dimensions, which is the maximum number for supergravity, is the minimum number in which one can obtain $\mathbf{SU(3)xSU(2)xU(1)}$ gauge fields by the Klein-Kaluza procedure. This coincidence suggests that the approach is worth serious consideration."

Indeed, this looks promising. The Standard Model, however, is more than its gauge bosons. Can fermions be included? As explained in Sect. 6.3, the observed quarks and leptons are chiral fermions, which means that the left-handed and right-handed fermions transform under different representations of the $\mathbf{SU(2)} \times \mathbf{U(1)}$ gauge group. And this chirality property cannot be embedded in a KK-type theory which is compactified on a manifold B^7. This disappointing fact was first noticed by Witten [564] and the arguments were refined ever since. More details to be found in the review article [20].

8.2.4 Instead of a Conclusion

The history of KK-ideas has been rather volatile. After initially receiving some attention, interest in the Klein-Kaluza approach faded away. Einstein himself in his attempts at finding a unified field theory now and then came back to Kaluza's idea about a further dimension. He gave it up in the mid-1940's maybe because it did not meet his objective to find a non-singular particle solution; see e.g. [516]. Nevertheless the idea did not fall into oblivion, due to the influence of the relativity textbook by P. Bergmann [43] on at least a generation of physicists.

The idea of extra dimensions witnessed a renaissance in the late 1960's within the dual string theories, which started off with 26 (25+1) dimensions in the bosonic string model and 10 (9+1) dimensions for the fermionic strings. In current versions of the supersymmetric string model, the ideas of extra dimensions and compactification are essential characteristics of the theory. In the late 1970's the idea of compactified dimensions received another revival within supergravity [11]. Reading the more recent literature on superstrings and supergravity permits one to delve into an industry of compactification and dimensional reduction of theories in dimensions $D \leqslant 11$ to lower, mostly four, dimensions. As of today, all theories (or rather models) in which gauge fields are unified with gravitation by means of extra, compact dimensions are called KK-models.

Whereas according to Klein one thought about a radius of compactification of the order of the Planck length, more recently researchers within the string- and brane-model community (for a popular account see [433]) came to propose non-compact length scales even within the micrometer range. Thus effects of extra dimensions could be visible at the Large Hadron Collider LHC.

Although many intrinsic problems with the Klein-Kaluza theories (interpretation of scalar fields, cylinder condition and compactification, vacuum configurations and their stability, inclusion of chiral fermions–not to mention renormalizability issues) have still not been solved completely or satisfactorily, the original idea with extra dimensions receives high attractiveness in contemporary theoretical research. An arXiv search with the entry *Kaluza* in the subject area "Titles" results in more than 600 hits.

I should mention two other versions of Kaluza-Klein theories aside from the compactified variant described before and which is the most prevalent today. In the so called projective theories the extra dimensions are nothing but a technical gimmick, and in noncompactified theories the extra dimensions are considered real but not necessarily a simple length or time; see e.g. [400].

The KK approach has very attractive features. Kaluza himself concluded his article by "*Trotz voller Würdigung der geschilderten physikalischen wie auch der erkenntnistheoretischen Schwierigkeiten, die sich vor der hier entwickelten Auffassung auftürmen, will es einem schwer werden, zu glauben, daß in all jenen an formaler Einheitlichkeit kaum zu überbietenden Beziehungen immer nur ein launischer*

Zufall sein lockendes Spiel betreibt[9]". Another kind of conclusion is drawn in [20]: "it seems quite likely that even if the original pure Kaluza-Klein theory cannot be sustained, extra spatial dimensions will play an important role in the eventual unified theory of interactions, …"

8.3 Supersymmetry

Supersymmetry is a symmetry that relates elementary particles with their given spins to hypothetical particles that differ by half a unit of spin. Thus every fermion (boson) is accompanied by a bosonic (fermionic) superpartner. Supersymmetry is a hypothetical symmetry in which bosons and fermions are indistinguishable. Such a symmetry is not yet observed, since so far there is no evidence of supersymmetric partners of the elementary particles in the Standard Model. Thus, if it is fundamentally present at all, supersymmetry is definitely a broken symmetry.

Supersymmetry (or SuSy for short) became a research topic in fundamental theoretical physics in the 1970's, although there are traces in mathematics still earlier. As with "ordinary" symmetries, there is a global and a local version of supersymmetry. Global supersymmetric models of unification are considered to be promising candidates for extending the Standard Model of particle physics. Local supersymmetry entails Einstein's theory of general relativity.

We saw in the previous chapters, that symmetry groups are related to geometric transformations in appropriate spaces, generators of Lie group algebras are associated to charges derived from Noether currents, field types are derived from representations of the symmetry group, etc. All these findings can be carried over to supersymmetry if one extends these considerations to superspace and superfields. Since these concepts are not presentable in a short way, they are dealt with in Appendix B.3. In this section, the superfield language is not used.

8.3.1 Why Supersymmetry?

There are both esthetic and technical motives for considering supersymmetry.

- As yet, elementary particle physics exhibits a dichotomy in that in the Standard Model, the constituent fields and the fields mediating the interactions are fermionic and bosonic, respectively. Is it conceivable that this can be overcome by a new form of symmetry? If such a symmetry exists it would relate fermions to bosons: For every fermion discovered in nature there is a corresponding boson with–at least if the symmetry would be perfect–exactly the same properties (otherwise this could not be called a symmetry).

[9] Here a translation [430] that however does not convey the intriguing pathetic tone of our predecessors: "In spite of all the physical and theoretical difficulties which are encountered in the above proposal it is hard to believe that the derived relationships, which could hardly be surpassed at the formal level, represent nothing more than a malicious coincidence".

- So far we have distinguished coordinate symmetries and internal symmetries. Is there a way to overcome this dichotomy in a non-trivial manner? The answer is "Yes and No": A successful amalgamation of the rotation group with the isospin-group in terms of $\mathbf{SU(4)}$ is known in nuclear physics. And the generalization of this approach to the flavor group led to a $\mathbf{SU(6)}$ symmetry that had some successes in particle physics, for instance neatly explaining regularities in the mass spectrum of pseudoscalar and vector mesons. However, as elucidated before, the spacetime symmetry relevant for a relativistic quantum field is not the rotation group, but the Poincaré group. S. Coleman and J. Mandula proved a no-go theorem [97] stating that in three spatial and one time dimension under rather general conditions (locality, causality, positive energy, finite number of different particles) there is no other unification of the Poincaré group $\mathbf{P}$ and a compact internal symmetry group $\mathbf{G}$ aside from the trivial $\mathbf{S} = \mathbf{P} \otimes \mathbf{G}$. The essential point in the no-go can be seen in that the Lorentz group is non-compact as compared to the rotation group. The proof of the Coleman-Mandula theorem is technically demanding; it can for instance be found in an appendix of Chap. 24 in [536]. Here is a simple argument which shows the consequences of the trivial product structure of the Poincaré group and the internal group. Let $\mathbf{G}$ have generators X^a with $\left[X^a, X^b\right] = if^{abc}X^c$. In the case of the trivial product structure we have

$$\left[X^a, T_\mu\right] = 0 = \left[X^a, M_{\mu\nu}\right].$$

Therefore the generators X^a commute with the Casimir operators of the Poincaré group:

$$\left[X^a, T^2\right] = 0 = \left[X^a, W^2\right].$$

which reveals that all members of the internal symmetry group must have the same mass and spin.

Supersymmetry offers a possible escape from the no-go theorem. In the proof of the theorem it is assumed that symmetries are realized in terms of Lie algebras. As it turned out, a genuine amalgamation of Poincaré symmetry and an internal symmetry is possible in the presence of fermionic symmetries. How and under which conditions this becomes true was elaborated by R. Haag, J.T. Łopuszanski and M. Sohnius [251]. The amalgamation occurs by allowing for graded Lie algebras (supersymmetric Lie algebras or superalgebras) on the background of Grassmann numbers, entities that contain both commuting and anti-commuting objects; see Appendix B.2. Anti-commmutators are of course not the essential news of supersymmetry, since it is known since the 1920's that a system with half-integer spins must be quantized with anti-commutators. New is how anti-commutators are related to symmetries. It turned out that only few anti-commuting algebras are compatible with the principles of quantum field theory [251].

- Supersymmetric field theories have a tendency to better convergence properties than ordinary field theories, the reason being that the contributions of the problematic boson loops are canceled by contributions from the fermion loops. It seems like a miracle that there are parameters in the theory which do not obtain radiative corrections. There are even supersymmetric field theories which are finite to all orders of perturbation–a well-investigated example being N=4 super-Yang-Mills theory.

- Supersymmetry modifies the running coupling constants. By using in the Standard Model the measured values of the three coupling constants and extrapolating these via the renormalization group according to (6.41) one arrives at the graphical representation of Fig. 8.1a. As can be seen, the coupling constants do not exactly meet at the same point. Taking supersymmetry into account, the three coupling constants no longer follow (6.41) but the counting yields

$$b_3 = 9 - n_f \qquad b_2 = 6 - n_f + \frac{1}{2}n_H \qquad b_1 = -n_f - \frac{3}{10}n_H .$$

  The changes come about because of contributions of the supersymmetric partners of the SM fermions, gauge bosons and Higgs scalars. These expressions for the leading order beta-functions should strictly speaking be applied only in the range where supersymmetry is valid, perhaps in a range between $1 < \mu < 10^{13}$ TeV, say. In any case, as a consequence of these changes the coupling constants merge at a reasonable value of energy at approximately 10^{16} GeV, illustrated in Fig. 8.1b.
- While there are infinitely many ways how nature could choose among Standard Model gauge groups, the strong constraints brought about by the representation theory of the superalgebra reveal that (in four dimensions) only eight possible versions of supersymmetry could exist.
- The following point will become fully understandable only at he end of this chapter: You may have come to know[10] that an open problem within the concordance model[11] of cosmology is the "dark side of the universe". This is the problem to find mechanisms leading to dark matter and dark energy. One of the dark matter candidates is the LSP (Lightest Supersymmetric Particle); more about this at the end of Sect. 8.3.3.

8.3.2 Compelling Consequences of Fermi-Bose Symmetry

If supersymmetry is valid in nature, a first compelling consequence is an equal number of bosonic and fermionic degrees of freedom in one multiplet. This can be seen from a simple example: Consider the Hamiltonian

$$H = \omega_a a^\dagger a + \omega_b b^\dagger b$$

with a bosonic operator a and a fermionic operator b obeying

$$\left[a, a^\dagger\right] = 1, \qquad \{b, b^\dagger\} = 1$$

[10] If not, see any modern account of cosmology, such as e.g. [445].

[11] This is so to speak the other standard model, namely that of cosmology.

(all other brackets vanishing). Define the fermionic operator

$$Q := b^\dagger a + a^\dagger b,$$

for which

$$\left[Q, a^\dagger\right] = b^\dagger, \qquad \{Q, b^\dagger\} = a^\dagger.$$

If $a^\dagger|0\rangle$ represents a one-boson state, then $Qa^\dagger|0\rangle$ represents a one-fermion state, and *vice versa*. The operator Q obeys

$$[Q, H] = (\omega_a - \omega_b)Q.$$

Thus for $\omega_a = \omega = \omega_b$ the Hamiltonian is supersymmetric. Furthermore in this case

$$\{Q, Q^\dagger\} = \frac{2}{\omega}H,$$

revealing that in the case of equal energy for the bosonic and fermionic states, the Q, $Q^\dagger$ and H form a closed algebra. In a certain sense the supersymmetry generator can be interpreted as the square root of the Hamiltonian.

The basic idea of supersymmetry is to allow for transformations of a bosonic field B into a fermionic field F and *vice versa*, in its infinitesimal form generically described by

$$\delta B(x) = \epsilon F(x). \tag{8.40}$$

This relation has some insightful consequences

- Since B is even/commuting and F odd/anti-commuting in the sense of a Grassmann algebra, the quantity ϵ must be Grassmann-odd. In order to comply with the spin-statistics theorem, ϵ must have half-integer spin s. The simplest choice is to opt for ϵ as a $s = 1/2$ spinor.
- Since the mass-dimensions of B and F in four spacetime dimensions are $[B] = 1$ and $[F] = 3/2$, the mass dimension of ϵ necessarily is $[\epsilon] = -1/2$. This in turn shows that the transformation of the fermionic field cannot be of the form $\delta F(x) \sim \epsilon B(x)$. A "factor" with mass dimension $+1$ is needed to accommodate for the missing unit of dimension on the right-hand side. If one does not want to introduce dimensioned constants, the only choice is

$$\delta F(x) = \epsilon \partial B(x). \tag{8.41}$$

- In a conceivable case, B is a real scalar field, and F is a Majorana spinor. Then ϵ must also be a Majorana spinor. Thus in order to cope with the appropriate indices and to retain a completely covariant form, the previous relations (8.40) and (8.41) become explicitly

$$\delta B = \bar{\epsilon}^{\alpha} F_{\alpha} \tag{8.42a}$$

$$\delta F_{\alpha} = -i(\gamma^{\mu}\epsilon)_{\alpha}\partial_{\mu}B; \tag{8.42b}$$

the factor $(-i)$ being a convention.

- If these transformations are to represent a symmetry operation they rather ought to obey the group postulates, or in this infinitesimal form they must obey a closed algebra. We find for instance the commutator

$$[\delta_{\epsilon}, \delta_{\epsilon'}]\, B = 2i\bar{\epsilon}\gamma^{\mu}\epsilon'\partial_{\mu}B$$

because according to (B.36) $\bar{\epsilon}\gamma^{\mu}\epsilon' = -\bar{\epsilon}'\gamma^{\mu}\epsilon$. The commutator reveals that two successive supersymmetry transformations on the bosonic field B lead to space-time translations $\partial_{\mu}B$. For this reason global supersymmetry is sometimes said to be the "square root" of translations. Having in mind that gauging translations is linked to general relativity, we surmise that local/gauged supersymmetry ("supergravity") can be grasped as square root of gravity (see Sect. 8.3.4).

8.3.3 Global Supersymmetry

The Wess-Zumino Model

We start the investigation of supersymmetry with the simplest supersymmetric field theory in four dimensions, the Wess-Zumino model [546], at first the free massless model, and then also the massive and interacting model. The Wess-Zumino model describes the dynamics of a real scalar field S, a real pseudoscalar field P and a Majorana spinor ψ.

The Lagrangian for the free massless case is

$$\mathcal{L}_{kin} = \frac{1}{2}(\partial_{\mu}S\partial^{\mu}S) + \frac{1}{2}(\partial_{\mu}P\partial^{\mu}P) + \frac{1}{2}i\bar{\psi}\partial\!\!\!/\psi. \tag{8.43}$$

The field equations are simply

$$[\mathcal{L}_{kin}]^{S} = -\Box S = 0 \qquad [\mathcal{L}_{kin}]^{P} = -\Box P = 0 \qquad [\mathcal{L}_{kin}]^{\Psi} = \frac{1}{2}i\partial\!\!\!/\psi = 0.$$

Of course the action belonging to the Lagrangian (8.43) is invariant under Poincaré transformations. But it is also invariant under the transformations

$$\delta_{\epsilon}S = \bar{\epsilon}\psi \qquad \delta_{\epsilon}P = i\bar{\epsilon}\gamma^{5}\psi \tag{8.44a}$$

$$\delta_{\epsilon}\psi = -i\partial\!\!\!/(S + i\gamma^{5}P)\epsilon \tag{8.44b}$$

where ϵ is a constant Majorana spinor. Namely, under these transformations the Lagrangian (8.43) transforms as

$$\begin{aligned}
\delta_\epsilon \mathcal{L}_{kin} &= \partial^\mu S \partial_\mu(\delta_\epsilon S) + \partial^\mu P \partial_\mu(\delta_\epsilon P) + \frac{1}{2} i(\delta_\epsilon \bar{\psi}) \partial\!\!\!/ \psi + \frac{1}{2} i \bar{\psi} \partial\!\!\!/ (\delta_\epsilon \psi) \\
&= \partial_\mu S \partial^\mu(\bar{\epsilon}\psi) + \partial_\mu P \partial^\mu(i\bar{\epsilon}\gamma^5\psi) \\
&\quad + \frac{1}{2} i \left[i\bar{\epsilon}\gamma^\nu\gamma^\mu \partial_\nu (S + i\gamma^5 P) \right] \partial_\mu \psi + \frac{1}{2} i \bar{\psi} \partial\!\!\!/ \left[-i\gamma^\nu \partial_\nu (S + i\gamma^5 P)\epsilon \right] \\
&= \partial_\mu \left[\bar{\epsilon} \left\{ \partial^\mu (S + i\gamma^5 P) - \frac{1}{2}\gamma^\nu\gamma^\mu \partial_\nu (S + i\gamma^5 P) \right\} \psi \right] \\
&= \partial_\mu \left[\frac{1}{2} \bar{\epsilon}\gamma^\mu \left\{ \gamma^\nu \partial_\nu (S + P\gamma_5) \right\} \psi \right],
\end{aligned} \tag{8.45}$$

where the identities $\bar{\epsilon}\psi = \bar{\psi}\epsilon$ and $\bar{\epsilon}\gamma^5\psi = \bar{\psi}\gamma^5\epsilon$ according to (B.36) were used. The result (8.45) exhibits another feature of supersymmetry (which it shares with diffeomorphism invariant theories): Their Lagrangians are quasi-invariant: $\delta_\epsilon \mathcal{L} = \partial_\mu F^\mu$.

The transformations (8.44) are the prime examples of supersymmetry transformations, and a realization of (8.42) within a field theory. In order to discover the symmetry group behind these infinitesimal transformations one calculates the commutators of two transformations with parameters ϵ_1 and ϵ_2: At first

$$\delta_2(\delta_1 S) = \delta_2(\bar{\epsilon}_1\psi) = -i\bar{\epsilon}_1\gamma^\mu \partial_\mu (S + i\gamma^5 P)\epsilon_2,$$

from which

$$(\delta_2\delta_1 - \delta_1\delta_2)S = -i\bar{\epsilon}_1 \partial\!\!\!/ (S + i\gamma^5 P)\epsilon_2 + i\bar{\epsilon}_2 \partial\!\!\!/ (S + i\gamma^5 P)\epsilon_1 = -2i\bar{\epsilon}_1\gamma^\mu\epsilon_2\, \partial_\mu S,$$

where again use has been made of (B.36). Similarly, one derives $[\delta_2, \delta_1]\, P = -2i\bar{\epsilon}_1\gamma^\mu\epsilon_2\, \partial_\mu P$. Finally

$$\begin{aligned}
[\delta_2, \delta_1]\, \psi &= -i\partial\!\!\!/ \delta_2 (S + i\gamma^5 P)\epsilon_1 - (1 \leftrightarrow 2) \\
&= -i\partial\!\!\!/ (\bar{\epsilon}_2\psi\epsilon_1 + i\gamma^5\bar{\epsilon}_2 i\gamma^5\psi\epsilon_1) - (1 \leftrightarrow 2) \\
&= -2i\bar{\epsilon}_1\gamma^\mu\epsilon_2\partial_\mu\psi + i\bar{\epsilon}_1\gamma^\nu\epsilon_2\gamma_\nu \partial\!\!\!/ \psi \\
&= -2i\bar{\epsilon}_1\gamma^\mu\epsilon_2\partial_\mu\psi + 2\bar{\epsilon}_1\gamma^\nu\epsilon_2\gamma_\nu [\mathcal{L}_{kin}]^\psi .
\end{aligned} \tag{8.46}$$

In the last line, a Fierz-identity was used together with (B.36), e.g. for the term

$$(\bar{\epsilon}_2\psi)\epsilon_1 - (1 \leftrightarrow 2) = -\frac{1}{4}(\bar{\epsilon}_2\gamma_\mu\epsilon_1)\gamma^\mu\psi - \frac{1}{4}(\bar{\epsilon}_2\gamma_{\mu\nu}\epsilon_1)\gamma^{\mu\nu}\psi.$$

In (8.46) we observe a phenomenon typically arising in supersymmetry: The commutator algebra only closes on-shell, that is on the space of solutions of the field equations, although the field equations themselves are invariant (the Lagrangian being quasi-invariant) independent of whether the field equations do hold or not.

This is an example of a general variational symmetry with an "open" algebra, mentioned in Sect. 3.3.4.

The non-closure of the algebra can be cured by launching additional fields. These moreover restore the off-shell balance of bosonic and fermionic fields. So far there is no balance, since the Majorana spinor has four components, but there are only two bosonic fields. Thus one introduces two additional, so-called "auxiliary", bosonic fields F_1 and F_2. Their name "auxiliary" stems from the fact, that due to their trivial field equations $[\mathcal{L}]^{F_i} = F_i = 0$ they are not dynamic and do not harm the dynamics of the other fields. The trivial field equations can be obtained from having terms proportional to F_i^2 in the Lagrangian; thus we consider

$$S^0_{WZ} = \int d^4x \big\{ \mathcal{L}_{kin} + \frac{1}{2}(F_1^2 + F_2^2) \big\}. \tag{8.47}$$

The extra fields ought to have mass dimension $[F_i] = 2$. On dimensional grounds the supersymmetry transformations of these fields are

$$\delta F_1 = -i\bar{\epsilon}\partial\!\!\!/\psi \qquad \delta F_2 = \bar{\epsilon}\gamma^5\partial\!\!\!/\psi, \tag{8.48}$$

and we assume that F_1 is a scalar and F_2 a pseudoscalar. The fields F_i cannot modify the transformations of S and P (as in (8.44a)), but they show up in the transformation of the Majorana field as

$$\delta_\epsilon \psi = -i\partial\!\!\!/(S + i\gamma^5 P)\epsilon + (F_1 + i\gamma^5 F_2)\epsilon. \tag{8.49}$$

The numerical coefficients in front of the F_i-terms are found from demanding that the commutator of two supersymmetry transformations (on all fields involved) closes off-shell:

$$[\delta_2, \delta_1]\, X = -2i\bar{\epsilon_1}\gamma^\mu\epsilon_2\partial_\mu X \qquad \text{for} \qquad X = \{S, P, \psi, F_i\}. \tag{8.50}$$

Before explaining the supersymmetry algebra in more detail in the next subsection, let us add mass and interaction terms to the theory so far defined by (8.47). Adding a mass term

$$\mathcal{L}_m = -\frac{1}{2}m_S^2 S^2 - \frac{1}{2}m_P^2 P^2 - \frac{1}{2}m_\psi\bar{\psi}\psi \tag{8.51}$$

to the Lagrangian, one shows that the action $\int d^4x(\mathcal{L}_{kin} + \mathcal{L}_m)$ is invariant under the first two transformations in (8.44) and a modified transformation

$$\tilde{\delta}_\epsilon \psi = -i(\partial\!\!\!/ - m)(S + i\gamma^5 P)\epsilon$$

only if $m_S = m_P = m_\psi = m$. This result, that all fields have the same mass, is typical of supersymmetry. It will later be interpreted in the sense that irreducible representations of the Poincaré superalgebra are mass-degenerate.

Wess and Zumino showed that one can add to the free theory $\mathcal{L}_{kin} + \mathcal{L}_m$ interaction terms in a way that the Poincaré supersymmetry is preserved, namely by an interaction term

$$\mathcal{L}_{int} = -g\Big\{\bar{\psi}(S - i\gamma^5 P)\psi + \frac{1}{2}g(S^2 + P^2)^2 + mS(S^2 + P^2)\Big\}$$

with some coupling g. The complete Wess-Zumino model is defined by the action

$$S_{WZ} = \int d^4x(\mathcal{L}_{kin} + \mathcal{L}_m + \mathcal{L}_{int}). \tag{8.52}$$

This action is invariant under supersymmetry transformations where the transformation of ψ is modified to

$$\tilde{\tilde{\delta}}_\epsilon S = \bar{\epsilon}\psi \qquad \tilde{\tilde{\delta}}_\epsilon P = i\bar{\epsilon}\gamma^5\psi \tag{8.53a}$$

$$\tilde{\tilde{\delta}}_\epsilon \psi = -i\Big[(\partial\!\!\!/ - m) - g(S + i\gamma^5 P)\Big](S + i\gamma^5 P)\epsilon. \tag{8.53b}$$

However, again with a lack of balance between the number of bosonic and fermionic degrees of freedom, the commutator of two supersymmetry transformations contains a term which is proportional to the field equation for ψ. What is the fully supersymmetric form of (8.52)? That is, in which form do the auxiliary fields F_i enter the Wess-Zumino action? The answer is

$$\begin{aligned}\mathcal{L}^a_{WZ} = &-\frac{1}{2}(\partial_\mu S\partial^\mu S) - \frac{1}{2}(\partial_\mu P\partial^\mu P) - \frac{1}{2}\bar{\psi}\partial\!\!\!/\psi + \frac{1}{2}(F_1^2 + F_2^2)\\ &+ m(F_1 S + F_2 P - \frac{1}{2}\bar{\psi}\psi) + g\Big[F_1(S^2 - P^2) + 2F_2 SP - \bar{\psi}(S - i\gamma^5 P)\psi\Big].\end{aligned} \tag{8.54}$$

I assume that you would not simply guess that this Lagrangian leads to the same equations of motion as (8.52). This comes about because of the field equations for the auxiliary fields are

$$F_1 = -mS - g(S^2 - P^2) \qquad F_2 = -mP - 2\,gSP. \tag{8.55}$$

Notice that supersymmetry enforces that there are only two dimensional parameter in the theory: the mass m and the coupling constant g which according to the coupling term cubic in the fields in the interaction part of the Lagrangian (8.54) can be interpreted as a Yukawa coupling.

Supersymmetric Yang-Mills Theory

Gauge theories can also be rendered supersymmetric. This was already exemplified in the classic article [546] for the case of electrodynamics. For a Yang-Mills theory with gauge group $\mathbf{G}$ the action

$$S = \int d^4x\left(-\frac{1}{4}F_a^{\mu\nu}F^a_{\mu\nu} + \frac{i}{2}\bar{\psi}^a(D\!\!\!\!/\,\psi)^a + \frac{1}{2}(D^a)^2\right)$$

is not only invariant under Poincaré transformations and **G** transformations but also under global supersymmetry transformations

$$\delta_\epsilon A^a_\mu = \bar{\epsilon}\gamma^\mu\gamma^5\psi^a \tag{8.56a}$$

$$\delta_\epsilon\psi^a = \frac{1}{2}\gamma^{\mu\nu}F^a_{\mu\nu}\gamma^5\epsilon - D^a\epsilon \qquad \gamma^{\mu\nu} = \frac{i}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu) \tag{8.56b}$$

$$\delta_\epsilon D = i\bar{\epsilon}(\not{D}\psi)^a. \tag{8.56c}$$

Here A^a_μ represent spin-1 gauge bosons of **G** and the ψ^a are their superpartners (*gauginos*). The D^a are auxiliary fields. It is instructive to count the number and type of fields. For a d-dimensional group there are $(3 \times d)$ components of the gauge fields (because for each of the A^a one component can be eliminated by a gauge choice) and d auxiliary fields. These $(3 \times d + d)$ bosonic fields are adjoined by the four components of the d Majorana spinors ψ^a. For all fields X present in the theory one can derive again the commutator relations (8.50). This points to a generic property of supersymmetry transformations which will next be exploited as the algebra of supercharges.

Super Poincaré Algebra

We know from Noether's first theorem that a global symmetry entails a conserved current and a Noether charge (under certain conditions on the current at infinity). This charge in turn is the generator of a symmetry transformation, and all charges constitute the Lie algebra belonging to the symmetry group. This was proven in Sect. 3.3. for "ordinary" symmetries. All this essentially holds also in the case of supersymmetry.

Thus we expect that the (super)transformations (8.44) are generated by spinorial (super)charges Q_a in the sense that for an infinitesimal Majorana spinor with components $\bar{\epsilon}^a$ we can write for any field Φ

$$\delta_\epsilon\Phi = \bar{\epsilon}^a Q_a\Phi.$$

For example in the Wess-Zumino model we read from (8.44) $\delta_\epsilon S = \bar{\epsilon}^a Q_a S = \bar{\epsilon}^a\psi_a$ or $Q_a S = \psi_a$. Now consider

$$[\delta_2, \delta_1]\,\Phi = [\bar{\epsilon}_2 Q, \bar{\epsilon}_1 Q]\,\Phi = \left[\bar{Q}\epsilon_2, \bar{\epsilon}_1 Q\right]\Phi = \left(\bar{Q}_a\epsilon^a_2\bar{\epsilon}^b_1 Q_b - \bar{\epsilon}^b_1 Q_b\bar{Q}_a\epsilon^a_2\right)\Phi$$
$$= -\bar{\epsilon}^b_1\epsilon^a_2\{Q_b, \bar{Q}_a\}\Phi.$$

Comparing this with (8.50), written in components as

$$[\delta_2, \delta_1]\,X = -2i\bar{\epsilon}^b_1(\gamma^\mu)_{ba}\epsilon^a_2\partial_\mu X$$

with the translation generator $T_\mu = i\partial_\mu$, we get the anti-commutation relation for the supercharges

$$\{Q_a, \bar{Q}_b\} = 2(\gamma^\mu)_{ab}T_\mu. \tag{8.57}$$

Thus the anti-commutator of two supersymmetry transformations results in a translation. Since the supercharge is a spinor it transforms as

$$\left[Q_a, M_{\mu\nu}\right] = \frac{1}{2}(\gamma_{\mu\nu})_a^{\ b} Q_b. \tag{8.58}$$

under Lorentz transformations. We saw previously on a simple example that the Hamiltonian $H = iT_0$ is supersymmetric if $[Q, H] = 0$. Thus covariance requests

$$\left[Q_a, T_\mu\right] = 0. \tag{8.59}$$

The relations (8.57–8.59) together with the Poincaré algebra formed by translations T_μ and Lorentz transformations $M_{\mu\nu}$ constitute the *super-Poincaré algebra*.

The supersymmetry algebra is invariant under $\mathbf{U(1)}$ transformations

$$Q_a \to e^{i\varphi} Q_a \qquad \text{or} \qquad Q_a \to e^{-iR\varphi} Q_a e^{iR\varphi}.$$

In comparing these expressions, we find that the generator of this unitary symmetry, termed *R-symmetry*, obeys the commutator relations

$$[Q_a, R] = Q_a \qquad\qquad \left[\bar{Q}_a, R\right] = -\bar{Q}_a.$$

The R-symmetry must not necessarily be a symmetry of the action; specifically the generator R does not commute with supersymmetry. R-symmetry plays thus a role in the breaking of supersymmetry.

So far, supersymmetry was exposed in terms Majorana spinors. However, it turned out in the development of research in supersymmetry that the use of Weyl spinors is much more appropriate. As described in Sect. 6.3, because of parity violation nature also seems to prefer 2-component objects. Then instead of charges Q_a with four components we have the two-component objects Q_α and $\bar{Q}_{\dot{\alpha}}$. The part of the supersymmetry algebra involving the fermionic charges is

$$\{Q_\alpha, \bar{Q}^{\dot{\alpha}}\} = 2(\sigma^\mu)^{\dot{\alpha}}_{\ \alpha} T_\mu \qquad \{Q_\alpha, Q_\beta\} = 0 \qquad \{\bar{Q}^{\dot{\alpha}}, \bar{Q}^{\dot{\beta}}\} = 0 \tag{8.60a}$$

$$\left[M_{\mu\nu}, Q_\alpha\right] = -\frac{1}{2}(\sigma_{\mu\nu})_\alpha^{\ \beta} Q_\beta \quad \left[M_{\mu\nu}, \bar{Q}^{\dot{\alpha}}\right] = \frac{1}{2}(\bar{\sigma}_{\mu\nu})^{\dot{\alpha}}_{\ \dot{\beta}} \bar{Q}^{\dot{\beta}} \tag{8.60b}$$

$$\left[T_\mu, Q_\alpha\right] = 0 \qquad\qquad \left[T_\mu, \bar{Q}^{\dot{\alpha}}\right] = 0. \tag{8.60c}$$

This algebra can of course be calculated directly from (8.57)–(8.59) but it is comprehensible from the transformation properties of the generators involved: The fact that the $(Q_\alpha, \bar{Q}^{\dot{\alpha}})$ are independent of spacetime coordinates entails (8.60c). The fact that Q_α transforms as a Weyl spinor means that $\left[M_{\mu\nu}, Q_\alpha\right] \propto (\sigma_{\mu\nu})_\alpha^{\ \beta} Q_\beta$ and a similar relation for $Q^{\dot{\alpha}}$ –and this is none other but (8.60b). For the anticommutator $\{Q_\alpha, \bar{Q}^{\dot{\alpha}}\}$ of two supercharges, we need an object antisymmetric in the indices $(\alpha, \dot{\alpha})$, and the only object that carries these indices is $(\sigma^\mu)^{\dot{\alpha}}_{\ \alpha}$. Now there is an extra Lorentz index which has to be contracted away, and the only Lorentz vector available is T_μ. Thus $\{Q_\alpha, \bar{Q}^{\dot{\alpha}}\} \propto (\sigma^\mu)^{\dot{\alpha}}_{\ \alpha} T_\mu$, which is the first anticommutator in (8.60). For the anticommutator $\{Q_\alpha, Q^\beta\}$, by a similar line of reasoning the most general possible expression is $\{Q_\alpha, Q^\beta\} = c_1(\sigma^{\mu\nu})_\alpha^{\ \beta} M_{\mu\nu} + c_2\delta_\alpha^\beta$. In taking the commutator of this expression with T_λ one finds that the constant c_1 must vanish. Furthermore, since $Q_\gamma := \epsilon_{\gamma\beta}Q^\beta$, the relation $\{Q_\alpha, Q_\gamma\} = c_2\epsilon_{\gamma\alpha}$ enforces $c_2 = 0$ since the left-hand side is symmetric and the right-hand side antisymmetric. Thus $\{Q_\alpha, Q_\beta\} = 0$, and similarly $\{\bar{Q}^{\dot{\alpha}}, \bar{Q}\dot{\beta}\} = 0$, which complete (8.60). The constants of proportionality in this algebra are fixed by the Jacobi identities up to the normalization of the supercharges.

Algebra of N-extended Supersymmetry

The Wess-Zumino algebra (8.60) can be extended, first by taking into consideration N types of fermionic charges/generators Q^i_α, $\bar{Q}^j_{\dot{\alpha}}$ where $i = 1, \ldots N$, and secondly by allowing for the possibility of further internal symmetries.

The first aspect was investigated by Haag, Łopuszanski, and Sohnius [251]. Their findings amount to the algebra

$$\{Q^i_\alpha, \bar{Q}^{\dot{\alpha}}_j\} = 2\delta^i_j(\sigma^\mu)_\alpha^{\ \dot{\alpha}} T_\mu \quad \{Q^i_\alpha, Q^j_\beta\} = \epsilon_{\alpha\beta}Z^{ij} \quad \{\bar{Q}^i_{\dot{\alpha}}, \bar{Q}^j_{\dot{\beta}}\} = \epsilon_{\dot{\alpha}\dot{\beta}}Z^{*ij} \tag{8.61a}$$

$$\left[M_{\mu\nu}, Q^i_\alpha\right] = -\frac{1}{2}(\sigma_{\mu\nu})_\alpha^{\ \beta} Q^i_\beta \qquad \left[M_{\mu\nu}, \bar{Q}^{\dot{\alpha}}_i\right] = \frac{1}{2}(\bar{\sigma}_{\mu\nu})^{\dot{\alpha}}_{\ \dot{\beta}} \bar{Q}^{\dot{\beta}}_i \tag{8.61b}$$

$$\left[T_\mu, Q^i_\alpha\right] = 0 \qquad \left[T_\mu, \bar{Q}^{\dot{\alpha}}_i\right] = 0 \tag{8.61c}$$

(together of course with the Poincaré algebra). The Z^{ij} are central charges–by definition commuting with all other symmetry generators–which, in contrast to a possible appearance in the Poincaré algebra, in general cannot be transformed away by redefining generators. They do have a physical meaning as quantum operators. For $N = 1$ there are of course no central charges, for $N = 2$ there is at most one, namely $Z^{12} = -Z^{21}$.

In the absence of central charges the N-extended supersymmetry algebra (8.61) is invariant under an automorphism group $\mathbf{U(N)}$ (reduced to $\mathbf{SU(N)}$ for the maximal theories) of internal symmetries.

$$Q^i_\alpha \to U^{ij} Q^j_\alpha$$

where U is a $N \times N$ unitary matrix. This is again an "R-symmetry".

A further extension of simple supersymmetry is possible by means of additional internal symmetries generated by charges X^a of a bosonic Abelian or non-Abelian algebra. These–as we know–commute with the Poincaré generators. By using the Jacobi identities their commutators with the supercharges are found to be

$$\left[X_a, Q^i_\alpha\right] = -(R_a)^i_{\ j} Q^i_\alpha \qquad \left[X_a, \bar{Q}^{\dot{\alpha}}_i\right] = \bar{Q}^{\dot{\alpha}}_j (\bar{R}_a)^j_{\ i} \tag{8.62}$$

where the $(R_a)^i{}_j$ are the representation matrices for the Lie-algebra generators in the representation according to the fermionic generators.

Summarizing, the N-extended super-Poincaré algebra with internal symmetries consists of

- bosonic generators T_μ and $M_{\mu\nu}$ of the Poincaré group
- bosonic Hermitean generators X^a ($a = 1, \ldots, \dim(\mathbf{G})$) of some Lie group $\mathbf{G}$
- fermionic generators Q^i_α (Weyl spinors with two components ($\alpha = 1, 2$) and $i = 1, \ldots, N$, belonging to a N-dimensional representation of $\mathbf{G}$, together with their conjugates $\bar{Q}^{\dot{\alpha}}_i$
- bosonic central charges Z^{ij}.

The part of the algebra involving the supercharges is given by (8.61) and (8.62).

Basic Consequences from the SuSy Algebra

The following conclusions can be derived immediately from the SuSy algebra (8.60):

- All particles belonging to an irreducible representation of the SuSy algebra have the same mass.
  This is simply a consequence of the commutators (8.61c), by which $\left[T^2, Q_\alpha\right] = 0$
- In a supersymmetric theory the energy is always non-negative.
  The contraction of $\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} = 2(\sigma^\mu)_{\alpha\dot{\alpha}} T_\mu$ with $(\bar{\sigma}^\nu)^{\dot{\alpha}\alpha}$, due to (B.20a), results in

  $$4T^\nu = (\bar{\sigma}^\nu)^{\dot{\alpha}\alpha}\{Q_\alpha, \bar{Q}_{\dot{\alpha}}\}.$$

  In taking the zero component and making use of the fact that the matrix $(\bar{\sigma}^0)$ is the unit matrix one further derives

  $$4T^0 = 4H = \sum_\alpha \{Q_\alpha, \bar{Q}_{\dot{\alpha}}\} = \sum_\alpha \{Q_\alpha, Q^\dagger_\alpha\} = \sum_\alpha (Q_\alpha Q^\dagger_\alpha + Q^\dagger_\alpha Q_\alpha)$$

  and thus for an arbitrary state $|\sigma\rangle$

  $$\langle\sigma|H|\sigma\rangle = \frac{1}{2}\sum_\alpha \sum_{\sigma'} \left|\langle\sigma'|Q_\alpha|\sigma\rangle\right|^2 \geqslant 0 \tag{8.63}$$

  because of the positivity of the Hilbert space.
- A supermultiplet contains an equal number of bosonic and fermionic states.
  Introduce the fermion number N_F as being 1 on fermionic states and 0 on bosonic states. Thus as an operator $(-1)^{N_F} B = B, (-1)^{N_F} F = -F$. Therefore, $(-1)^{N_F} Q_\alpha = Q_\alpha (-1)^{N_F - 1}$. The trace of $(-1)^{N_F}$ in an arbitrary (finite-dimensional) representation is

  $$\mathrm{Tr}\left[(-1)^{N_F}\right] = \sum_B \langle B \mid (-1)^{N_F} \mid B\rangle + \sum_F \langle F \mid (-1)^{N_F} \mid F\rangle = n_B - n_F$$

which counts the difference of bosonic and fermionic degrees of freedom. Due to the cyclic property of the trace

$$\begin{aligned}0 &= \mathrm{Tr}\left[-Q_\alpha(-1)^{N_F}\bar{Q}_{\dot\beta} + (-1)^{N_F}\bar{Q}_{\dot\beta}Q_\alpha\right]\\ &= \mathrm{Tr}\left[(-1)^{N_F}\{Q_\alpha,\bar{Q}_{\dot\beta}\}\right] = 2\sigma^\mu_{\alpha\dot\beta}\mathrm{Tr}\left[(-1)^{N_F}T_\mu\right].\end{aligned}$$

This must hold for arbitrary momentum, and thus $n_B = n_F$.

- The supercharges Q_1^i and $Q^{i\dot{2}}$ raise the z-component of the spin (resp. the helicity) whereas Q_2^i and $Q^{i\dot{1}}$ lower it by half a unit.
  This follows from (8.61b) using $M_{12} = J_3$: $\left[J_3, Q_1^i\right] = \frac{1}{2}Q_1^i$ and $\left[J_3, Q_2^i\right] = -\frac{1}{2}Q_2^i$, etc.

Representations of Super-Poincaré Transformations

In Sect. 5.1.4, we dealt with the classification of particles within relativistic field theories by the Wigner method of "little groups" or the induced representations: Particle states by definition transform in unitary irreducible representations of the (3+1)-dimensional Poincaré algebra. We saw that finding the irreducible representations of a group is eased by first discovering its Casimir operators. For the Poincaré group these were the square of the momentum operator, i.e. $C_1 = P^2$ and the square of the Pauli-Łubanski vector i.e. $C_2 = W^2$. The operator P^2 relates to the mass of the particle, and the little group depends on the mass being zero or not. For massive particles the little group is $\mathbf{SO(3)}$. The representations of $\mathbf{SO(3)}$ are labelled by the spin $j \in {}^1\!/_2\,\mathbb{Z}$ of the $(2j+1)$-dimensional representation. The little group for massless particles is $\mathbf{ISO(2)}$ and the eigenstates are labelled by a single eigenvalue, the helicity.

Since the SuSy algebra contains the Poincaré algebra as a subalgebra, any representation of a SuSy algebra is also a representation of the Poincaré algebra, although in general not an irreducible one. Each irreducible representation of the SuSy algebra corresponds to several particles. These are related by supersymmetry transformations, in the sense that they can be transformed into each other by the Q^i_α and $\bar{Q}^{\dot\alpha}_j$, and thus have spins differing by units of one half. These collections of supersymmetry-related particles are called supermultiplets.

We already made use above from the fact that due to $\left[P_\mu, Q_a\right] = 0$ the object P^2 is also a Casimir operator of the super-Poincaré algebra. This has the immediate consequence that the field components (or states) in a supermultiplet have the same mass.

However, the Pauli-Łubanski vector $W_\mu := \frac{1}{2}\epsilon_{\mu\nu\varrho\sigma}M^{\nu\varrho}P^\sigma$ cannot serve to define another Casimir operator since $\left[W^2, Q_a\right]$ has terms proportional to the supercharges. It is possible to generalize the Pauli-Łubanski vector (see the original publication [457]) to

$$V_\mu = W_\mu - \frac{1}{8}\bar{Q}\gamma_\mu\gamma^5 Q$$

which obeys $\left[V_\mu, Q_a\right] = -\frac{i}{2}(\gamma^5 Q)_a P_\mu$. Now the tensor $C_{\mu\nu} := V_\mu P_\nu - V_\mu P_\nu$ commutes with all supercharges and $C^2 = C_{\mu\nu}C^{\mu\nu}$ is another Casimir operator. As a matter of fact it has a similar meaning as the W^2 operator. For example in the frame $P^\mu = (m, \vec{0})$.

$$C^2 \mid m, \sigma, \ldots\rangle = -m^2\sigma(\sigma+1) \mid m, \sigma, \ldots\rangle$$

with a "superspin" σ.

As for the Poincaré group, we need to distinguish the massless and the massive case. But additionally we may consider $N > 1$ (extended supersymmetries) and for these in turn the cases with vanishing and non-vanishing central charges.

Massless States

For massless particles $(P^2 = 0)$, we boost to the frame in which the four-momentum is $(P_\mu) = (E, 0, 0, E)$. Let us assume that there are no central charges (as a matter of fact, this is not a restriction, since one can show that for the massless case the central charges must vanish because of the positivity of the Hilbert space). The anticommutators in the supersymmetry algebra (8.61a) become

$$\{Q^i_1, \bar{Q}^{\dot{1}}_j\} = 2\delta^i_j E$$
$$\{Q^i_2, \bar{Q}^{\dot{2}}_j\} = 0.$$

The latter is consistent with $Q^i_2 = 0 = \bar{Q}^{\dot{2}}_i$ for all representations, since for a positive definite Hilbert space

$$0 = \langle\Phi|\{Q^i_2, \bar{Q}^{\dot{2}}_i\}|\Phi\rangle = \| Q^i_2 \|^2 + \| \bar{Q}^{\dot{2}}_i \|^2 .$$

Thus we are left with only the Q^i_1 and $\bar{Q}^{\dot{1}}_j$. In defining

$$a_i := \frac{1}{\sqrt{4E}} Q^i_1, \qquad a_i^\dagger := \frac{1}{\sqrt{4E}} \bar{Q}^i_{\dot{1}}$$

which obey

$$\{a_i, a_j^\dagger\} = \delta_{ij} \qquad \{a_i, a_j\} = 0 = \{a_i^\dagger, a_j^\dagger\},$$

we see the a_i and $a_j^\dagger$ act as annihilation and creation operators, respectively. This leads us to introduce a lowest helicity state $|\lambda_0\rangle$ with helicity λ_0 in the sense that $a_i|\lambda_0\rangle = 0$ for $\forall i$. Each application of one of the operators $a_j^\dagger$ raises the helicity by $^1\!/_2$. Thus one obtains successively the states

$$
\begin{array}{ll}
|\lambda_0\rangle & \\
a_j^\dagger|\lambda_0\rangle = |\lambda_0 + \frac{1}{2}, j\rangle & N \text{ states} \\
a_{j_1}^\dagger a_{j_2}^\dagger|\lambda_0\rangle = |\lambda_0 + 1, j_1 j_2\rangle & \frac{1}{2}N(N-1) \text{ states} \\
a_{j_1}^\dagger \cdots a_{j_k}^\dagger|\lambda_0\rangle = |\lambda_0 + \frac{k}{2}, j_1 \cdots j_k\rangle & \binom{N}{k} \text{ states} \\
a_1^\dagger \cdots a_N^\dagger|\lambda_0\rangle = |\lambda_0 + \frac{N}{2}, j_1 \cdots j_N\rangle . &
\end{array}
$$

Due to the antisymmetry in (i, j) there are $\binom{N}{k}$ states with helicity $(\lambda_0 + k/2)$. Summing over all k, this gives a total of 2^N states. These are the group-theoretically originating states which constitute a representation of N-extended supersymmetry for massless states. We get a stack of helicities ranging from λ_0 to $\lambda_0 + N/2$. In general, the helicities will not be distributed symmetrically around zero. Thus this stack cannot be CPT invariant, since CPT flips the sign of helicities. In order to restore CPT invariance we require that if the helicity λ shows up in the stack, the parity reflected state with opposite helicity $-\lambda$ must complement the multiplet. A restriction motivated by physics originates from the fact that renormalizable interacting field theories can consistently only be formulated for fields with spin $< {}^3\!/_2$. Thus renormalizable field theories only exist for $N \leqslant 4$. The only known consistent (classical) couplings for $N = {}^3\!/_2$ fields occur in supergravity and gravity theories. But in these theories fields with spin $\geqslant {}^5\!/_2$ seem to have inconsistent couplings. Therefore, the highest supergravity theory is restricted to $N = 8$.

Here are a few examples for which the previous procedure at first delivers two states, namely aside from the start state with helicity λ_0 the state with $\lambda_0 + {}^1\!/_2$, and which contingently must be supplemented by further states in order to be invariant with respect to CPT transformations.

- $N = 1$ and $\lambda_0 = -{}^1\!/_2$ plus the CPT conjugate constructed from $\lambda_0 = 0$

Helicity	$-{}^1\!/_2$	0	${}^1\!/_2$
No. of states	1	2	1

This is nothing but the chiral supermultiplet underlying the Wess-Zumino model and corresponding to the degrees of freedom associated with the two scalar fields S and P (or, as used in another context, a complex scalar A) and a Majorana fermion ψ.

- $N = 1$ and $\lambda_0 = -1$ plus the CPT conjugate constructed from $\lambda_0 = {}^1\!/_2$

Helicity	-1	$-{}^1\!/_2$	${}^1\!/_2$	1
No. of states	1	1	1	1

This corresponds to the degrees of freedom associated with a Majorana fermion ψ and a vector boson A_μ in supersymmetric electrodynamics.

- $N = 1$; the graviton multiplet
  corresponding to the graviton and the gravitino.

Helicity	-2	$-3/2$	$3/2$	2
No. of states	1	1	1	1

- $N = 2$; $\lambda_0 = -1/2$

Helicity	$-1/2$	0	$1/2$
No. of states	1	2	1

  called the hypermultiplet and consisting of two Majorana spinors and a complex scalar. Observe that this multiplet is CPT-invariant from the outset.
- $N = 4$ with $\lambda_0 = -1$

Helicity	-1	$-1/2$	0	$1/2$	1
No. of states	1	4	6	4	1

  corresponding to a vector supermultiplet.
- $N = 8$ with $\lambda_0 = -2$

Helicity	-2	$-3/2$	-1	$-1/2$	0	$1/2$	1	$3/2$	2
No. of states	1	8	28	56	70	56	28	8	1

  that is a multiplet with 1 graviton, 8 gravitinos, 28 vector fields, 56 Majorana spin-$1/2$ fields, and 70 scalars. This is the only supermultiplet that can be constructed.

You may easily convince yourself that in four dimensions, there are the multiplets

- N=1 with chiral, vector, and gravity multiplet
- N=2 with hyper, vector, and gravity multiplet
- N=4 with vector and gravity multiplet
- N=8 with gravity multiplet.

Massive states in the absence of central charges
Consider next the case of a massive particle in the absence of central charges. In the rest frame $(p_\mu) = (m, 0, 0, 0)$

$$\{Q^i_\alpha, \bar{Q}^{\dot{\alpha}}_j\} = 2\, m \delta^i_j \delta_\alpha^{\ \dot{\alpha}}.$$

The representations of this algebra are easy to construct, since it is the algebra of two fermionic creation and annihiliation operators (up to a rescaling of the supercharges by $\sqrt{2\,m}$). For simplicity take first the case $N = 1$. Assume that the Q_α annihilates a state $|j_0\rangle$

$$Q_\alpha |j_0\rangle = 0.$$

Then we find the representations by a line of reasoning similar to the massless case: Apply successively the operators $\bar{Q}^{\dot\alpha}$ on the "ground state" $|j_0\rangle$. The charges $\bar{Q}^{\dot\alpha}$ are in the $\mathbf{2}_L$ representation which transforms as spin $^1\!/_2$ under rotations. Thus by the rule for the addition of angular momenta, the states $\bar{Q}^{\dot\alpha}|j_0\rangle$ have spins $(j_0 + {}^1\!/_2)$ and $(j_0 - {}^1\!/_2)$ for $j_0 \neq 0$. For $j_0 = 0$ they have only spin $^1\!/_2$. The combinations $\bar{Q}^{\dot\alpha}\bar{Q}^{\dot\beta}$ transform in the $\mathbf{2}_L \otimes_A \mathbf{2}_L = \mathbf{1}$ representation since the $\bar{Q}$'s anticommute. Thus they transform as a singlet under rotations, and so the state $\bar{Q}^{\dot\alpha}\bar{Q}^{\dot\beta}|j_0\rangle$ has the same spin as $|j_0\rangle$. Explicitly

$$
\begin{array}{lr}
|j_0\rangle & \text{spin} \quad j_0 \\
\bar{Q}^{\dot\alpha}|j_0\rangle & \text{spin} \quad j_0 \pm \dfrac{1}{2} \quad \text{if} \quad j_0 > \dfrac{1}{2} \\
\bar{Q}^{\dot\alpha}\bar{Q}^{\dot\beta}|j_0\rangle = -\dfrac{1}{2}\epsilon^{\dot\alpha\dot\beta}(\bar{Q}_{\dot\gamma}\bar{Q}^{\dot\gamma}|j_0\rangle) & \text{spin} \quad j_0.
\end{array}
$$

As examples, consider the following multiplets

- massive chiral multiplet

Spin	0	$^1\!/_2$
No. of states	2	1
Degrees of freedom	2	2

- massive vector multiplet

Spin	0	$^1\!/_2$	1
No. of states	1	2	1
Degrees of freedom	1	4	3

For the extended superalgebra, one similarly constructs 2^{2N} states out of a state $|j_0\rangle$ with total spin j_0 by applying successively the operators $\bar{Q}_i^{\dot\alpha}$. The maximal spin state $\bar{Q}_1^{\dot 2}\bar{Q}_2^{\dot 2}\ldots\bar{Q}_N^{\dot 2}|j_0\rangle$ has spin $j = j_0 + N/2$. From this it is obvious that massive multiplets of $N > 1$ supersymmetry always contain spin larger or equal to 1.

The previous results only hold if there are no central charges. Massive states in the presence of central charges are treated for instance in Sect. 4.3.3 of [47]

Superspace, Superfields, and Superactions

It is astounding that the complete chain of symmetry arguments we are acquainted from relating Poincaré symmetry to Minkowski space, to covariant fields, and to variational principles can be carried over to supersymmetry. This is in some more detail explained in Appendix B.3.

- Superspace
  Superspace is an extension of Minkowski space with its ordinary/even/bosonic coordinates x^μ by uncommon/odd/fermionic coordinates θ^a. The odd character of the θ-coordinates is due to their Grassmann nature: $\theta_a\theta_b + \theta_b\theta_a = 0$. Both types of coordinates are embraced by the superspace coordinates $z^M \sim (x^\mu, \theta_a)$. We can also compare coset spaces:
  – Minkowski spacetime = Poincaré group$\big/$Lorentz group
  – Superspace = Super-Poincaré group$\big/$Lorentz group
- Superfields
  These are fields living in superspace, that is they are functions $F(z^M) = F(x^\mu, \theta_a)$. If expanded in the odd coordinates θ^a, their Taylor expansion terminates at the fourth order because of the Grassmann nature of the θ-coordinates. The coefficients in such an expansion are Minkowski fields (e.g. scalar fields, spin-$^1\!/_2$ fields, vector fields). Thus a superfield represents a multiplet of fields with different spins. In general such a multiplet is reducible.
- Superactions
  Just as in “ordinary” field theories, the action encodes the field equations, variational symmetries, and the quantization, a superaction is the most compact device to encode a theory. A superaction is an expression
  $$S = \int dz^M \mathcal{L}$$
  where the super-Lagrangian $\mathcal{L}$ is a function of superfields and their (super)-derivatives.
- Supercovariance
  Although this is actually a topic of the next subsection, I remark here that supergravity can be formulated as a geometrical theory in superspace, if one requires that a theory should be covariant with respect to general coordinate transformations in superspace: $z^M \to z^M + \xi^M(z)$.

Breaking of Supersymmetry

Since in the world around us, bosons and fermions are quite different entities, supersymmetry–if latently present at all–must definitely be broken. There are various mechanism by which symmetry breaking can occur.

The most appealing version is “dynamical supersymmetry breaking” (DSB) since it provides an attractive way to achieve a large hierarchy between the Planck scale and the scale of supersymmetry breaking as first shown in [564]. In light of the well-understood mechanism of spontaneous symmetry breaking, one also reflects about spontaneously broken supersymmetry. However, spontaneous breaking of global supersymmetry entails a massless spin-$^1\!/_2$ Goldstone fermion, and such a particle has not been observed. (The neutrino can be ruled out.)

Another type of symmetry breaking is known by the term "soft supersymmetry breaking", in which supersymmetry is explicitly broken by additional terms in the Lagrangian. This seems less aesthetic, but it simply parametrizes our ignorance of how many Higgs-like particles and which kind of interactions are to be used as an input for spontaneous symmetry breaking.

The superfield formulation of supersymmetric theories suggest two symmetry breaking mechanism. You find these in the literature under the names Fayet-Iliopoulos (D-term) and O' Raiffeartaigh (F-term) symmetry breaking.

MSSM

As its name denotes, "Minimally Supersymmetric Standard Model" is the supersymmetric extension of the Standard Model with the minimal number of additional fields and the minimum number of couplings. The latter means in technical terms (see below) that one imposes R-parity. The model is displayed in all established books on supersymmetry, like [47], [369]; see also the review [252] and the "Supersymmetry Primer" [362]. At first it is plausible that the number of fields in the SM has to be doubled in going to the MSSM.

SM field	Spin	SuSy field	Spin
Leptons	$^1\!/_2$	sleptons	0
Quarks	$^1\!/_2$	Squarks	0
Gluons	1	Gluinos	$^1\!/_2$
W bosons	1	Winos	$^1\!/_2$
Z boson	1	Zino	$^1\!/_2$
Photon	1	Photino	$^1\!/_2$
Higgs boson	0	Higgsino	$^1\!/_2$
Graviton	1	Gravitino	$^3\!/_2$

Don't be surprised to see other beasts in the literature like binos (the SuSy partners of the electroweak B-bosons), neutralinos (subsuming the SuSy partners of the Z, the photon and the Higgs), charginos (instead of winos), gauginos (the fermionic SuSy partners of the gauge bosons). And don't be surprised not to find the masses of these exotic animals in the literature because so far these were not observed.

"Minimal" also refers to the number of Higgs fields in the MSSM. Whereas one doublet is the minimum in the SM, two of them are needed in case of supersymmetry in order to give masses to all fermions in the model (including the gluinos and Higgsinos). While the most general supersymmetric extension of the Standard Model would include terms leading to baryon and lepton non-conservation these are excluded in MSSM by imposing an additional discrete symmetry on the Lagrangian called R-parity and introduced by P. Fayet [175]. This is a $\mathbf{Z_2}$ symmetry under which each field has charge

$$R = (-1)^{(3B+L)+2s}$$

where s is the spin of the particle/field. All Standard Model elementary particles by this definition have $R = +1$, all sparticles have $R = -1$. This means that if two particles interact, they at most produce a pair of sparticles. And if a sparticle decays, the result is an odd number of sparticles and an even number of particles. As a consequence of R-parity conservation the lightest supersymmetric particle LSP must be stable–because there is no less massive object with $R = -1$. According to the rules of quantum mechanics a superposition of the photino, the zino and the Higgsino should exist. The LSP is the lightest of these "neutralinos". The LSP is a good candidate for the non-baryonic dark matter. The signature of an LSP in astrophysics observations comes from its creation in the decay of a next-lightest neutralino together with an electron-positron pair.

R-parity is a remnant of the continuous $\mathbf{U_R(1)}$ R-symmetry mentioned before. We saw that a particle and its supersymmetric partner transform differently. Specifically gaugino fields are charged under this symmetry. Symmetry breaking gives mass to the gauginos and the discrete R-symmetry remains. But be aware that this chain of arguments is only phenomenological. There is no fundamental reason why R-parity should be conserved. (The LSP only exists if R-parity is conserved). R-parity violation may be the source of Majorana masses and mixings for neutrinos.

Even the "minimal" model described above has 124 free parameters: Aside from the 19 parameters of the Standard Model there are additionally 5 real parameters and 3 CP-violating phases in the gaugino-Higgsino sector, 21 masses, 36 mixing angles and 40 CP-violating phases in the squark and slepton sector. There are various variants of the MSSM and their predictions are tested at the LHC. Amongst others the collision of two protons should produce either two gluinos or two squarks. These decay into lighter supersymmetric particles and into quarks. At the very end stands the decay into a quark and an LSP, and thus the signature of supersymmetry is some missing energy due to an undetected LSP.

8.3.4 Local Supersymmetry and Supergravity

Whither Supergravity?

You have no doubt gained the impression that symmetry arguments carry us up and away: Lorentz-invariance (and assumptions about mass dimensions) fix the Lagrangians for free spin-0 and spin-$^1/_2$ fields, which in turn have global phase symmetries. Enlarging these to local symmetries necessitates the introduction of gauge boson (spin-1) fields. In this subsection we start still another symmetry-inspired journey: We will find that spin-3/2 and spin-2 fields appear when globally supersymmetric theories are required to be locally supersymmetric. And as exposed in Subsect. 7.6.1, spin-2 fields are a signal of gravity!

There is a simple heuristic argument to show how gravity "arises" from local symmetry: Take the commutator of two successive local supersymmetry transformations of a field X:

$$\left[\delta_{\epsilon(x)}, \delta_{\epsilon'(x)}\right] X = 2\bar{\epsilon}(x)\gamma^{\mu}\epsilon'(x) P_{\mu} X$$

(in the four-component notation of spinors). On the right-hand side, we observe a spacetime dependent translation, and this is a characteristic sign of general relativity.

On the other hand, we know from the representation theory of the super-Poincaré group that multiplets arise in the form $(j, j + {}^1\!/\!_2)$. If there is a graviton with $j = 2$ (or $j + {}^1\!/\!_2 = 2$) there should be a supersymmetric partner with spin ${}^5\!/\!_2$ (or spin ${}^3\!/\!_2$). A spin-${}^5\!/\!_2$ theory is ruled out because such a fermion does not consistently interact with other fields. Consequently, one is dealing with a multiplet composed of a graviton and a gravitino, described by a symmetric tensor $h_{\mu\nu}$ and a Majorana spinor which also carries a Lorentz-index $(\psi_{a\mu})$. The latter is the Rarita-Schwinger field whose action and field equations are described in Sect. 5.3.5.

There are various ways to approach supergravity. One is to use superspace techniques. I will not treat this in this book, but refer to [545]. Another way is to more or less straightforwardly follow the recipe of gauging a global symmetry. The "more or less" in the previous sentence is used not without tongue in a cheek: The original work in this spirit [194] necessitated elaborate–even computer-supported–calculations. Nevertheless, they paved the way for a technique now also called the *Noether coupling method*, as particularly promoted in [547]. I will try to communicate the spirit of this approach in the next subsection. Still another way is to comprehend supergravity as a gauged super-Poincaré theory. This is treated in a separate subsection below.

Local Supersymmetry is Supergravity

The heading of this section and partly also the chain of arguments is taken from [47]. At this place we apply the "Noetherization" procedure illustrated for a $\mathbf{U(1)}$ symmetric theory in Sect. 5.3.4. The idea is to start with a globally supersymmetric matter action, coupling its Noether currents to a gauge field and determining successively extra terms in the action and in the transformations of the fields until local supersymmetry is established. In the following, I do not go through all the laborious details, but only want to convey the arguments why the gauging of a global supersymmetry necessitates the existence of a gravitino and a graviton.

Start from the simplest supersymmetric matter action, namely the kinetic term part of the Wess-Zumino model, that is (8.43):

$$S_0 = \int d^4x \left[\frac{1}{2}(\partial_\mu S \partial^\mu S) + \frac{1}{2}(\partial_\mu P \partial^\mu P) + \frac{1}{2} i \bar{\psi} \partial\!\!\!/ \psi\right].$$

This action is invariant with respect to the global transformations (8.44). If one allows the Majorana spinor ϵ to become spacetime dependent, S_0 ceases to be invariant. Rather

$$\delta S_0 = \frac{1}{2} \int d^4x \, (\partial_\mu \bar{\epsilon}) \gamma^\nu \gamma^\mu \left[\partial_\nu (S + i\gamma^5 P)\right] \psi =: \int d^4x \, (\partial_\mu \bar{\epsilon}_a) J^\mu_a(x)$$

with a spinorial current J_a^μ. As exemplified in Sect. 5.3.4, where a gauge field is introduced in order to compensate for the non-invariance, here a new field $\Psi_a^\mu(x)$–the gravitino–is brought in. It transforms as

$$\delta\Psi_a^\mu = \frac{2}{k}\partial^\mu \epsilon_a. \tag{8.64}$$

Additionally a further action term is introduced

$$S_1 = -\frac{k}{2}\int d^4x\, \bar{\Psi}_\mu J^\mu.$$

The constant k takes on the role of the gravitino coupling constant. Since Ψ^μ as a spinor has mass dimension $3/2$ and the mass dimension of the Majorana spinor ϵ is $(-1/2)$ the constant k has mass dimension (-1). Now

$$\delta(S_0 + S_1) = \int d^4x\, (\partial_\mu \bar{\epsilon}_a) J_a^\mu(x) - \frac{k}{2}\int d^4x\, (\delta\bar{\Psi}_\mu) J^\mu - \frac{k}{2}\int d^4x\, \bar{\Psi}_\mu \delta J^\mu.$$

Of course the procedure operates in such a way that the first two terms cancel. There is a remaining non-vanishing part, which among others contains terms quadratic in the scalar field S:

$$\begin{aligned}\delta(S_0 + S_1) &= -ik\int d^4x\, \bar{\Psi}^\mu \gamma^\nu \epsilon(x)\Big[\partial_\mu S \partial_\nu S - \frac{1}{2} g_{\mu\nu}\partial_\lambda S \partial^\lambda S + \dots\Big]\\ &=: -ik\int d^4x\, \bar{\Psi}^\mu \gamma^\nu \epsilon(x) T_{\mu\nu} + \dots.\end{aligned}$$

Here, the energy momentum tensor $T^{\mu\nu}$ for the scalar field arises (see Sect. 7.5.2). According to the Noetherization procedure it becomes possible to get rid of this term by introducing a further gauge field $g_{\mu\nu}$ which transforms as

$$\delta g_{\mu\nu} = -ik\big(\bar{\Psi}_\mu \gamma_\nu \epsilon(x) + \bar{\Psi}_\nu \gamma_\mu\big)\epsilon(x). \tag{8.65}$$

Thus the graviton enters the scene as the supersymmetric partner of the gravitino. (Remember that the photon enters the scene as the partner of the electron and positron.) In the spirit of the Noetherization procedure a further action term

$$S_2 = -\int d^4x\, g_{\mu\nu} T^{\mu\nu}$$

is introduced such that

$$\delta(S_0 + S_1 + S_2) = -\int d^4x\, g_{\mu\nu}\delta T^{\mu\nu}.$$

The procedure has still not terminated at this stage. Among other things, it turns out that one has to modify the transformation law (8.64) to

$$\delta\Psi_a^\mu = \frac{2}{k}\mathcal{D}^\mu \epsilon_a := \frac{2}{k}\big\{\partial^\mu \epsilon_a + \frac{1}{2}\omega_\mu^{\ ab}\sigma_{ab}\big\}, \tag{8.66}$$

where ω_μ^{ab} are the components of the spin connection. Additionally, further action terms quartic in the fields arise. The whole story becomes very complicated; details can be found in [178]. And this is not even the end; as already observed–typically for global supersymmetry–the resulting action is only invariant on-shell, signaling the need for auxiliary fields. In any case, the final action contains as a part the action for $N = 1$ supergravity

$$S_{SG} = \int d^4x\sqrt{|g|}\Big[\frac{1}{2k^2}R - \frac{i}{2}\epsilon^{\mu\nu\varrho\sigma}\bar{\Psi}_\mu\gamma_5\gamma_\nu D_\varrho\Psi_\sigma\Big] \tag{8.67}$$

which is invariant on-shell with respect to the transformations (8.65, 8.66). The first term is built out of the metric and the connection. The second term is the kinetic term for the gravitino and is the curved-space version of the Rarita-Schwinger action [434]; compare (5.78) with $m_{3/2} = 0$. The gravitational field equations are recovered if we identify the constant arising in the transformation of the Rarita-Schwinger field (see (8.64)) with the square root of the gravitational constant: $k = \sqrt{\kappa} = 2M_{Pl}^{-1}$. Then (8.67) represents the coupling of Einstein-Cartan gravitation with a Rarita-Schwinger field (Einstein-Cartan, since in building the scalar curvature R no assumption about a vanishing torsion was made).

Almost at the same time in 1976 as the previously-described construction of supergravity by D.Z. Freedman, S. Ferrara, and P. van Nieuwenhuizen [194] there appeared a derivation of a Lagrangian similar to (8.67) by S. Deser and B. Zumino [115], now in a first-order formulation.

Supergravity from a Gauged Supergroup

One may very well ask whether the previously described procedure has a structural legitimation in terms of symmetry groups and geometry. These structures were uncovered by S.W. MacDowell and F. Mansouri in their pioneering paper [356] (an independent related approach is due to A.H. Chamseddine and P.C. West [86]). They built the action (8.67) by starting from a super-Lie group **G** and introducing gauge fields for the generators of the associated Lie super-algebra. (The most relevant super-Lie groups are treated in Appendix B.4.) The next steps are to use these gauge fields for defining covariant derivatives, curvatures and actions in the spirit of (Poincaré) gauge theory. MacDowell and Mansouri derived N=1 supergravity with a cosmological constant from gauging $\mathbf{OSp(1/4)}$ with its four supersymmetric generators and ten generators for $\mathbf{Sp(4)}$. Now it is the case that $\mathbf{Sp(4)}$ is locally isomorphic to the de Sitter group $\mathbf{O(2,3)}$. The objective of gauging $\mathbf{OSp(1/4)}$ is to be able to make a Inönü-Wigner contraction to the super-Poincaré algebra. A version of a de Sitter gauge theory was described in Sect. 7.6.3 and one can immediately extend most of the results to the supersymmetric case: Aside from the gauge potential components A^i and A^{ij} associated to the generators of de Sitter algebra, we introduce further fermionic ones, labeled as $\overline{\psi}^\alpha$ such that (7.125) becomes extended to

$$A_\mu = A^i_\mu\Pi_i + \frac{1}{2}A^{ij}_\mu J_{ij} + \bar{\psi}^\alpha_\mu Q_\alpha.$$

The group algebra of $\mathbf{OSp(1/4)}$ determines the field strength

$$F^{i}_{\mu\nu} = \partial_\mu A^{i}_{\nu} + A^{i}{}_{k\mu} A^{k}_{\nu} - \frac{i}{4}\bar{\psi}_\mu \gamma^i \psi_\nu - (\mu \leftrightarrow \nu) \tag{8.68a}$$

$$F^{ij}_{\mu\nu} = \partial_\mu A^{ij}_{\nu} + A^{i}{}_{k\mu} A^{kj}_{\nu} - \frac{1}{\mathcal{R}}\bar{\psi}_\mu \gamma^{ij} \psi_\nu - \frac{1}{\mathcal{R}^2} A^{i}_{\mu} A^{j}_{\nu} - (\mu \leftrightarrow \nu) \tag{8.68b}$$

$$\Phi^{\alpha}_{\mu\nu} = \partial_\mu \bar{\psi}^{\alpha}_{\nu} + \bar{\psi}^{\beta}_{\mu} (\gamma_{kl})^{\alpha}_{\beta} A^{kl}_{\nu} + \frac{1}{\mathcal{R}} (i\gamma)^{\alpha}_{\beta} A^{i}_{\mu} \psi^{\beta}_{\nu} - (\mu \leftrightarrow \nu) \tag{8.68c}$$

with the de Sitter radius $\mathcal{R}$. These expressions exhibit the extension of (7.126) to the supersymmetric case. Pretty much as in the case of the de Sitter gauge theory, the condition $F^{i}_{\mu\nu} = 0$ is imposed on the translational field strengths and one obtains an invariant YM-type action

$$S_{SdS} = \int d^4x\, \epsilon^{\mu\nu\rho\sigma} \left(\epsilon_{ijkl} F^{ij}_{\mu\nu} F^{kl}_{\rho\sigma} + \frac{i}{2} \bar{\Phi}_{\mu\nu} \gamma_5 \Phi_{\rho\sigma} \right).$$

This action consists of five terms [85] ranging in powers of $\mathcal{R}$ from $\mathcal{R}^0$ to $\mathcal{R}^{-4}$. The term proportional to $\mathcal{R}^0$ is identical with (7.129a) and corresponds to the Gauß-Bonnet topological invariant. The term with $\mathcal{R}^{-1}$ is another boundary term. The next term, proportional to $\mathcal{R}^{-2}$, delivers the supergravity action (8.67). There is a further mass-like term for the gravitino (proportional to $\mathcal{R}^{-3}$), and finally the cosmological term.

Beyond $N = 1$, $D = 4$ Supergravity

- Extended supergravity
  The version of supergravity described above contains only one graviton and one gravitino. As such it cannot be considered a unified model. This original version is now called $N = 1$ supergravity, and it has been extended to N-supergravity, where N runs from 1 to 8. Mathematically, N can go beyond eight, but as argued before, for physical reasons the maximum is $N = 8$.
  N-supergravity possesses aside from the graviton also N gravitinos plus further fields. For example, $N = 2$ supergravity theory holds one graviton, two gravitinos and one spin-1 field. Further on, $N = 3$ supergravity has 1 graviton, 3 gravitinos, 3 spin-1 particles and 1 spin-$^1\!/_2$ field. For the largest theory ($N = 8$) there is 1 graviton, 8 spin gravitinos, 28 vectors, 56 Majorana spin-$^1\!/_2$ fields and 70 scalars, all massless. This theory has $\mathbf{O(8)}$ as a "hidden" gauge group.
  In the early 1980's great expectations were held pertaining to the $N = 8$ model being the ultimate unified theory of nature: (1) It is the largest extended supergravity based on a unique irreducible multiplet in which the graviton appears together with particles of lower spin. "Matter" is present in the multiplet and cannot–or that is to say, must not be coupled separately. (2) It contains 28 gauge bosons

–even more than needed within the Standard Model, and 56 spin-$^1/_2$ fermions –seemingly more than enough to fit into the Standard Model. But the great expectations faded away for several reasons. Firstly, the gauge group $\mathbf{O(8)}$ does not contain $\mathbf{SU(3)xSU(2)xU(1)}$ as a subgroup. Furthermore, after symmetry breaking, the 56 fermions turn out to be insufficient in number. Next, pretty much as all other extended supergravity models, the $N = 8$ model is non-chiral, and no simple reasonable way is known to arrive at the chiral Standard Model. And last but not least, although supergravity models tend to be "more finite" than the non-super ones, there were indications in the late 1970's that supergravity is UV infinite beyond the two-loop level. There was still hope that the $N = 8$ model eventually might prove to be finite because its cousin, the global maximal $N = 4$ super Yang-Mills is finite to all orders. The finiteness issue of maximal supergravity is still open. The respective calculations are quite difficult because of the contribution of tens of thousands of Feynman graphs already at the three-loop level. Therefore (and also because superstrings seemed to offer promising opportunities in 1984, after a decade of silence) the so-far glaring hopes of supergravity dimmed away. However, a wind of change began blowing rather recently with an idea of improved bookkeeping device for Feynman graphs [46]; see [387] where H. Nicolai puts this into context.

- Supergravity in higher dimension
  Supersymmetry can be formulated in dimensions other than four. However, its formulation and properties change qualitatively between different dimensions. Most important, it depends on the dimension which types of spinors are allowed to exist. So while in $D = 4$, both Weyl and Majorana spinors are possible, in $D = 11$ only Majorana spinors exist. Also the form of the anticommutators of the supercharges and the automorphism group depend on the dimension. While in four dimensions the automorphism group is $\mathbf{U(N)}$, in eleven dimensions it is $\mathbf{SO(N)}$. And, last but not least, the types of multiplets depend on both D and N; see e.g. [496].
  The interest in higher dimensions came about because some extended supergravity theories in $D = 4$ were found to be equivalent to certain higher-dimensional supergravity theories. W. Nahm showed that the maximal dimension in which supergravity can be formulated is $D = 11$ [378]. (This holds true if one restricts the theories to a $(+, -, -, \ldots -)$ signature of the metric, that is to only one time dimension.) For higher dimensions–in order to get a consistent theory–massless fields with spin greater than two are needed. In general, the formulation of a higher-dimensional supergravity needs fewer field types than its N-extended equivalent. Supergravity in higher dimension is the revival–or the revitalization–of Kaluza-Klein models; thus it is also called " Modern Kaluza-Klein" [11]. In their review on Kaluza-Klein Supergravity [141], the authors admit that "We do not yet know whether supergravity is a theory of the real world, nor whether the real world has more than four dimensions as demanded by Kaluza-Klein theories. However, our research in this area has convinced us that the only way to do supergravity is via Kaluza-Klein and that the only viable Kaluza-Klein theory is supergravity."

In 1978 E. Cremmer, B. Julia, and J. Scherk [105] [104] showed that $(D = 11, N = 1)$ supergravity corresponds to $(D = 4, N = 8)$ supergravity after compactification on a B^7 manifold. While the field content of the $(D = 4, N = 8)$ theory is rather complicated and knotty, in its $D = 11$ formulation this is neatly arranged: There are three types of fields namely 11-beins e^A_M, spinorial fields Ψ_M, and antisymmetric three-index tensor fields A_{MNP}. Here M, N denote the indices for 11-dimensional curved spacetime, the indices A, B enumerate the 11-dimensional flat spacetime. The graviton described by e^A_M has $\{(D\text{-}1)(D\text{-}2)\}/2\text{-}1 = 44$ degrees of freedom. Since in $D = 11$ the Lorentz group is $\mathbf{SO(10, 1)}$ and thus a Dirac spinor has $2^5 = 32$ components, the gravitinos described by the Majorana $^3\!/_2$-spinor Ψ_M represents $^1\!/_2 \cdot \{(D\text{-}2) \times 32\text{-}32\} = 128$ degrees of freedom. The 128-44 = 84 missing bosons can be assembled in the completely antisymmetric tensor field A_{MNP} with three indices, since this field has $\binom{D-2}{3}$ components. Since in 11 dimensions no matter and no Yang-Mills supermultiplets exist, the source of gravity is fixed by supersymmetry to consist of the A_{MNP} and the Ψ_M fields. The very fact that "force" and "matter" unambiguously determine each other—they constitute the supergravity multiplet—makes this theory unique.
Later, it was found that there is a further version of $D = 11$ supergravity [118] being a gauged version with local $\mathbf{SO(8)}$. These theories are, as you can believe, quite complex, and it even took some time to entangle their symmetries. Cremmer and Julia found a "hidden" $\mathbf{E_7}$, and only during the last ten years or so have yet larger symmetries been found (infinite Lie groups or Kac-Moody algebras[12]).
In 1980 P. Freund and M.A. Rubin proved that 11-dim supergravity can spontaneously compactify into the form $V_4 \times V_7$ [197]. Their scheme was described in a previous section (see Sect. 8.2.3): Either V_4 contains the time dimension, in which case V_4 is an AdS space and V_7 is an arbitrary small Einstein 7-manifold. Or else V_7 is an AdS-space, and V_4 an Einstein 4-manifold. The first case seems off course to be more appropriate to describe the world we know of (apart from the large cosmological constant arising the model), but as yet there are no established arguments to rule out the second case. Modulo this ambiguity it seems that through supersymmetry the dimension of spacetime can be calculated! The details of the 4D physics depend on the topology of V_7. Thus for instance the two different versions of $N = 8$ supergravity are recovered by either the topology of a 7-torus or a 7-sphere.

From 'Flat Space Theory' to 'Supergravity'

There are two ways to implement supergravity: You either start from a theory that is supersymmetric and impose general coordinate invariance, or else you make a "flat" theory generally covariant and afterwords impose supersymmetry.

---

[12] If you are mathematically inclined and want to learn more about E_{10}, you may try [388]; its presentation here would go far beyond the scope of this book.

$$\begin{array}{ccc} \text{FLAT SPACE THEORY} & \xrightarrow{\text{supersymmetry}} & \text{SUPERSYMMETRIC THEORY} \\ \downarrow & & \downarrow \\ \text{general coordinate invariance} & & \text{general coordinate invariance} \\ \downarrow & & \downarrow \\ \text{GRAVITY} & \xrightarrow[\text{supersymmetry}]{} & \text{SUPERGRAVITY} \end{array}$$

In previous sections the first approach was pursued. An instructive example ('supergravity in 1 dimension') for the latter approach is treated in Sect. 3.7 of [232].

8.3.5 Instead of a Conclusion

Supersymmetry has a wide scope, ranging from super-mathematics to super-phenomenology. A balanced account between these two poles is given in the classic [545]. It is indeed possible to transfer notions from standard differential geometry, group theory, functional analysis to e.g. supernumbers, superfunctions, supermanifolds, super Lie groups, super Hilbert spaces, comprehensively treated in [122]. For the mathematics of graded Lie algebras see also [101]. On the other hand, since supersymmetry is definitely not an exact symmetry of nature, it is obvious to ask where there could be signals in experiments (such as those performed at the LHC). This phenomenological aspect is the case of the minimally extended Standard Model (MSSM), for a short and topical account see [45], [371]. A good pedagogical exposition is given by [362] and a balanced theoretical and phenomenological presentation is to be found in [47]. The most influential "classic" articles from the seventies and the early eighties are collected in [176], and the standard reference for supergravity is the Physics Reports article by P. van Nieuwenhuizen [519]. A compact, but very readable overview is [179].

Let me finish this subchapter on supersymmetry with two quotations, expressing opinions which I also share: (1) H. Nicolai [387]: "For sure, the discovery of supersymmetry at the Large Hadron Collider (LHC) would open many new avenues and revolutionize particle physics. But even if no supersymmetric particles are found, it appears that supersymmetry is here to stay: it has become an integral part of mathematics, having inspired several Field Medalists, and will surely continue to play a key role in our search for a consistent (finite) theory of quantum gravity." (2) In the preface of his Volume III field theory book [536] S. Weinberg concedes "‥, and many other physicists are reasonably confident that supersymmetry will be found to be relevant to the real world, and perhaps soon"[13].

---

[13] I lost a bet that LHC detects indications of supersymmetry prior to discovering any Higgs.

8.4 Further Speculations

This chapter is rather speculative compared to the previous ones: experimentally, thus far there are no signs for X-particles, of extra dimensions, or for supersymmetric partners of each and every elementary particle. All these phantasies are inspired by ideas of symmetry. And there are even more phantasies.

8.4.1 Compositeness and Technicolor

The history of physics of the last 150 years or so revealed an increasingly granular picture of matter: From molecules to atoms to nuclei to nucleons to quarks (leaving aside the leptons). Is it possible that quarks and leptons can be resolved into substructures? Although the experiments with the largest microscopes reveal that these building blocks must be smaller than 10^{-15} mm, they can of course not be excluded. Prior to the successful formulation and experimental confirmation of the Standard Model, there were ideas about substructure particles generically called "preons". The various versions of compositeness of all particles in the Standard Model leaves open many issues, nevertheless preons arise now and then in the literature [116].

The motivation of the technicolor models is a dissatisfaction with the Higgs sector in the Standard Model. Indeed the mechanism of electroweak symmetry breaking is rather *ad hoc* and artificial[14]. Why for instance does one have to introduce by hand a tachyon in the beginning, aside from thus eventually getting the non-trivial vacuum? Technicolor aims for dynamical electroweak symmetry breaking, and considers the Higgs to be a composite of more fundamental fermions. The basic idea of technicolor can be illustrated on a model close to QCD: Consider a $\mathbf{SU(N_{tc})}$ gauge theory with techni-fermions U and D in the fundamental representation:

$$\begin{pmatrix} U \\ D \end{pmatrix}_L \qquad \text{and} \qquad U_R, D_L.$$

Similar to QCD there is a kinetic Lagrangian term

$$\mathcal{L}_{tc} = i\bar{U}_L \not{D} U_L + i\bar{U}_R \not{D} U_R + i\bar{D}_L \not{D} D_L + i\bar{D}_L \not{D} D_L$$

which has a chiral $\mathbf{SU(2)_L} \times \mathbf{SU(2)_R}$ symmetry. And similar to the situation with pions arising as pseudo-Goldstone bosons, the breaking of the technicolor-originating chiral symmetry gives rise to three Goldstone bosons. In order to avoid anomalies the left-handed techni-fermions are attributed hypercharge zero and the right-hand ones hypercharge $\pm 1/2$. By spontaneous symmetry breaking the would-be Goldstone bosons become the W- and Z-bosons. In order to yield the observed masses, the scale for technicolor must be in the range of the parameter $v \approx 250$ GeV. In the

[14] In German we have the slogan "von hinten durch die Brust ins Auge" which I do not translate.

GSW model the Higgs boson also has the role of giving masses to the fermions. This happens by renormalizable Yukawa interaction terms. Since with technicolor the Higgs is a composite of two techni-fermions, the interaction is of the four-fermion type, and thus non-renormalizable. The most popular means to get around this is *extended technicolor*, where the extension refers to the introduction of new broken gauge interactions between the known fermions and the techni-fermions; for more details see Chap. 8 in [369]. Interestingly one of the father's of technicolor utters "…(extended technicolor)… have potential problems with flavor-changing neutral current weak interactions, and though these problems may be surmounted, the added complications reduce their attractiveness." [536] (end of Sect. 21.4).

8.4.2 Strings and Branes

String theory, as it is called today, originates from an observation made in the 1960's that hadronic interactions exhibit a feature called "duality": Scattering amplitudes with two incoming and two outgoing particles could be described either by an exchange of particles in the s-channel or in the t-channel (where s and t are the Mandelstam variables built from the four-momenta of the particles involved.) One of the explanations of this duality culminated in the idea of Regge poles in S-matrix theory[15]. G. Veneziano rediscovered a formula for Euler's beta function which exhibits just this feature. It was found that the Euler-Veneziano formula could be interpreted in terms of mesons as oscillation modes of a relativistic string. This view obtained a further twist after it was found that the dynamics of the string can be derived from an action which is proportional to the area swept out by the string in the course of its history; this goes by the names of Y. Nambu, T. Goto, H. B. Nielsen, and L. Susskind. The idea was born that elementary particles are not point-like but are one-dimensional objects.

Assume that the string propagates in a D-dimensional spacetime. Let the world-sheet/area swept by the string be characterized by functions $X^\mu(\sigma, \tau)$, where the timelike τ and the spacelike σ parametrize the area (see Fig. 8.2); denote $(\zeta^\alpha) = (\tau, \sigma)$. The worldsheet metric is given by

$$g_{\alpha\beta} = \eta_{\mu\nu} \frac{\partial X^\mu}{\partial \zeta^\alpha} \frac{\partial X^\nu}{\partial \zeta^\beta}$$

and the Nambu-Goto string action is proportional to the area of the world-sheet:

$$S_{NG} = \frac{1}{2\pi\alpha'} \int d^2\zeta \sqrt{|\det g_{\alpha\beta}|}.$$

Here $(2\pi\alpha')^{-1}$ is the string tension in terms of the slope α' of the Regge trajectories. Notice that the Nambu-Goto action is nothing but GR in two dimensions. Indeed, in

[15] My thesis back in 1970 dealt with Regge trajectories in pion-nucleon scattering.

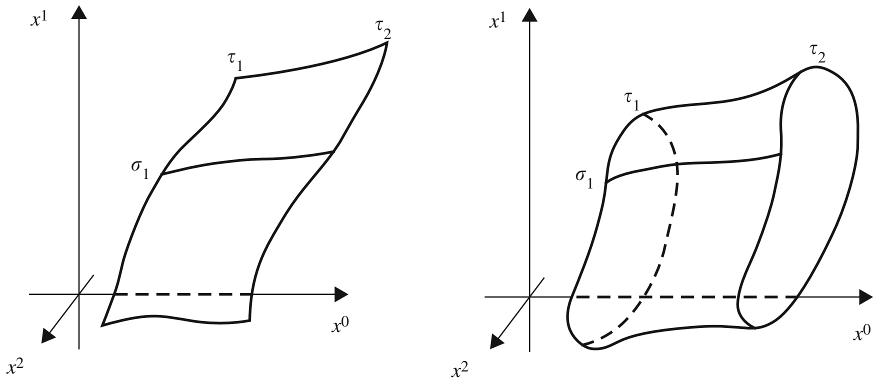


Fig. 8.2 Worldsheets for Open and Closed Strings

two dimensions a curvature term $\sqrt{-g}R$ is a total derivative, and thus the Nambu-Goto action represents the cosmological term.

There is another formulation of the relativistic string named after A. Polyakov:

$$S_P = \frac{1}{4\pi\alpha'}\int d^2\zeta\sqrt{-h}h^{\alpha\beta}\partial_\alpha X^\mu\partial_\beta X_\mu = \frac{1}{4\pi\alpha'}\int d^2\zeta\sqrt{-h}h^{\alpha\beta}g_{\alpha\beta}. \qquad (8.69)$$

Here $h^{\alpha\beta}$ is the intrinsic metric on the worldsheet, and not the induced one (namely $g_{\alpha\beta}$). The Polyakov action resembles a description of two-dimensional gravity coupled to D scalar fields. It is invariant under two-dimensional diffeomorphism $\hat{\zeta}^\alpha = \zeta^\alpha + \xi^\alpha$ where the metric $h_{\alpha\beta}$ and the X^μ indeed transform like a symmetric tensor and like scalars:

$$\bar{\delta}_\xi h_{\alpha\beta} = -h_{\alpha\gamma}\xi^\gamma_{,\beta} - h_{\gamma\beta}\xi^\gamma_{,\alpha} - h_{\alpha\beta,\gamma}\xi^\gamma \qquad\qquad \bar{\delta}_\xi X^\mu = -X^\mu_{,\alpha}\xi^\alpha.$$

The Polyakov action is also Weyl invariant, that is invariant under local scale transformations

$$\hat{h}_{\alpha\beta} = e^{2\Phi(\zeta)}h_{\alpha\beta}$$

which is specific to two dimensions because $\sqrt{-h}h^{\alpha\beta}$ itself stays invariant. Furthermore, there is the invariance with respect to global Poincaré transformations (i) $X^\mu \to X^\mu + a^\mu$ (ii) $X^\mu \to X^\mu + \omega^\mu_{\ \nu}X^\nu$.

Since there is no kinetic term for the intrinsic metric, the field equations with respect to the metric $h^{\alpha\beta}$ are simply algebraic relations

$$0 = \frac{\partial S_P}{\partial h^{\alpha\beta}} \propto \left(g_{\alpha\beta} - \frac{1}{2}h_{\alpha\beta}h^{\gamma\delta}g_{\gamma\delta}\right)$$

from which

$$\det g_{\alpha\beta} = \frac{1}{4}h(h^{\gamma\delta}g_{\gamma\delta})^2.$$

Inserting this into the original action one recovers the Nambu-Goto action. Thus the two actions are equivalent on the classical solutions. The theories defined by S_{NG} and by S_P are not necessarily equivalent on the quantum level, since the path integral also has contributions from non-solutions. The Polyakov action is preferred since it does not have the nasty square root in the string fields, and since it is immediately amenable to path integral and BRST techniques.

The energy-momentum tensor is calculated to be

$$T_{\alpha\beta} := \frac{2}{\sqrt{-h}} \frac{\partial S_P}{\partial h^{\alpha\beta}} = (g_{\alpha\beta} - \frac{1}{2} h_{\alpha\beta} h^{\gamma\delta} g_{\gamma\delta}).$$

It satisfies $\partial_\alpha T^{\alpha\beta} = 0$ and $T^{\alpha}_{\ \alpha} = 0$ which we know from Sect. 3.5.1 to be signatures of conformal symmetry. On-shell, the energy-momentum tensor vanishes

$$T_{\alpha\beta} \doteq 0.$$

This again is fully consistent with 2D gravity for which the "left-hand" side of Einstein's field equations vanishes.

The gauge freedom originating from the diffeomorphism and the scaling symmetry can be used to choose $h_{\alpha\beta} = e^{2\Phi} \eta_{\alpha\beta}$ for which the Polyakov action simplifies to

$$S_{BS} = \frac{1}{4\pi\alpha'} \int d^2\zeta \, \eta^{\alpha\beta} \partial_\alpha X^\mu \partial_\beta X_\mu = \frac{1}{4\pi\alpha'} \int d\tau d\sigma (\dot{X}^2 - \check{X}^2)$$

with the definitions $\dot{X}^\mu := \frac{\partial X^\mu}{\partial \tau}$ and $\check{X}^\mu := \frac{\partial X^\mu}{\partial \sigma}$. This action has a remaining conformal symmetry, which might be broken by a further gauge condition. The field equations now become the string (*sic!*) equations

$$\ddot{X}^\mu - \check{\check{X}}^\mu = 0.$$

These are to be solved with the constraints resulting from the vanishing of the energy-momentum tensor

$$\dot{X} \cdot \check{X} = 0 \qquad \dot{X}^2 + \check{X}^2 = 0. \tag{8.70}$$

Further there are the boundary conditions for
–closed strings: $X^\mu(\tau, \sigma + \pi) = X^\mu(\tau, \sigma)$ (periodicity condition)
–open strings: either $\partial_\sigma X^\mu(\tau, 0) = 0 = \partial_\sigma X^\mu(\tau, \pi)$ (Neumann condition)
or $X^\mu(\tau, 0) = a^\mu,\ X^\mu(\tau, \pi) = b^\mu$ (Dirichlet condition).
For the open string one can also have the situation that one string end obeys the Neumann and the other end the Dirichlet boundary condition. The Dirichlet boundary condition became investigated in the relativistic string model only in the early 1980's, but it experienced a boom with the notion of D-branes; a little more on this below. Interestingly, this condition arrives on the scene because of duality properties within the family of string models.

Later, we will need the Hamiltonian for the bosonic string in the conformal gauge:

$$H := \int_0^\pi d\sigma(\dot{X}^\mu \Pi_\mu - L_{BS}) = \frac{1}{4\pi\alpha'}\int_0^\pi d\sigma(\dot{X}^2 + \check{X}^2)$$

where the canonical momenta to the fields are $\Pi_\mu := \partial\mathcal{L}_{BS}/\partial\dot{X}^\mu$.

The next steps will only be sketched, for details see any book on string theory [123], [240], [308], [352], [416], [583], or the review article [349]. In the following, I also will only deal with the open string. The closed string can be treated in a similar way. The general solution of the string equations for the open string is of the form

$$X^\mu(\sigma,\tau) = q^\mu + 2\alpha' p^\mu \tau + i\sqrt{2\alpha'}\sum_{n\neq 0}\frac{\alpha_n^\mu}{n}e^{-in\tau}\cos n\sigma \tag{8.71}$$

where q^μ and p^μ are interpreted as the position and momentum of the center of mass, and the α_n^μ are the Fourier modes of the oscillator. Because the X^μ are real the coefficients obey $\alpha_{-n}^\mu = (\alpha_n^\mu)^*$.

The conditions (8.70), expressed by (8.71) become the *Virasoro operators*

$$\begin{aligned} L_m &:= \frac{1}{8\pi\alpha'}\int_0^\pi d\sigma\left[e^{im\sigma}(\dot{X}^\mu + \check{X}^\mu)^2 + e^{-im\sigma}(\dot{X}^\mu - \check{X}^\mu)^2\right] \\ &= \frac{1}{2}\sum_{-\infty}^{\infty}\alpha_{m-n}\alpha_n \qquad\qquad \text{where} \quad \alpha_0^\mu := \sqrt{2\alpha'}p^\mu, \end{aligned}$$

by which the Hamiltonian is identical with the Virasoro operator L_0:

$$H = \frac{1}{2}\sum_{-\infty}^{\infty}\alpha_{-n}\alpha_n.$$

Assuming the ("equal τ") Poisson brackets of the string field X^μ and its momenta Π_μ

$$\{X^\mu(\sigma,\tau), \Pi_\nu(\sigma',\tau)\} = \delta_\nu^\mu\delta(\sigma,\sigma')$$

with a delta-function adapted to the boundary conditions (see [491]), the Poisson brackets of the coefficients in (8.71) become, with $\Pi^\mu = (1/2\pi\alpha')\dot{X}^\mu$

$$\{\alpha_m^\mu, \alpha_n^\nu\} = -im\,\delta_{(m+n)0}\,\eta^{\mu\nu} \qquad\qquad \{q^\mu, p_\nu\} = \eta^{\mu\nu}.$$

These lead to the Virasoro algebra

$$\{L_m, L_n\} = -i(n-m)L_{m+n}$$

which was derived in 3.5.1. as typical for the generators of conformal symmetries in two dimensions.

To prepare the transition to the quantized bosonic string, define $\alpha_m^\mu = \sqrt{m} a_m^\mu$, $\alpha_{-m}^\mu = \sqrt{m} a_m^{\mu\dagger}$. The a_m^μ and $a_m^{\mu\dagger}$ have the algebra of annihilation and creation operators, respectively, with commutators $[a_m^\mu, a_n^{\mu\dagger}] = \delta_{mn}\eta^{\mu\nu}$ typical for a collection of harmonic oscillators. Define the ground state of the theory as that which is annihilated by the operators a_m^μ. If one chooses this to be also an eigenstate of the center of mass operator we have

$$a_m^\mu |0, p^\mu\rangle = 0 \text{ (for } m > 0) \qquad\qquad \hat{p}^\mu |0, p^\mu\rangle = 0.$$

A spectrum of states is generated by successively applying creation operators. Now one is facing two problems. For one, the states created this way do not obey the conditions set by the Virasoro operators. Furthermore, the states may have negative norm due to $\langle 0|a_m^\mu a_m^{\mu\dagger}|0\rangle = \eta^{\mu\mu}\langle 0|0\rangle$. These two problems turn out to be related. The quantum version of the Virasoro algebra receives a central charge because of normal ordering:

$$\left[\hat{L}_m, \hat{L}_n\right] = (m-n)\hat{L}_{m+n} + \frac{D}{12}m(m^2-1)\delta_{(m+n)0}, \qquad \hat{L}_m := \frac{1}{2}\sum_{-\infty}^{\infty} :\alpha_{m-n}\alpha_n:.$$

As a consequence one cannot impose the condition $\hat{L}_m|\Phi\rangle = 0$ for all m on physical states. The best one can do is to define a physical state as

$$\hat{L}_m|phys\rangle = 0 \qquad \text{for} \quad m > 0 \qquad (\hat{L}_0 - a)|phys\rangle = 0$$

where a is some constant, later determined from consistency. (This procedure resembles the Gupta-Bleuler quantization of electrodynamics where the positive frequency part of the Lorenz condition $\partial_\mu A^\mu$ is requested to annihilate on physical states.) Now it turns out that by this definition of a physical state all negative-norm states are absent if the bosonic string is embedded in 26 dimensions and if $a = 1$. These conditions also arise in other quantization procedures. One might for instance try to solve the Virasoro constraint equations and to describe the theory directly in terms of physical degrees of freedom only. (This also has an analogy in QED if one introduces the transverse and longitudinal parts of the vector potential and its canonical conjugates.) For the string one can define the true degrees of freedom in certain gauges (remember that there is still the conformal symmetry in the theory) and the light-cone gauge became an established one for this purpose. I do not go into details here–they are found in any book on strings. While the Gupta-Bleuler quantization is covariant and plagued by possible negative-norm states, the light-cone quantization has no problems with negative-norm states, but is not manifestly Lorentz invariant. Lorentz invariance must be proven at the end, and it was found that the Lorentz generators indeed obey the Lorentz algebra if $D = 26$ and $a = 1$. These numbers appear again in the BRST quantization of the bosonic string in that the BRST charge is nilpotent only for these values. Notice that the existence of a critical dimension for the string has its origin in the conformal invariance of the model.

As it turns out (and this is best be seen in the light-cone gauge), the lowest-lying states for the string are a tachyon and 24 massless vector particles. All states can be represented on a Chew-Frautschi plot (spin versus mass) as lying on a set of straight and parallel trajectories. These are none other than the Regge trajectories from hadronic physics, and the number $a = 1$ carries the meaning of what was in the late 1960's named the intercept in the dual resonance models.

As stated before, the previous considerations can be extended to the closed string. Because of the other boundary condition (namely periodicity), the expansion has now two series of coefficients α_n^μ and $\bar{\alpha}_n^\mu$, and accordingly a further set of Virasoro generators. Once more the theory is only consistent for $D = 26$ and $a = 1$. The lowest-lying states of the closed string are again the tachyonic ground state and a massless symmetric traceless tensor (spin-2).

The dual string model was *en vogue* around the time when the Gell-Mann/Zweig quark model aroused attention. Therefore it is understandable that one tried to merge both models: Mesons were pictured as a system of a quark and an antiquark connected by a string[16]. Indeed the $\mathbf{SU(3)}$ symmetry could be imposed on the string by so-called Chan-Paton factors. In the 1970's fermions were brought into the string model (P. Ramond, J. Schwarz, A. Neveu) and the critical dimension became reduced to $D = 10$. This inclusion of fermions was one of the very first realizations of supersymmetry. Another welcomed side-effect of going from strings to superstrings was the farewell to the tachyonic state in the spectrum.

We noticed that the bosonic string action (8.69) looks like the coupling of D scalar fields X^μ to 2-dimensional gravity with metric $h_{\alpha\beta}$. And we know from GR that if one wants to include fermions ψ^μ into a D-dimensional theory one needs to introduce D-beins in the game; for the bosonic string these are zweibeins e_α^a. In order to save the Bose-Fermi balance further fermionic fields χ_α are needed. The complete process of constructing the full superstring action including its bouquet of symmetries (reparametrizations, 2-dimensional Lorentz transformations, Weyl- and super-Weyl transformations, supersymmetry) is described in [352]. The gauge freedom can be exploited to define the superconformal gauge $e_\alpha^a = e^\Phi \delta_a^\alpha$, $\chi_\alpha = \rho_\alpha \lambda$ as the generalization of the conformal gauge. In this gauge the fermionic string is given by the action

$$S_{FS} = \frac{1}{4\pi\alpha'} \int d^2\zeta \left[\partial_\alpha X^\mu \partial^\alpha X_\mu - i\bar{\psi}^\mu \rho^\alpha \partial_\alpha \psi_\mu\right]$$

where ρ^α are two-dimensional Dirac matrices. In the extension of the bosonic string one has two choices of boundary conditions for the fermionic fields ψ_0 and ψ_1. They are named after their first investigators P. Ramond, and A. Neveu and J. Schwarz, respectively:

$$(R) : \psi_0(\pi, \tau) = \psi_1(\pi, \tau) \qquad\qquad (NS) : \psi_0(\pi, \tau) = -\psi_1(\pi, \tau).$$

These different boundary conditions give rise to two different oscillator-mode expansions for the fermionic field. Accordingly they lead to modified Virasoro operators

[16] In 1976 I published work on a (not quite satisfactorily functioning) string-quark baryon.

and an enlarged Virasoro algebra. The details are not of importance here. The most significant point is the reduction of the critical dimension from 26 to 10. Further $a = 0$ for (R) and and $a = 1/2$ for (NS).

Right from the beginning, the pioneers of the string model were haunted by string states with zero mass and spin 2, which does not have a corresponding particle in the hadronic spectrum. String theory faded away, while the quark model of the strong interactions was successfully implemented. Nevertheless many insights into the rich mathematical structure of the early string model were absorbed in other fields, especially supergravity. J. Scherk and J. Schwarz gave convincing arguments that the massless spin-2 state should be interpreted as a graviton. In 1984 M. Green and J. Schwarz showed that string theory is free of anomalies if it has an internal gauge group $\mathbf{O(32)}$ or $\mathbf{E_8} \times \mathbf{E_8}$. From this it seemed that the string model was an excellent candidate not only for quantum gravity, but also a version of a unification of all fundamental forces. The theory is completely finite and has no free parameters. There was even the hope that it leads to a unique theory in which consistently all constants of nature could be derived from the dynamics. At the end of the 1980's five versions of consistent string theories were known to exist, named Type I, Type IIA, Type IIB, Heterotic-O(32), Heterotic-E8 × E8. I would need much more space to describe how they are defined. This can be done properly only by a further reformulation of the string model, called the Green-Schwarz model, in which ten-dimensional space-time supersymmetry–and not only two-dimensional superconformal symmetry–becomes manifest as a symmetry of the action.

String theorists still have to solve two essential problems: First, how to get rid of the "superfluous" dimensions, and second, where exactly in the rich string dynamics does the Standard Model reside. The first point of course gave rise to a revival to the idea of compactification, investigated in Kaluza-Klein models. In Sect. 8.2.2, I described how the topology of the compactified space relates to the Yang-Mills content of the full theory, and I illustrated this for a sphere and a torus in 2D. You can imagine that a six-dimensional space can be compactified in manifold ways. The Heterotic-E8 × E8 superstring indeed has solutions which come close to our known world with its four dimensions, Standard Model gauge group, three families ..., this is good news. The bad news is that superstrings do have many other solutions as well. Guesses are around 10^{500} – 10^{1500} depending on the way of how six of the 10 dimensions are compactified. And the theory is also unable to explain why the 10D-space compactifies as $M^4 \times C^6$. (It is an astounding state of affairs that string theories in 10 dimensions vibrate with symmetries, but that after compactification we may get into worlds with no symmetries at all).

After the so-called first string revolution in 1984, string theory took off, and became an own chapter in mathematical physics. If you open a string theory textbook you will certainly be impressed by the wealth of mathematical notions in use, far beyond the repertoire of an "ordinary" theoretician. Not only was the systematics of string compactifications (based on so-called Calabi-Yau manifolds and orbifolds) developed, but one discovered numerous interrelations among the five string models. These are essentially duality relations in the sense that the strong coupling regime of one theory is essentially the same as the weak coupling of another with respect to a certain compactification. String theorists experienced a second string revolution in 1995, when E. Witten announced his M-theory program by which he conjectured that the five seemingly different string theories are just manifestations of a single theory. Here, even 11-dimensional supergravity made its re-appearance, in that for instance dimensional reduction on a circle yields the low energy limit of Type IIA string theory. Unlike strings, the fundamental objects in M-theory are believed to be extended in not only one but two spatial directions. Such objects are called supermembranes.

One of the pillars in quantum field theory is locality; the fields represent point particles. This is relinquished in string theory in favor of objects that are extended in one dimension. But if one is willing to give up locality, it becomes reasonable to ask "Why stop at one dimension?". Indeed, already in the 1930th there were ideas of describing elementary particles as three-dimensional objects. These reappeared in another form in string theory with the notion of D-branes. Previously I mentioned the Dirichlet boundary conditions for open strings. These violate translational invariance … unless the points to which the string is fixed is a dynamical object itself. Thus a D-brane is simply an object upon an open string ends. It need not be two-dimensional membranes but can be of lower and higher dimension of any shape and size. As stressed by J. Polchinski, the introduction of D-branes is not optional, rather they are needed for a consistent and complete description of string theory. Interestingly, the requirement of supersymmetry puts limitations on the allowed dimensionality of supermembranes. Thus, where for instance in 4D there are 0-, 1-, 2-, 3- dimensional superbranes, in 8D only a 4-brane can exist. A complete "brane scan" (space-time dimension *vs.* allowed types of p-branes) is known [139]. Additionally, it is known that the five superstring theories host different types of branes.

8.4.3 Gauge/Gravity Duality Conjecture

Let me take it directly from the person whose name is associated with this duality. J. Maldacena writes in the abstract of [358]: "The gauge/gravity duality is an equality between two theories: On one side we have a quantum field theory in d spacetime dimensions. On the other side we have a gravity theory on a d+1 dimensional spacetime that has an asymptotic boundary which is d dimensional. It is also sometimes called AdS/CFT, because the simplest examples involve anti-deSitter spaces and conformal field theories. It is often called gauge-string duality. This is because the gravity theories are string theories and the quantum field theories are gauge theories. It is also referred to as 'holography' because one is describing a $d+1$ dimensional gravity theory in terms of a lower dimensional system, in a way that is reminiscent of an optical hologram which stores a three dimensional image on a two dimensional photographic plate. It is called a 'conjecture', but by now there is a lot of evidence that it is correct." (Despite the last sentence, one should be aware that there are hard-core field theorists who doubt that the conjecture makes sense.) In essence this conjecture tells us that theories of gravity on curved space-times are equivalent to quantum field theories living on the boundaries of those spaces. The example Maldacena refers to is *AdS/CFT*, realized for instance by a five-dimensional anti-deSitter space AdS_5 (whose boundary is a four-dimensional spacetime) and the maximally-symmetric ($N = 4$) super Yang-Mills theory which is a conformal field theory with proper UV behavior.

This subsection ended the unification chapter with speculations which are mathematically very advanced, but phenomenologically hard to swallow. Nevertheless I admire the braveness of those actively involved in the super-string-brane-gravity industry. Their ideas, although fitting eminently well with the spirit of this monograph, reach far beyond its scope. The most important publications on supergravity, supermembranes and M-theory are reprinted in [140]. Extra dimensions, M-theory and warped dimensions already found its way into the popular literature such as [241], [433]. Since the string community did not fulfill the glorious promises made during both its revolutions, there are also detractors of " string theory" like [476], [566], calling it a "Theory of Anything".

Chapter 9
Conclusion

... the unreasonable effectiveness of symmetry in our understanding of Nature.

9.1 Symmetries: The Road to Reality

9.1.1 Symmetry: The Golden Thread

The Growing Awareness of Symmetries

Although even the classical physical disciplines, namely Newton's mechanics and Maxwell's electrodynamics carried already the germ of symmetry, it was the 20th century which deserves the title "Century of Symmetries". Among the most influential was Einstein's special relativity, in which among other things the Galilei group of symmetry transformations was replaced by the transformations of the Poincaré group. Because of the unitarity properties of quantum physics it was the irreducible unitary representations of the Poincaré group that led E. Wigner to the definition of particles, and their classification in terms of mass and spin. Symmetry requirements and the necessity of renormalizability restrict the type of fields and their possible interactions (in terms of the action functional). After roughly half a century, the revival of Weyl's gauge argument for electrodynamics in the form of its non-Abelian generalization by Yang and Mills arrived on the scene. They and the forerunners of the quark color-charge model paved the way for our understanding of what we today call the 'internal symmetries'. It took, however, a detour from symmetries to the understanding of symmetry breaking before one was able to completely formulate the 'standard model' of particle physics. As of today, this experimentally extraordinarily well-verified model stands side by side with another very successful theory,

This citation from A. Zee [578] alludes to the title of an article by E. Wigner on the "Unreasonable Effectiveness of Mathematics in the Natural Sciences"

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6_9,

namely general relativity.[1] Thus it is quite natural that many researchers are trying to overcome conceptual weaknesses within the standard model (e.g. through ideas of grand unification), and there is also a large community of physicists working on amalgamating quantum field theory and general relativity in the form of a not-yet-achieved quantum gravity theory. And, given the success of symmetry arguments, it comes as no surprise that many of these attempts again employ such arguments (e.g. supersymmetry and Kaluza-Klein models).

Symmetries as Encoded in Actions

There is no doubt that symmetry is a mathematical notion and that it is intimately related to group theory. It seems that symmetry in fundamental physics must be formulated on the level of functional integrals to cope with the quantum aspect of nature. Except for some rare situations (called "anomalous") the symmetry of a theory can be formulated just as well on the level of its action functional. Indeed, for reasons essentially unknown, fundamental physics can be derived from appropriate actions by Hamilton's principle of least action. The starting point of all symmetry considerations in this book are actions

$$S = \int d^D x \, \mathcal{L}(Q, \partial Q, \partial\partial Q),$$

which are functionals defined in terms of Lagrangians $\mathcal{L}$ depending on fields Q^α and their first (and possibly, if admissible, second) derivatives. The variation of an action is denoted as

$$\delta S = \frac{\delta S}{\delta Q^\alpha} \delta Q^\alpha = S_{,\alpha} \delta Q^\alpha.$$

From the principle of least action, the field equations of the theory are given by $S_{,\alpha} = 0$.

If, under infinitesimal variations of the fields, characterized by (infinitesimal) ϵ_r and written as

$$\delta_\epsilon Q^\alpha(x) = \mathcal{R}^\alpha_r(Q) \cdot \epsilon_r = \sum_r \int d^D y \, R^\alpha_r(x, y) \epsilon_r(x),$$

the variation of the action vanishes ($\delta_\epsilon S = S_{,\alpha} \delta_\epsilon Q^\alpha \equiv 0$), we are dealing with a variational symmetry. This type of symmetries plays a pivotal role in fundamental physics, and therefore predominates in this book. When a variational symmetry is present, the field equations are not independent, but instead comply with

$$S_{,\alpha} \mathcal{R}^\alpha_r = \int d^D x \, S_{,\alpha} R^\alpha_r(x, y) \equiv 0, \tag{9.1}$$

[1] Observe that one speaks of the standard "model" of particle physics and the "theory" of general relativity. This is perhaps unreflected by many, but it reveals that GR, simply due its elegance–derived from one of the greatest symmetries imaginable–is far more respected as an ultimate theory than the standard model, because as discussed above, the latter has many unexplained features and more than a dozen parameters of unknown origin.

the so-called Noether identities. The requirement that symmetry transformations constitute a group can be formulated in terms of the commutator of two infinitesimal symmetry transformations. This results in consistency conditions on the functionals/operators $\mathcal{R}_r^\alpha$. In some theories, the commutators close with structure constants; then the symmetry group is a finite Lie group, the constants being the structure constants of the Lie algebra associated to the symmetry group. These are the genuine Noether symmetries, appearing for instance as the internal symmetries of particle physics. In other cases, and typically for reparametrization-invariant theories (the most prominent example being general relativity), the commutator algebra has spacetime-dependent structure functions. It even may be the case that the algebra only closes on-shell, as for many supersymmetric models.

In terms of the Lagrangian $\mathcal{L}$, a symmetry variation δ_ϵ leads to the identity

$$0 \equiv \Delta_\epsilon \mathcal{L} = [\mathcal{L}]_\alpha \bar{\delta}_\epsilon Q^\alpha + \partial_\mu J_\epsilon^\mu. \tag{9.2}$$

Here the variation $\bar{\delta}_\epsilon Q^\alpha = \delta_\epsilon Q^\alpha - Q^\alpha_{,\mu}\delta_\epsilon x^\mu$ allows for a variation of the coordinates as well. The expressions $[\mathcal{L}]_\alpha$ are the Euler-Lagrange derivatives of the Lagrange density $\mathcal{L}$. The on-shell $([\mathcal{L}]_\alpha \doteq 0)$ conserved currents J_ϵ^μ are in general

$$J_\epsilon^\mu \doteq \frac{\partial \mathcal{L}}{\partial(\partial_\mu Q^\alpha)} \bar{\delta}_\epsilon Q^\alpha + \mathcal{L}\delta_\epsilon x^\mu - \Sigma_\epsilon^\mu.$$

Here Σ_ϵ^μ is a term that potentially appears in the current if $\mathcal{L}$ is quasi-invariant, that is, if $\delta_\epsilon \mathcal{L} = \partial_\mu \Sigma_\epsilon^\mu$.

In the case of a global symmetry (the ϵ being constants), the Noether current can be expanded as $J_\epsilon^\mu = \epsilon^r j_r^\mu$. Under suitable boundary conditions on the fields, the on-shell conserved currents imply conserved charges

$$C_r = \int d^{D-1}x\, j_r^0.$$

In the case of local symmetries, the currents in general contain terms with derivatives of the ϵ's, $J_\epsilon^\mu = \epsilon^r j_r^\mu + \epsilon^r_{,\nu} k_r^{\mu\nu}$, say. Now (9.2) leads to the chain of Klein-Noether identities from which ensue the Noether identities (9.1) in the form $\mathcal{R}_r^\alpha \cdot [\mathcal{L}]_\alpha \equiv 0$. Formally, charges

$$C\,[\epsilon] = \int k_r^{i0} \epsilon^r dS_i$$

can be defined in terms of the superpotentials $k_r^{\mu\nu}\epsilon^r$ integrated over a (D-2) dimensional hypersurface.

Quantum Physics Requires a Bridge from Symmetries to Group Representations

According to a theorem by E. Wigner, symmetry transformations in quantum physics are represented either by linear and unitary, or anti-linear and anti-unitary operators.

This implies that the group structure of Noether symmetry transformations and the algebra of its Lie generators is largely mirrored in quantum mechanics:

$$g(\epsilon) = e^{i\epsilon_r X^r} \quad \longleftrightarrow \quad \hat{U}_\epsilon = e^{i\epsilon_r \hat{X}^r}.$$

Here, on the left-hand side is the relation between a group element near the identity and its Lie algebra generators X^r. The right-hand side of the expression shows the relation between the (anti)unitary symmetry operator $\hat{U}_\epsilon$ and the Hermitean operator $\hat{X}^r$. The group operation is mirrored in the unitary representation as

$$\hat{U}(g) \cdot \hat{U}(g') = e^{i\phi(g,g')}\hat{U}(g \circ g').$$

The phase $\phi(g, g')$ occurs because quantum states are rays in Hilbert spaces. This necessitates the investigation of ray or projective representations of symmetry groups.

Symmetries Restrict Actions and Field Equations

The representation theory of the Lorentz group leads to a classification of possible field variants with half-integer and integer spin. The lowest irreducible unitary representations are realized by scalar fields, Weyl fields, and vector fields (corresponding to spin 0, $^1\!/_2$, 1). Actions formed from these fields and their derivatives must be Lorentz scalars. Together with the requirement of renormalizability (through the dimensional renormalization criterium), the allowed action terms are largely fixed.

From Global to Local Symmetries

Any theory with global symmetries can be rendered locally symmetric with a procedure first noted by H. Weyl for electrodynamics: Replace the derivatives ∂ for the fields in the globally-symmetric theory by a covariant derivative D defined in terms of a connection:

$$\partial \to D = \partial + \text{Connection}.$$

In the case of Yang-Mills theories, the connections correspond to the vector potentials A, in the case of metric general relativity these are the Levi-Civita connections Γ; and in the case of tetrad gravity, we have the spin connections ω. The connections are interpreted as the gauge fields of the locally-symmetric theory. The second Noether theorem together with the requirements of Lorentz invariance and of renormalizability restrict the form of the kinetic term for the gauge fields and as well the form of possible interaction terms. This seems like magic, but it works: Imagine a world with only spin-0 and spin-$^1\!/_2$ fields (scalars and spinors) and suppose that it exhibits global phase symmetry. For promoting these to local symmetries, we need a spin-1 field (a Yang-Mills field). In relativistic field theories the spin (0, $^1\!/_2$, 1) system is Lorentz invariant. Enhancing this global theory to a local one, we need to introduce further connections and by this spin-2 fields (representing gravity). The spin-2 fields also come into existence if the globally supersymmetric (0, $^1\!/_2$, 1) system is made locally supersymmetric. This then necessitates spin-$^3\!/_2$ fields (called gravitinos).

Spacetime Symmetry Groups

H. Minkowski's discovery of the merits of the Poincaré group compared to the Galilei group opened further avenues. For one thing, the investigation of the invariance groups of spacetimes possessing the minimal set of requirements from special relativity (the kinematical groups) revealed a network of algebras and their mutual relations. Among these is the de Sitter group, of which the Poincaré group is a contraction. Another generalization of the Poincaré group is the conformal group.

Theories of Gravitation: Highly Symmetric–not Renormalizable?

Einsteins' s general relativity (GR) is the theory of gravitation *per se*, in the sense that it describes all known gravitational processes with very high precision. The Hilbert-Einstein action is invariant under general coordinate transformations constituting the 'highly' symmetric diffeomorphism group. It seems that GR is the simplest among all theories showing this symmetry. However, GR is not an acceptable theory if one takes seriously the requirement for renormalizability. On the other hand, waiving this requirement gives one the freedom to investigate a plethora of theories which extend GR, as for instance the so-called Poincaré gauge theories.

9.1.2 The "Weltgesetze" and Their Symmetries

The title of this subsection is–of course–chosen deliberately. It mimics the wording and the spirit of the Göttingen mathematicians around D. Hilbert and F. Klein (E. Noether also served them by contributing material for their arguments), who aimed at laying the 'foundation of physics' through a mathematical inspired motivation. But around this time (the early 1920's), Hilbert could not have been aware that nature holds more in store than just gravitation and electromagnetism.

The previous subsection summarized the power of variational symmetries in general. Although, as spelled out before, the field variants and their appearance in action functionals is largely restricted, these generic arguments cannot tell us for instance how many different interactions there are and which symmetry groups are preferred by nature.

Action

In this book, depending on the context, different splittings of the "world action" $S_{\mathcal{W}}$ were used:

$$S_{\mathcal{W}} = S_G + S_{\mathcal{M}} = S_G + (S_F + S_B) = S_G + (S_F + (S_{GB} + S_H)).$$

Here S_G denotes the pure gravitational action term, while $S_{\mathcal{M}}$ stands for the matter and energy coupled to the gravitational field. The further division of $S_{\mathcal{M}}$ into S_F and S_B indicates the fermionic action part (including its coupling to the gauge bosons via the covariant derivative) and the kinetic gauge boson part S_{GB} plus the scalar

(Higgs) part including the Yukawa coupling to the fermions. More explicitly, as of today, the world action is given by

$$S_{\mathcal{W}} = \int_M d^4x (\det e) \left\{ \frac{1}{2\kappa} R[e] + \frac{1}{g^2} F^2 + \overline{\psi} \not{D} \psi + (D\phi)^2 + V(\phi) + \overline{\psi} \phi \psi \right\}. \quad (9.3)$$

This action is invariant under diffeomorphisms, local Lorentz transformations and the local internal symmetries $\mathbf{SU(3)} \times \mathbf{SU(2)} \times \mathbf{U(1)}$. In the shorthand notation used in (9.3), the first term is the gravitational part in terms of tetrads e (with the Newton gravitational constant $G = \kappa/(8\pi)$). The second term contains the Yang-Mills field strength F for the gauge bosons (photon, weakons, and gluons), with g denoting collectively the three field strength of the internal product group. The third term describes the dynamics of the fermion fields ψ (electron and neutrino families, quarks) via the covariant derivatives D. The next two terms are due to scalar Higgs field ϕ, and the last term is the Yukawa term responsible for mass generation. The covariant derivatives D are to be built with respect to the spin- and the gauge-connections.

Symmetries

The world action (9.3) has many global and local symmetries. The global symmetries for e.g. the standard model of particle physics are hidden in the detailed form of the fermionic terms in (9.3). The local symmetries are

- Diffeomorphism, mappings of the spacetime manifold M upon itself:

$$d : M \to M, \qquad x \mapsto d(x)\,.$$

Under this mapping the physical fields behave as

$$\varphi(x) \mapsto \varphi(d(x)) \quad (9.4a)$$
$$\psi(x) \mapsto \psi(d(x)) \quad (9.4b)$$
$$A^a_\mu(x) \mapsto \mathcal{K}^\varrho_\mu\, A^a_\varrho(d(x)) \quad (9.4c)$$
$$e^I_\mu(x) \mapsto \mathcal{K}^\varrho_\mu\, e^I_\varrho(d(x)) \quad (9.4d)$$
$$\omega^I_{\mu\,J}(x) \mapsto \mathcal{K}^\varrho_\mu\, \omega^I_{\varrho\,J}(d(x)) \quad (9.4e)$$
$$g_{\mu\nu}(x) \mapsto \mathcal{K}^\varrho_\mu \mathcal{K}^\sigma_\nu\, g_{\varrho\sigma}(d(x)) \quad (9.4f)$$
$$\Gamma'^\mu{}_{\nu\lambda} \mapsto \mathcal{J}^\mu_\varrho \mathcal{K}^\sigma_\nu \mathcal{K}^\tau_\lambda\, \Gamma^\rho{}_{\sigma\tau} + \mathcal{J}^\mu_\varrho \partial_\nu \mathcal{K}^\varrho_\lambda, \quad (9.4g)$$

where

$$\mathcal{J}^\varrho_\mu(x) =: \frac{\partial d^\varrho(x)}{\partial x^\mu} \qquad \mathcal{K} = \mathcal{J}^{-1}.$$

Diffeomorphism "see" only spacetime indices. Therefore the scalar and the spinor fields transform as Riemann scalars, the Yang-Mills gauge fields, the tetrads and the spin connections transform as covariant Riemann vectors, and the metric as a Riemann tensor. The Levi-Civita connection $\Gamma^\mu{}_{\nu\lambda}$ transforms inhomogeneously.

- Lorentz-Transformations, mappings from the manifold to the Lorentz group:

$$\lambda : M \to \mathbf{SO(3,1)}.$$

Under this mapping the physical fields behave as

$$\varphi(x) \mapsto \varphi(x) \tag{9.5a}$$
$$\psi(x) \mapsto S(\lambda(x))\psi(x) \tag{9.5b}$$
$$A^a_\mu(x) \mapsto A^a_\mu(x) \tag{9.5c}$$
$$e^I_\mu(x) \mapsto \Lambda^I{}_J(x)e^J_\mu(x) \tag{9.5d}$$
$$\omega_\mu{}^I{}_J(x) \mapsto \Lambda^I{}_K(x)\omega_\mu{}^K{}_L(x)\Lambda^L{}_J(x) + \Lambda^I{}_K(x)\partial_\mu\Lambda^K{}_J(x). \tag{9.5e}$$

Lorentz transformations "see" only the tangent space indices: Scalar fields and internal gauge fields belong to the trivial representation, spinor fields to the spinor representation $S(\lambda)$, and the tetrads to the fundamental representation $\Lambda^I{}_J$ of the Lorentz group. The spin connection $\omega_\mu{}^I{}_J$ transforms inhomogeneously.
- G-Transformations, internal symmetry transformations with respect to the gauge group $\mathbf{G}$ are mappings

$$g : M \to \mathbf{G}$$

under which

$$\varphi(x) \mapsto D_\varphi(g(x))\varphi(x) \tag{9.6a}$$
$$\psi(x) \mapsto D_\psi(g(x))\psi(x) \tag{9.6b}$$
$$A^a_\mu(x) \mapsto \mathrm{Ad}(g(x))A^a_\mu(x) + g(x)\partial_\mu g^{-1}(x) \tag{9.6c}$$
$$e^I_\mu(x) \mapsto e^I_\mu(x) \tag{9.6d}$$
$$\omega_\mu{}^I{}_J(x) \mapsto \omega_\mu{}^I{}_J(x). \tag{9.6e}$$

Local gauge transformations "see" only the group indices: D_φ und D_ψ are representations of $\mathbf{G}$ to which the φ and the ψ belong, and $\mathrm{Ad}(g)$ is the adjoined representation of $\mathbf{G}$. The gauge group connection A^a_μ transforms inhomogeneously.

In their local (or infinitesimal) form, the previous mappings become

1. For $d(x) : x'^\mu = x^\mu + \epsilon^\mu(x)$, we have for the Riemann scalars

$$\delta_\epsilon\varphi = 0, \qquad \delta_\epsilon\psi = 0,$$

for the Riemann vector fields $G_\mu{}^\alpha := \{A^\alpha_\mu, e^J_\mu, \omega_\mu{}^I{}_J\}$

$$\delta_\epsilon G^\alpha_\mu = -G^A_\nu\ \epsilon^\nu_{,\mu}, \tag{9.7}$$

and for the metric tensor

$$\delta_\epsilon g_{\mu\nu} = -g_{\varrho\nu}\ \epsilon^{\varrho}_{,\mu} - g_{\mu\varrho}\ \epsilon^{\varrho}_{,\nu}. \tag{9.8}$$

The Levi-Civita connection transforms infinitesimally as

$$\delta_\epsilon \Gamma^{\mu}{}_{\nu\lambda} = \Gamma^{\varrho}{}_{\nu\lambda}\,\epsilon^{\mu}_{,\varrho} - \Gamma^{\mu}{}_{\varrho\lambda}\,\epsilon^{\varrho}_{,\nu} - \Gamma^{\mu}{}_{\nu\varrho}\,\epsilon^{\varrho}_{,\lambda} + \epsilon^{\mu}_{,\nu\lambda}. \tag{9.9}$$

2. If the Lorentz transformations near the identity are denoted as $\Lambda^I_K = \delta^I_K + \lambda^I{}_K(x)$ with $\lambda^I{}_K + \lambda_K^I = 0$ one derives the infinitesimal form

$$\delta_\lambda \varphi = 0, \qquad \delta_\lambda A^a_\mu = 0 \tag{9.10}$$

$$\delta_\lambda \psi = \frac{1}{4}\Sigma^{IJ}\lambda_{IJ}\,\psi \tag{9.11}$$

$$\delta_\lambda e^I_\mu = \lambda^I{}_K e^K_\mu \tag{9.12}$$

$$\delta_\lambda \omega_\mu{}^I{}_J = \lambda^I{}_{J\,,\,\mu} + \omega_\mu{}^I{}_K\lambda^K{}_J + \omega_\mu{}^K{}_J\lambda^I{}_K \equiv D_\mu[\omega^I{}_J]\lambda^I{}_J, \tag{9.13}$$

where $D_\mu[\omega^I{}_J]$ is the covariant derivative with respect to the spin connection.

3. Infinitesimal gauge transformations near the group identity are $g(x) = 1 + i\theta^a(x)T^a$. Specifically for a Yang-Mills gauge field, we obtain the infinitesimal form

$$\delta_\theta A^a_\mu = \partial_\mu\theta^a + f^{abc}\theta^b A^c_\mu = D_\mu[A^a]\theta^a.$$

Here, f^{abc} are the structure constants in the Lie algebra generated by the T^a and $D_\mu[A^a]$ is the covariant derivative with respect to the gauge connection. A coupling constant g is absorbed in the definition of the gauge connection and therefore appears as g^{-2} multiplying the F^2-term in the Lagrangian. In particular, from this expression one can recognize the complete analogy of the gauge group **G** and the Lorentz group through the similarity to (9.13) .

9.1.3 History of Symmetry Considerations

In the history of symmetry considerations, I see three lines of evolution which merged in the last century. These are related to (1) our understanding of symmetries with respect to space and time, (2) to so-called internal symmetries discovered in parallel with the advances in particle physics, and (3) to the abstract notion of groups, a notion perfectly suited to stating precisely what symmetries are all about.

- space-time symmetries
  In the understanding of space-time symmetries, we can observe an evolution going from Aristotle to Galilei to Einstein. Aristotelian space-time can be thought of as $\mathcal{A} = \mathbb{E}^3 \times \mathbb{E}^1$, that is the Cartesian product of a three-dimensional Euclidean space

with a one-dimensional line, representing time. Space and time are completely unrelated. The isometry group is $(\mathbf{SO}(\mathbf{3}) \ltimes \mathbb{R}^3) \times \mathbb{R}^1$, the first factors standing for rotations and translations in $\mathbb{E}^3$, and the latter describing translations in $\mathbb{E}^1$. Due to the principle of Galilean relativity, the dynamical laws of classical mechanics are unchanged when transformed uniformly. This gives Galilean space-time the structure of a fibre bundle with base space $\mathbb{E}^1$ and fibre $\mathbb{E}^3$, locally isomorphic to $\mathcal{A}$. Galilean space-time has the Galilei group as its isometry group: $\mathbf{Gal} = (\mathbf{SO}(\mathbf{3}) \ltimes \mathbb{R}^3) \times \mathbb{R}^1 \times \mathbb{R}^1$. Here, the additional factor represents Galilei boosts. With the advent of special relativity we now understand spacetime to be a Minkowski space with the Poincaré group as isometry group. The different space-time notions according to Aristoteles, Galilei and Einstein are lucidly explained in Chap. 17 of [410].

- internal symmetries
  The internal symmetries are entirely a story of the twentieth century. Here, two strands came together in the 1970's: One of these arose from the observation of W. Heisenberg that protons and neutrons share so many properties that it seems reasonable to consider them as two faces of the same coin–or in the language of symmetries, they can be transformed into each other as representations of $\mathbf{SU}(\mathbf{2})$. After the observation of still more "elementary particles", this symmetry was extended to $\mathbf{SU}(\mathbf{3})$. A breakthrough was the perception by M. Gell-Mann and by G. Zweig that a more appropriate assignment of the multiplets should refer to the quark constituents of the elementary particles instead of the particles themselves, previously thought to be elementary. The other strand originates from considerations by H. Weyl about allowing scale (or as he called it "gauge") transformations as local symmetries for a gravitational theory, and the later insight that this idea of replacing a global symmetry by a local one necessarily requires the existence of gauge fields. This all culminated in what is today called 'Yang-Mills gauge theory', together with a good understanding of how to associate three families of three colored quarks to an $\mathbf{SU}(\mathbf{3})$ gauge theory. An excellent account about the different avenues leading to the accepted picture of gauge theories is to be found in [429]. That monograph also contains the pioneering original articles.
- group theory
  As we saw in previous chapters, the mathematics of symmetry is group theory–as it also becomes evident from the length of Appendix A. Group theory has its origin in the analysis of solvability of polynomial equations by radicals. However, group theory did not enjoy popularity in physics before the 1930's and 1940's. It was quantum theory which forced group theory upon physics, starting from the book of H. Weyl [550], but even more influenced by E.P. Wigner's book [553]. These works initiated a plethora of articles, many of them of a difficult mathematical character. Therefore, group theory was not easily accepted in the physics community. Opponents used the word *Gruppenpest*, which means the plague or pestilence of group theory. Sometimes this is attributed to W. Pauli,[2] but in an interview with T.S. Kuhn within the oral history project of the American Institute of Physics,

[2] W. Pauli was known to be harsh, so this sounds like Pauli. In any case, at an International Congress of Philosophers, held in Zürich in 1954, he stated: "It seems likely to me, that the reach

Wigner was not sure whether it was not actually E. Schrödinger who coined the term.[3] Indeed, some eminent physicists (like E. Condon, G. Shortley, J. Slater) were proud to derive results in atomic physics without group-theoretical methods. Specifically, there was a widespread skepticism about the use of representation theory. This changed after M. Gell-Mann and Y. Ne'eman successfully used $\mathbf{SU}(\mathbf{3})$ representations to classify strongly interacting particles in the early 1960's. In the 1970's there was a real boom in the literature, a search in a Berlin central library catalogue lists nearly fifty books. An internet search with "symmetries particle physics" yields more than 1,500,000 entries.

As a result of the prominent role played by E. Wigner in applying group theory and symmetry considerations, for which he received the Nobel price in physics in 1963, the "Group Theory and Fundamental Physics Foundation" has conferred the Wigner medal since 1978 "to recognize outstanding contributions to the understanding of physics through group theory".

Today, group theory is pivotal to fundamental physics.[4] A. Zee [579] illustrates this nicely as follows: "I am reminded of the story of a visitor to a joke-tellers" convention. One comedian would shout out "C-46!"and the other comedians would laugh appreciatively. Someone else would stand up and shout out "S-5!" and everyone would laugh. The puzzled visitor asked what was going on, and his friend explained: "All possible jokes, not counting minor variations of course, have been classified and numbered, and we know all of them by heart." Similarly, all groups have been classified and numbered by mathematicians. When a physicist comes to my office, she might mutter $\mathrm{SO}(3)$ or $\mathrm{E}(6)$, and I would nod appreciatively. The physicist is telling me her guess which group Nature uses in Her design. Incidentally, I should give the punch line to the story. Finally, a comedian got up and shouted "G-6!" and everyone really cracked up. The visitor asked why this particular joke was so extraordinarily funny, and his friend replied, "Oh, that is Joe Schmo; he is so dumb that he doesn't know there is no such thing as "G-6!" Similarly, if I were to mention $\mathrm{G}(6)$ in a seminar, my colleagues would raise their eyebrows in surprise! Anyhow, all groups have been classified and named."

Timeline

This timeline only covers the essential milestones concerning symmetries in fundamental physics, but not those for instance in atomic physics, molecular physics, many-body physics, or crystallography. Since it is a timeline, it should not come as a

---

of the mathematical group concept in physics is not yet fully exploited." The essential role which "symmetries" were playing in the work of Pauli is investigated in [222].

[3] http://www.aip.org/history/ohilist/4965.html

[4] Perhaps I can be allowed to make a personal remark: Between leaving the Gymnasium in the early 1960's and starting to study physics, I was stationed with the German army near Göttingen. In an antiquarian bookshop, I found a small brochure about group theory, which in its introduction remarked that group theory is of a certain importance to physics. I became curious, and I'm still curious…

surprise that developments in different areas, such as particle physics and spacetime physics, are collocated together.

1632 Galileo Galilei realizes that it is impossible to tell whether a given frame is moving uniformly or is at rest. Although he does not state this precisely as a quantitative law, he illustrates it by the example of ships moving relative to each other or to the coast line [208]. Today, this is called Galilei invariance, and is expressed by the symmetry of the laws of classical mechanics with respect to inertial systems.

1632 I. Newton lays down in his *Principia* [385] the three basic laws for classical mechanics. The first law states the conservation of momentum due to the homogeneity of space, which–as we know today–is related to translational invariance.

1832 The genesis of group theory is the investigation of algebraic solutions of polynomial equations by E. Galois. He is the first to coin the name "group", defining this algebraic structure, and using normal subgroups for coset decomposition. Galois finds the relation between the algebraic solutions and the structure of permutation groups associated with the equation. This work was not published until 1846. In fact, N. Abel already had a proof that quintic equations cannot be solved by radicals. Galois shows why they can't. It took more then forty years until in **1878**, A. Cayley formulates the abstract group concept as we use it today.

1862 J.C. Maxwell succeeds in combining the accumulated knowledge about electric and magnetic phenomena into a set of coupled differential equations for the electric and the magnetic field and their sources. The equations contain a constant with the dimension of a velocity, which had to be identified with the vacuum velocity of light. They predict electromagnetic waves propagating at the speed of light; it was suggested that the medium in which these waves propagate is an "aether". One should definitely also mention that it was M. Faraday who developed our current understanding of fields.

1868 Initiated by C.F. Gauß, B. Riemann in his inaugural lecture "*Uber die Hypothesen, welche der Geometrie zu Grunde liegen*" [442] constitutes those differential geometric concepts which are indispensable in modern theoretical physics, among others the definition of manifolds, metric, curvature, embedding.

1872 F. Klein presents an outline ("program") of his future research to the faculty at the Erlangen University [319]. This famous *Erlanger Programm* had emerged from his discussions with S. Lie. The program describes an application of group theory to geometry: geometries are classified by invariance groups and groups define geometries.

1887 W. Killing completes the classification of all simple Lie groups. This classification is later streamlined by Eli Cartan in his dissertation from **1894** (and became known under his name).

1888–1893 S. Lie and F. Engel publish '*Theorie der Transformationsgruppen*' [343]. Interestingly enough, Lie's mostly known contributions to group theory and algebra originates in finding all symmetries of differential equations.

1886–1904 In this period, the scientific community becomes pregnant, giving birth to special relativity. In **1887** A. Michelson and E. Morley try to detect a possible effect of Earth's motion through the "aether wind" on the speed of light. However, no difference between the speed of light moving in two perpendicular directions could be detected. In **1892** G. FitzGerald and independently H. Lorentz aim to resolve this by proposing that objects contract along their direction of motion due to the pressure of the "aether wind". This was later called the FitzGerald-Lorentz contraction. In **1897** J. Larmor introduced the transformations which make up what is now called the Lorentz group, as those transformations which leave Maxwell's equations invariant.

1905 This is the year called *annus mirabulis*, because A. Einstein publishes four essential papers concerning critical open issues of that time. In the context of symmetries, it is the publication '*Zur Elektrodynamik bewegter Körper*' [151], in which, starting from a set of assumptions on space and time, he directly derives the Lorentz transformations. Today, this field is called special relativity.

1906 F. G. Frobenius and I. Schur publish '*Über die reellen Darstellungen der endlichen Gruppen*' in which they prove pivotal results in the theory of group representations (specifically the theory of characters).

1908 H. Minkowski gives his famous speech '*Raum und Zeit*' in Cologne. His four-vector notation arising from the notion of Minkowski (sic!) spacetime and the use of what are now called Lorentz tensors renders previous calculations of relativistic mechanics and electrodynamics more transparent. Furthermore, these notations and techniques also paved Einstein's road towards general relativity.

1915, 25st Nov. A. Einstein presents his ultimate field equations of gravity to the Prussian Academy of Science [152]. The exact date is important here for two reasons. For one, Einstein gave another version two weeks before this date (in which the assumption was made that general coordinate transformations should be volume preserving, e.g. $\sqrt{|g|} = 1$.) The other reason is that D. Hilbert presented an action functional for gravity in Göttingen a few days earlier. This led to some dispute in the history of science community of who discovered general relativity; for the latest point of view, cf. [100]. Interestingly enough Hilbert never claimed priority; instead he said: "*Jeder Strassenjunge in unserem mathematischen Göttingen versteht mehr von der vierdimensionalen Geometrie als Einstein. Aber trotzdem hat Einstein die Sache gemacht, und nicht die grossen Mathematiker*".[5]

[5] "Every urchin in our mathematical Göttingen knows more about four-dimensional geometry than Einstein. Nevertheless, it was Einstein who did the work and not the great mathematicians.

1918 E. Noether describes how symmetries and conserved quantities are related in the case of a finite continuous symmetry group, uncovering the deep connection between abstract variational symmetries of actions, and well-known–but perhaps at that time inexplicable–conservation laws. Furthermore–and this has drastic consequences for our theories of fundamental physics–Noether also shows that for infinite continuous groups, the field equations are not independent. Her work did not immediately penetrate into physics, but remained unnoticed for nearly half a century; see [327].

1918 H. Weyl claims to have found a classical unified field theory for gravitation and electromagnetism. It is based on local scale transformations on the metric, and assumes that the full theory is invariant with respect to these transformations. Weyl called it an *Eichtheorie*. Although Einstein, with whom Weyl corresponded seriously in this matter was at first very enthusiastic, he in the course of this correspondence acquired the opinion that the model proposed by Weyl had too many unphysical length and time scalings. Finally he accepted it for publication [548], however with an addendum: "Except for the agreement with reality, it is in any case a grand intellectual achievement."

1921 A. Einstein presents to the Prussian Academy a paper by Th. Kaluza with the title '*Zum Unitätsproblem der Physik*', based on a manuscript which he had received already two years before. Kaluza was able to formally derive from five-dimensional general relativity the gravitational and electromagnetic theory in four dimensions assuming specifically that the field quantities do not depend on the fifth dimension. This unification approach was rediscovered, reinterpreted and extended by Oscar Klein in **1926** [321]. He, among other things, makes Kaluza's cylinder condition precise and points out the fact that the extra dimension is of the order of the Planck length.

1922 E. Schrödinger reformulates Weyl's 1918 proposal in terms of phase transformations instead of rescalings, and in **1927** F. London identifies the phase transformations as transformations of the Schrödinger wave function; see the reprinted articles and their placement within a science-history context in [429]. This gave rise to Weyl's classic *Elektron und Gravitation* in **1929**, which can be seen as the origin of our current comprehension of gauge theories: turning global symmetries of a theory into local symmetries by introducing appropriate gauge fields.

1924 S. Bose introduces what is now called Bose-Einstein statistics for photons. In **1925** there is a generalization by A. Einstein to those particles or quanta we now call 'bosons'. Their many-quanta states are invariant under all permutations.
1926 Fermi-Dirac statistics is introduced (by E. Fermi and P.A. Dirac—*sic*) for those particles we now call 'fermions'. Their many-particle states change sign under odd permutations. This statement includes the Pauli exclusion principle.

1928 P. A. M. Dirac proposes a relativistic wave equation for spin-$^1/_2$ particles. Its meaning is at first not clear because of negative energies in the theory. Only in **1931** was it interpreted as a theory of the electron and its antiparticle. From a mathematical

point of view, the first appearance of spinors shows the importance of half-integer representations of the Lorentz group.
H. Weyl publishes '*Gruppentheorie und Quantenmechanik*', a book with alternating sections on quantum physics and on group theory.

1930 In order to save the otherwise well-established conservation law for energy and momentum, W. Pauli hypothesizes the existence of a new particle, the 'neutrino' (name coined by E. Fermi after the discovery of the neutron in **1932**). To me, this is the first instance of a successful practice in the 20th century: Hypothesize a new particle each time a symmetry seems to be broken. The neutrino (or more precise– given the six neutrinos we know today: the $\bar{\nu}_e$) was detected in **1956** by a group led by C. L. Cowan and F. Reines using the inverse beta decay.

1930 L. Rosenfeld pioneers the phase-space formulation of locally symmetric field theories. His ambitious attempt to canonically quantize the Einstein-Maxwell-Dirac system was probably ahead in time. It is revisited only twenty years later.

1931 E. Wigner investigates the consequences of quantum-mechanical principles for symmetry operators and presents what is now known as Wigner's theorem in his book '*Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*'. With this theorem, he also discloses the specific role of time-reversal symmetry.

1932 W. Heisenberg proposes a symmetry between protons and neutrons in nuclear physics; later called isospin symmetry.
C. D. Anderson finds the positron in a cosmic-ray experiment, the first of the antiparticles to be detected (and as predicted by Dirac in 1928).

1933 E. Fermi formulates a Lagrangian for β-decay, including an interaction term with proton, neutron, electron, and neutrino fields.

1934 W. Pauli and V. Weisskopf show that the "merger" of special relativity and quantum physics necessarily leads to the notion of antiparticles, if one identifies the Noether charges in a complex scalar field theory with electrical charges [408].

1935 H. Yukawa proposes to interpret the short range of the strong interaction by the exchange of massive particles (the π-meson).

1937 The discovery of the muon (by C. A. Anderson) in cosmic rays, at first believed to be the π-meson predicted by Yukawa, is the first manifestation for the existence of families in the particle spectrum. The associated μ neutrino was discovered only in **1962** (and Yukawa's hypothetical particle was in the meantime discovered in **1947**).

1939 Given the prominent role of the Poincaré group in classical relativistic physics, E. Wigner investigates its unitary representations [555]. The results allow the classification of all relativistic wave equations and the transformation properties of quantum fields.

1940 W. Pauli proves the spin-statistics theorem: Particles with half-integer spin–the fermions–obey Fermi-Dirac statistics, those with integer spin–the bosons–obey Bose-Einstein statistics.

around 1950 The Hamiltonian formulation of field theories with local symmetries (already in the focus of Rosenberg's article of 1930) receives a further impetus by work of P. A. M. Dirac and by the Syracuse group around P. G. Bergmann.

1952 G. C. Wick, A. Wightman and E. P. Wigner introduce the notion of superselection rules, related to the ray representation of a symmetry group.

1954 C. N. Yang and R. L. Mills generalize the approach of H. Weyl for deriving the electromagnetic interactions from locally gauging a **U(1)**-symmetry to a **SU(2)** local isospin transformations [570], the first example of what are now called non-Abelian gauge theories. Astoundingly, this type of theory is preferred by nature as proven by the success of the standard model of particle physics. And in **1956**, R. Utiyama formulates gauge theories for arbitrary semisimple Lie groups [513], including **SO(3,1)**.

1954–55 The CPT theorem, involving space inversion (P), charge conjugation (C) and time reversal (T) is proven by G. Lüders and W. Pauli: In a local relativistic quantum field theory the product CPT of these transformations is always a symmetry.

1956–57 Parity violation in weak interaction processes is envisioned by T. D. Lee and C. N. Yang and verified experimentally by C. -S. Wu *et al.*. Lee and Yang shared the 1957 Nobel prize for physics. This led to the $(V\text{-}A)$ model of weak interactions due to R. Marshak and C. G. Sudarshan, as well as to R. Feynman and M. Gell-Mann.

1958 P. A. M. Dirac and the ADM-trio (R. Arnowitt, S. Deser, and C. W. Misner) present Hamiltonian versions of general relativity. A (3+1)-decomposition of the metric clarifies the role of the three-metric and the extrinsic curvature as phase-space variables, and the lapse and shift functions as multipliers of first-class constraints.

1959–61 J. Goldstone and Y. Nambu suggest that the ground state (vacuum) of quantum field theory may lack the full symmetry of the Hamiltonian, and that massless excitations (later to be called Goldstone bosons) must accompany this 'spontaneous symmetry breaking'.

1961 M. Gell-Mann and independently Y. Ne'eman suggest **SU(3)** as a symmetry for strong interactions (the "Eightfold Way"). This includes the isospin symmetry in a larger symmetry group which also acts on the strangeness quantum number. Baryons and mesons are arranged in multiplets of **SU(3)**.

1964 M. Gell-Mann and G. Zweig propose independently a new, deeper level of quantum particles, namely the quarks, to account for the observed **SU**(**3**) symmetry in particle physics. The up, down and strange quarks successfully describe the pattern of hadrons known thus far. This early quark model even predicted the existence of

a further resonance state with specific quantum numbers and in a certain mass range. The discovery of the Ω^- then paved the way for the establishment of the quark model. It is found that for spontaneously-broken gauge symmetries there are no Goldstone bosons but instead massive vector mesons. Today, this is called the Higgs phenomenon after P. Higgs, one of the discoverers of this mechanism.
The CP breaking part of the weak interaction is established experimentally by a group around J. W. Cronin and V. L. Fitch.
J. Bjorken and S. Glashow speculate about the existence of a further quark (the 'charmed quark') in order to restore lepton-quark symmetry. Up to that year, only three quarks were known, but already four leptons (electron, muon and their neutrinos). A quantitative prediction (GIM mechanism) is made by S. Glashow, J. Iliopoulos and L. Maiani in **1970**.

around 1970 A theory of electroweak unification emerges from work of S. Glashow, A. Salam and S. Weinberg. It combines the Yang-Mills field theory aspect with the idea of spontaneous symmetry breaking.

1970–1971 G. t Hooft proves the renormalizability of Yang-Mill field theories.

1972 The strong interaction is conceived as a Yang-Mills theory by M. Gell-Mann, H. Fritzsch and H. Leutwyler.

1973 Discovery of asymptotic freedom of non-Abelian gauge theories by D. Gross, F. Wilczek, and independently by D. Politzer.

1973 A Grand Unified Model based on $\mathbf{SU}(\mathbf{5})$, the minimal group embracing the standard model symmetry group, is devised by H. Georgi and S. Glashow. Among other things, it allows the calculations of the lifetime of the proton. In **1999** the results from the Super-Kamiokande detector point to a lifetime longer than the predicted one.

1973 Supersymmetry becomes a topic of interest. After some initial ideas by Yu.A. Golfand and E. P. Likhtman (1970) and by D. V. Volkov and V. P. Akulov (1972), the first supersymmetric field theory is formulated by J. Wess and B. Zumino [546]. The superalgebras compatible with the principles of quantum field theory is classified in **1975** [251]. This ignites an inflation of 'super'physics, culminating in supergravity and the superstring model.
M. Kobayashi and T. Maskawa predict the existence of at least one additional family of quarks and leptons in order to explain CP violation within the standard model.

1973–74 The essential features of the currently-accepted standard model of particle physics are established. A fifth and a sixth quark ('bottom' and 'top') are found in **1977** and **1995**, respectively. The third member in the lepton family, the tauon, was indirectly discovered in the seventies, its associated neutrino only in **2000**.

1974 Two experimental groups, one at SLAC headed by B. Richter, and the other one headed by S. Ting from MIT, detect a resonance state dubbed J/ψ, and identified as “charmonium”, a bound state of a charm quark and its antiparticle. This discovery became known as the “November Revolution” in particle physics, since it marks the breakthrough of the quark model.

1975 C. Becchi, A. Rouet, R. Stora and independently I. V. Tyutin discover a Grassmann-odd symmetry on the Faddeev-Popov path integral of gauge theories. This leads to a conceptually new approach for quantizing gauge symmetries, an approach completed by I. Batalin and G. Vilkovisky in the early **1980**s for all theories with variational symmetries.

1976 S. A. Hojman, K. Kuchař, and C. Teitelboim reveal the universal structure of the Hamiltonians and the constraint algebras of diffeomorphism-invariant theories.

1984 M. Green and J. Schwarz prove that the superstring model is free of anomalies. This “first string revolution” initiates an intense boom in ‘string theory’. A “second string revolution” took place in 1996 with E. Witten’s proposal of an M-theory.

1987 A. Ashtekar finds a “new” Hamiltonian formulation of general relativity which has greater similarities to Yang-Mills gauge theories than the ADM version. This influences heavily the development of loop quantum gravity.

1998 In attempts to measure the ‘deceleration’ rate of the cosmic evolution, it is found that the universe is expanding at an increasing rate. This phenomenon is parameterized in cosmological models by a ‘dark energy’. Although the origin of the dark energy is not settled, this might very well be a signal that the Minkowski geometry assumed in special relativity needs to be replaced by a de Sitter geometry.

2012 CERN announces the discovery of a scalar boson with a mass of 126 GeV, having properties of the Higgs boson.

Virtual History …

Virtual (or: counterfactual) history asks the question: What would have happened later, if… Thus imagine another course of events. I like the following virtual his/herstory: Imagine that Einstein discovered special relativity prior to Maxwell’s discovery of the basic equations of the electrodynamic field. This could very well have been the case, since–as explained in Subsect. 3.2.1–special relativity can be derived without recourse to the velocity of light. With Einstein’s insight, it would have been possible for Dirac to discover “his” equation. Observing that this equation is invariant with respect to global $\mathbf{U(1)}$-transformations and rendering this into a local symmetry would have lead to the introduction of a gauge field A_μ—and eventually to Maxwells theory! As a matter of fact, this change in the logic underlies a formulation of H. Weyl: “… that electromagnetism is an accompanying phenomenon of the material wave field…”[6] [551]

[6] “*… daß das elektrische Feld ein notwendiges Begleitphänomen … des materiellen durch ψ dargestellten Wellenfeldes ist.*”

…and Future Insights

It would of course be brilliant to be able to predict the future of symmetry-related topics in fundamental physics; although, as we have seen in the chapter on 'symmetries of everything' (Chap. 8), theoreticians are prepared for new kinds of symmetries while they are awaiting further experimental results from the LHC at Geneva. The recent discovery of a scalar boson–maybe the Higgs boson in its minimal manifestation in the standard model–gave rise to a lot of excitement even in the public, but it left those working on fundamental physics rather unexcited, and rather uneasy instead. This is because, as discussed in Sect. 6.6., the standard model cannot be the ultimate story in any case. No one would be surprised if in the new hitherto unexplored energy range of the LHC, phenomena are discovered which are unprecedented by any kind of imaginings of the theoreticians–even when they are inspired by symmetry considerations. And perhaps, after a century of symmetries in fundamental physics, a century of symmetry breaking will follow. My personal bets? The speculative supersymmetry will be established (off course broken by a large margin) and the extremely successful Poincaré symmetry will turn out to be broken (by a very slight margin).

9.2 Are Symmetries a Principle of Nature?

This is a question which leads out of physics. Indeed, it is a philosophical question.

9.2.1 . . . and Other Philosophical Questions

Physics and philosophy have a common origin. Or, to be more precise, originally there was no such discipline called physics. It seems not unreasonable to consider Galileo Galilei to have been the very first physicist–at least taking him as a representative of a new discipline with its aspirations to perform systematic measurements on natural phenomena and to cast the results into mathematically-formulated laws. Nevertheless, the common origin of physics and philosophy has never been lost. Some of the outstanding physicists from the past five centuries also came onto the scene as philosophers. On the other hand, philosophers have turned eagerly to topics raised by developments in physics at the beginning of the last century: These are especially the interpretational problems of quantum physics and the ever-changing role of space and time in the relativity theories.

Also, many results from fundamental physics give rise to questions which commonly are considered to be a domain of philosophers. S. Weinberg formulated a harsh critique towards philosophy (in fact entitled "Against Philosophy") [532], in the sense that philosophy has in his opinion never really helped physics, and sometimes even hindered its development. He concedes that, "Physicists do of course carry around with them a working philosophy", but he is rather critical of the role of philosophers as being beneficial to physicists. On the other hand, a representative of the philosophical discipline summarized the current situation as “Philosophers of science have barely scratched the surface of the topic of laws, symmetries, and symmetry breaking.” [144]. From my point of view, there are certain questions which belong undoubtedly to the domain of physics, other questions certainly belonging to philosophy, and there are questions on the border between these disciplines. Nevertheless, being myself a physicist, I claim that a philosopher should not ignore the experimental results and the established theories in physics, and thus has to listen to the physicists–and not the other way round. Weinberg writes “In our hunt for the final theory, physicists are more like hounds than hawks; we have become good at sniffing around on the ground for traces of the beauty we expect in the laws of nature, but we do not seem to be able to see the path to the truth from the heights of philosophy.” However, after reaching the goal, one might very well ask which path is the logically more satisfying one.

Indeed many physics/philosophy questions are triggered by the issue of symmetries [57], aside from the one in the title of this section: What is a law of Nature[7]? Are there laws of Nature[8]? Are there principles of Nature? Is a principle a more fundamental notion than a law of Nature? Which principles can serve to explain our current knowledge? Can physics be completely understood by principles? What is the minimal set of principles? … Does Nature know its laws and principles?

In the introduction I made E. P. Wigner witness of the important role of symmetries by citing “... if we knew all the laws of nature, or the ultimate Law of nature, the invariance properties of these laws would not furnish us new information.” This is taken from [558], where in a broader context Wigner is reasoning on a hierarchy of knowledge, which means the “... progression from events to laws of nature, and from laws of nature to symmetry or invariance principles.” So what he has in mind is that natural laws are a means to describe a set of data, and principles are a means to reflect about natural laws. And he sees “... a great similarity between the relation of the laws of nature to the events on one hand, and the relation of symmetry principles to the laws of nature on the other.” This similarity is expressed in the next sentence: “... if we had a complete knowledge of all events in the world ... there would be no use for the laws of physics, ... if we knew all the laws of nature ... the invariance properties of these laws would not furnish us new information.” At a first glance these considerations sound plausible, but they immediately raise questions: Is this threefold hierarchy events–natural laws–principles actually all there is, and can the levels strictly be defined. Are laws of nature merely a clever bookkeeping device about events ? What’s about their predictive power ? Can one be satisfied about the state of physics by merely throwing a bunch of natural laws on the floor ?

---

[7] In this philosophical context I will write Nature, World, Universe with capitals.

[8] Despite the discussion in the literature of the philosophy of science, I do not distinguish between laws of Nature and laws of physics.

No surprise, there is already some tradition in the philosophical community if it comes to the question of laws and principles of physics [368, 518] and to the topic of symmetries [57]; the latter reference actually contains a very readable introduction to the theme.

Let us start our philosophical excursion with a topic which in philosophy is usually treated by questions related to what means progress in science and theory reduction.

9.2.2 Symmetries and the Unification of Physics

Undoubtedly, symmetries allow one to make seemingly unrelated entities look more uniform and unified. Therefore, many researchers directly relate symmetries to the vision of unification of physics. This is meanwhile already a theme of popular science literature, such as ‘Dreams of a Final Theory’ by S. Weinberg [532], or ‘New theories of everything–the quest for ultimate explanation’ by J. D. Barrow [31]. What could possibly be meant by ‘unified physics’?

Looking into a school textbook, we see physics as nothing but a collection of seemingly diverse disciplines, described by headings as mechanics, acoustics, fluid dynamics, heat, electricity, magnetism, optics, atomic physics, nuclear physics, elementary particle physics…Where is ‘unification’ in light of this variety of branches of physics? And where is a notion of ‘unification’, considering the serious and ever-increasing specialization of physicists[9]? This is a point which has also been raised by Frieder–you remember my skipper Frieder, whom I introduced in the Preface. Nevertheless, he objects to my distinction of electricity and magnetism–no wonder, he is an electrical engineer. Of course he is right, since from the days of Faraday and Maxwell, the two disciplines have been unified into electromagnetism. And the last century saw even more ‘unification’ of originally separate disciplines.

Although he did not state it in this context, A. Einstein believed in unification. After reflecting on the fragmentation of physics into separate disciplines, he devotes his considerations to the foundations of physics [161]: “…, from the beginning there was always present the attempt to find a unifying theoretical basis for all these single sciences, consisting of a minimum of concepts and fundamental relationships, from which all concepts and relationships of the single disciplines might be derived by logical process. This is what we mean by the search for a foundation of the whole of physics. The confident belief that this ultimate goal may be reached is the chief source of the passionate devotion which has always animated the researcher”.[10] In the same

---

[9] I, for instance, do not even understand the abstract of the PhD thesis of my niece Vera, which somehow deals with microelectronics–but I’m not sure because I don’t understand her!

[10] “*Daneben bestand aber auch von Anfang an das Bestreben, eine alle diese Einzelwissenschaften vereinigende theoretische Basis aufzufinden, bestehend aus einem Minimum von Begriffen und Grundrelationen, aus denen alle Begriffe der Einzelwissenschaften sich auf rein logischem Wege sollen ableiten lassen. Es ist das Suchen nach einem Fundament der ganzen Physik. Das Vertrauen in die Erreichbarkeit dieses höchsten Zieles ist eine Hauptquelle der leidenschaftlichen Hingabe, welche die Forscher von jeher beseelt hat.*”

article, Einstein takes a tour through the history of physics in order to see lines of convergence towards a unifying foundation. And of course, we educated physicists are able to look beyond the school textbooks. We know that the falling apple and the movement of the planets have the same cause. We furthermore know that the dynamics of fluids and acoustics can be formulated in the language of Newtonian mechanics. The merging of electric and magnetic phenomena embraces optics, as well. Thermodynamic phenomena are explained in terms of ‘statistical mechanics’. And the very fact that burning substances glow opens a window onto quantum physics, situated at the interface of thermodynamics and electromagnetism.

All this is standard knowledge to anyone with a background in natural sciences. And that person will agree that there is a line of convergence in our understanding of seemingly unrelated phenomena, for the moment vulgarized as ‘unification’. In this book, unification became more abstract: Phenomena on quite different energy scales obey laws with a comparable appearance. By this I mean that gravitational, electromagnetic, weak and strong interactions seem to share similar concepts such as gauge bosons, connections, field strengths …

What Made Unification Possible?

There are various meanings of ‘unification’, so let me try out this one: A theory T’ enlarges another theory T in the sense that T is a special case or an approximation of T’. Examples are (1) classical mechanics as limiting case of relativistic mechanics, which itself is a special case of special relativity; (2) Newton’s law of gravitation as a limiting case of general relativity; (3) wave optics as a special case of electrodynamics, and geometrical optics as an approximation to wave optics; (4) Newtonian mechanic as an approximation of quantum mechanics, and classical physics (mechanics and electrodynamics) as approximations to QED; (5) QED as a low-energy limit of the electroweak theory. Each unification step was also accompanied by the creation of new concepts and the shift of the meaning of earlier ones.

Unification also meant the discovery of the importance of certain ‘principles’. (Here, I use quotation marks; later in this chapter I try to be more precise about principles.) We saw that in the unification of the fundamental interactions, these principles are related to the existence of an action, its variational symmetries and the renormalization requirement for a quantum field theory.

With increasing unification of physics, we observe an increasing conceptual simplification. Here again, action functionals play a central role, since these–in a condensed form–entail the symmetries of the theory, the dynamical equations, and also the quantum field-theory version. To an astounding degree, the action functionals for the electromagnetic, the weak and the strong interactions are structurally identical (being of the Yang-Mills type with specific symmetry groups). Simplification goes hand in hand with the beauty or aesthetics of the description.

The successful unification of physics over the course of time makes one wonder of whether an ultimate unification can possibly be achieved, and how far we are on the way towards this 'unity of physics'. Be aware that at this very moment we are pondering about unity of fundamental physics. But what is possibly meant by the 'unity of physics'?

'Unity of Physics' in Terms of Metaphors

Metaphor "World formula on a dinner napkin":
In a strict sense, this metaphor is meaningless since you can always write a formula $W = 0$ onto a napkin (or if you prefer–onto your T-shirt). The point is to explain what is meant by the expression W, how its components map to the real world, and whether they are amenable to experimental observation. This explanation definitely does not fit on the T-shirt of a theoretical-physical layman. Nevertheless, it is possible to render the vision of a brief world formula more precise: if we take for instance the already short formula (9.3)[11] it is obvious that it has too many terms. I even cheated, since the term which is quadratic in the field strength hides the fact that there are three interactions in the Yang-Mills sector. It was discussed in Sect. 7.6.3 why–at least for general relativity–the first and the second terms are structurally different. And finally, the potential term $V(\phi)$ and the Yukawa coupling term $\overline{\psi}\phi\psi$ look rather constructed and patched onto the rest of the action.

In this book, I have described speculative attempts to arrive at a more compact expression for the action. Grand unified theories as well Kaluza-Klein theories and superstring models aim, among other things, at overcoming the separate terms for the four interactions.

Metaphor "Axiomatization":
This metaphor is inspired by the *Elements* of Euclid, an axiomatization of geometry. The idea is to logically deduce from a set of independent and consistent set of axioms, in which the basic entities of a domain are encoded, theorems which hold in this domain. You know that Euclid introduced with his fifth axiom a postulate referring to parallel lines, and there were various attempts to show that this can be derived from the other four axioms. Failures to do so led to the insight that besides Euclid's geometry other geometries can be imagined. I mention this because it led to a reflection about the axiomatic approach, especially by D. Hilbert who then formulated another axiomatization of geometry.[12] Hilbert also reflected about the 'Foundations of Physics' in the spirit of axiomatization [465], [99], [440]. Among the twenty-three unsolved problems which Hilbert presented to the International Mathematical Congress in Paris in 1900, the sixth problem asked whether physics can be axiomatized. Indeed you might understand Newton's *Principia* as an axiomatization of mechanics. This was criticized in the 19th century (L. Boltzmann, E. Mach, H. Hertz). The early 20th century saw the axiomatization of special relativity (D. Hilbert, C. Carathéodory, H. Reichenbach), thermodynamics (E. Mach,

[11] Your T-shirt could have the patch $\delta S_{\mathcal{W}} = 0$.

[12] The axiomatization doctrine received a setback by Gödel's incompleteness theorem from 1931. It is not clear whether this also applies at the same time to physics, because the 'axiomatization' schemes may differ from those in mathematics.

C. Carathéodory), quantum mechanics (D. Hilbert, P. Jordan, J. von Neumann), quantum field theory (A. Wightman, R. Haag, D. Kastler), general relativity (M. Bunge, and J. Ehlers, F. Pirani, A. Schild).

From these examples, you may notice that attempts at axiomatization were more or less successful for those branches of physics which show a certain degree of maturity, in that basic postulates can be identified and distinguished from heuristic principles. There is a difference between axiomatization in mathematics and physics, in any case. While in mathematics, the objects and relations in the axioms and the ensuing theorems are in the first instance void of any interpretation in the real world, an axiomatic physics needs to model reality. (Here I'm using 'reality' in a broad sense, not distinguishing common-sense reality from scientific reality.) One might prefer to base an axiomatization on observables only, or also give non-observables an ontological status. Thus for instance there are both axiomatizations of field theories avoiding the idea of a 'field' from the outset, and others taking 'fields' as part of their basic furniture. And given that in many domains of physics, principles seem to be at work, there are attempts to start from these as kind of axioms in the sense of premisses. Thus an axiomatic special relativity can be founded on the relativity principle and the constancy of the vacuum velocity of light[13] instead of on the 3+1 Minkowski geometry of space-time.

Metaphor "God had no choice":
… as a possible answer to the question "*Hätte Gott die Welt auch anders erschaffen können?*" which Einstein posed in the 1940's to his co-worker E. Straus [195]. Einstein's question may sound presumptuous, but it could be posed because of his experience with general relativity, in which only a few presuppositions yield a unique theory. If God did not have a choice, the full theory would be highly self-contained and self-consistent: Not only the structure of the natural laws, but also the type and number of fields, any natural constants, and even the dimension and the signature of spacetime would be fixed. This is not only a unity of physics, but 'uniqueness' of physics [466].

To me it seems that we are furthest from this option, because before contemplating about His ultimate design, there are many more down-to-earth questions to answer. We do not know why **U(1)**, **SU(2)**, **SU(3)** or other internal symmetry groups were chosen by Him. And we have only some idea about why the Poincaré group is present in symmetry transformations in spacetime. He might have opted for other kinematical groups; and maybe He has, in favoring the more general de Sitter group. Furthermore, to describe our Universe we need four spacetime dimensions, three particle families, four interactions, around twenty parameters in the standard model of particle physics, and a handful of parameters in the concordance model of cosmology. We must exclude right-handed leptons from the game (by brute force) and enforce

[13] Another axiomatization could be based on the 'relativity without light' derivation of Lorentz transformations as given in Subsect. (3.2.1).

CP violation on the model. By which deeper reason was He compelled to create two long-range and two short-range forces, and which rules of the Genesis rendered gravitation so much weaker than the other interactions?

There was a short period in physics during which theorists thought to have found a theory–namely the 1986 string theory–which made the 'God had no choice' option viable. But soon it was realized that uniqueness could not be achieved. This gave rise to the attitude that maybe 'God had a few choices', with the option that various world designs are conceivable–one of them being our particular World. But this attitude is not really satisfactory in that "few" needs to be quantified (possibly ranging from two to a dozen to ...), and one needs to worry about the fate of the other Worlds. To me, it seems a curiosity in the history of science that the attitude in string theory completely changed, in that many of its adepts are now willing to believe in a very large "string theory landscape" of admissible worlds [466]; more about this below.

In a special issue on the achievements and non-achievements of string theory in the journal 'Foundations of Physics', L. Smolin, one of the heavy critics behind anthropomorphic principles, states: "I should stress that the landscape problem, while it arose in string theory, is likely to be there whatever the fundamental unification of physics turns out to be. There is no evidence from any approach to quantum gravity that mathematical consistency or the existence of a low energy limit restricts the matter content of a theory."[477]

To conclude these reflections about God's design of the World, let me adopt this instructive Genesis version from A. Zee [578]: "How did the mind of the Creator work when He designed our cosmos? Did He say "Let there be light!" or did He say, "Let symmetries be local! "?" Zee offers still another Genesis: "When He supposedly said, "Let there be light!" perhaps He actually said, "Let there be an SU(5) Yang-Mills theory with all its gauge bosons, let the symmetry be broken down spontaneously, and let all but one of the remaining massless gauge bosons be sold into infrared slavery. That one last gauge boson is my favorite. Let him rush forth to illuminate all of my creations."

Although it looks quite appealing, the 'World formula on a napkin' is the weakest form of unity in physics. Even if such a comprehensible and compact 'World formula' would be found, one would not necessarily understand the principles behind it (if they exist at all). This brings forth the question of why the formula is at it is; or how one can understand that this formula is at it is. And if an axiomatization is found, it leaves open the question of how many different worlds are possible according to the axioms. The strongest metaphor is 'God had no choice', supposing that the axioms/principles are so stringent that only one world, namely our own, can exist.

"The End of Physics"

The successes in fundamental physics in the second half of the last century led theorists to serious speculations about the end of physics. These were particularly expressed in the supergravity community (you saw in Subsect. 8.3.4, how stringent the most symmetric supergravity model is). Some remarkable early achievements of the superstring model in the late 1980's and 1990's made some researchers in fundamental physics nearly euphoric about the imminent complete unity of physics. Even prominent physicists had "Dreams of a Final Theory" and thought about the "Ultimate Laws of Nature" (S. Weinberg). S. Hawking affirmed in his inaugural lecture as Lucasian professor in Cambridge in April 1980 the question "Is the end in sight for theoretical physics?" with "…some reasons for a cautious optimism that we will see a complete theory within the lifespan of some sitting in this audience" [256]. Nearly twenty years later, he changed his point of view radically and gave a talk about "Gödel and the End of Physics" [257]. The title is a bit deceptive, because as a matter of fact he argues with Gödel's theorems that a physical theory as a formal system, is either inconsistent or incomplete. His essay concludes with the statement "Some people will be very disappointed if there is not an ultimate theory that can be formulated as a finite number of principles. I used to belong to that camp, but I have changed my mind. I'm now glad that our search for understanding will never come to an end, and that we will always have the challenge of new discovery."

Although not expressed explicitly, those speculating about the ultimate theory had in mind the 'God had no choice' metaphor. But, what if 'God had a choice'? Then He was able to create arbitrary and completely different universes, differing not only by their natural constants, but also by spacetime dimension, types of interactions, different kinds of symmetries, etc. Interestingly enough, we can observe these days that more and more serious scientists are dealing with the idea of 'multiverses' [76]. And also interestingly, many of them became friendly with the idea that our universe is nothing but an arbitrary point in the space of multiverses [76] as a result of their frustration at the failure of the string model which once raised the hope that it would yield the 'God had no choice' theory.

The end of physics[14] has been proclaimed to be near several times in the history of science. A prominent example is the story told about M. Planck, who was discouraged from studying physics by his high school teacher. And although heavily impressed by the unification progress based on the emphasis of symmetry arguments, I personally doubt that we ever will see a termination of physics. It is less than hundred years ago that we became aware of the strong and the weak interactions. And we still seem to understand only five percent of the matter/energy content of our universe. How can we be sure that we already know all essential facts for formulating the final theory.

The very idea of effective field theories (as explained in Subsect. 5.6.) allows to describe our current knowledge within a seemingly closed theory/model, although in reality–whatever that means–the currently-favored theory is part of an ongoing and maybe never-ending story. According to this comprehension, physics can be understood as a chain of effective field theories, the chain elements building the bridge between various levels of description. So one may ask whether in moving up in energies (or, synonymously going to smaller and smaller distances) this chain

[14] The physicist and philosopher C.F. von Weizsäcker defined the end of physics as that state of affairs where all questions that one directs to Nature are outside of physics [543].

terminates. It might for instance terminate at the Planck energy, where our notions of space and time and the laws of relativistic quantum field theory break down, or because a genuine (renormalizable) fundamental theory is reached. But, we must also be prepared that it will not terminate at all, and this would bury all aspirations of a final or unified theory.

9.2.3 Laws of Nature and Principles of Physics

In reflecting about the laws of Nature we need to deal according to [368] with three aspects: "First, whether the observed regularities observed in physics are based on strict 'laws of nature' that hold rigorously and without any exception. Second, … comparing this concept with invariance principles, causality principles, teleological principles and means of predicting future events. Finally … the ambitious and intricate third question, why the laws of nature hold." (quoted from the book jacket). Another good source on the current physics-philosophy debate on the nature and the knowability of laws of Nature is [542].

Apparently, we are living in a world in which laws of physics can be mathematically formulated, experimentally tested and applied for technical purposes. These laws may be grossly heuristic and basically incomplete, like Ohm's law. On the other hand, they may be encoded in a full-fledged 'theory' like the standard model of particle physics.[15] Thus asking questions about the nature of "Laws of Nature", I indeed have in mind networks of laws derived in a theory or constituting the theory. Furthermore, if the laws of fundamental physics–along the lines of this book–are identical to the laws of Nature, these could become identical to the World action.

There are at least three different notions of laws of Nature, namely whether they are ontological (part of Nature), epistemological (helping us understand Nature), or methodological (helping to predict the behavior of Nature). The weakest version is the methodological one, and this is definitely true. The physical laws which mankind has found during the last centuries are valuable tools to describe natural phenomena or to construct and control technical artifacts. The epistemological notion is the basis for our strong belief that the laws we discover in our terrestrial laboratories are universal and hold in our solar system, our galaxy and in the visible universe. Otherwise, one would have to write different physics books for different spatial-temporal regions. But what about regions we can never observe? Previously, I mentioned that these days there is growing attention to the multiverse hypothesis, which comes in different forms. In its extreme form, each of the single universes needs its own distinct physics books. The strongest version of a law of Nature is the ontological one; this would allow a statement such as "Nature knows its laws".

[15] One should keep in mind that many predictions of the standard model are not derived completely analytically, but are very often approximate results derived perturbatively, partly based on computer simulations within lattice gauge theories.

Principles of Physics

The laws themselves or even the network of laws encoded in the Lagrangian of a theory do not tell us why they are what they are. In this book it was stressed that Lagrangians are what they are because we imposed certain principles. Thus the principles seem more fundamental than the laws–maybe they are the laws of Nature themselves.[16] This gives rise to further questions, namely as to what characterizes a principle, how many principles are needed, whether there is a relation or hierarchy among the principles. I did not really count, but I suspect that in all of physics one finds some dozens of principles. In the context of this book, in order to cover the scope of variational symmetries, we need at least

- The existence of an action: P. Ramond starts his book on field theory [431] with an hallelujah on the action, which I slightly modify: It is a beautiful and awe-inspiring fact that all the fundamental laws of physics can be understood in terms of one mathematical construct called the action. It yields the dynamical equations of the classical theory, and analysis of its invariances leads to conserved currents. It encodes the phase-space description of the theory with its canonical structure. In addition, the action acquires its full importance in quantum physics: it provides a clear and elegant language to effect the transition between classical and quantum physics through the use of Feynman path integrals.
- A gauge principle: Any global variational symmetry needs to be localized in order to explain the force-field variants observed in experiments.
- 'Minimal coupling' between matter fields and force fields. This is a means to restrict possible interaction terms in a Lagrangian. Technically, it leads to a simple recipe to replace an ordinary derivative by a covariant one.
- A preference for second order dynamical equations, as a necessary condition for a well-posed initial value problem, such that Cauchy data allow to integrate the equations.
- Quantum principles, as for example the identification of observables with Hermitean operators in a Hilbert space;
- Locality and renormalizability in order to be able to compute measurable quantities in a quantum field theory. Again this restricts otherwise possible interaction terms in the action.
- A sort of simplicity: scalars, spinors, vectors, …, and no fields with spin>2.

These are already seven principles,[17] and behind any of these might lurk further principles–principles yet to be found: The principle that Nature needs to be described by actions (to make quantization possible [279], because dynamics is computation [506], . . .), a principle by which Nature necessarily is quantized, a principle by which fields with spin larger than two cannot interact, …

[16] The interested reader is referred to an article by D. Deutsch [117] on an examination of the slogan "law without law" brought into the world by J. A. Wheeler.

[17] There could be another more general principle as a guiding theme in the fundamental interactions: Physics should not depend on how we describe it.

Merely possessing a long list with principles is of course unsatisfactory. One would like to find out how they are related, and specifically how the shortlist of independent principles looks. How short might this list be? In the axiomatization metaphor of the unity of physics, there should be only few principles. Why is this so? In an interview that Einstein gave in 1954, and which was published only after his death, he stated “I want to know how God created this world. I’m not interested in this or that phenomenon, in the spectrum of this or that element. I want to know His thoughts; the rest are details.” (E. Salaman, ‘A Talk With Einstein’, The Listener 54 (1955), pp. 370–371). This fits well with a previous quote “…a minimum of concepts and fundamental relationships, from which all concepts and relationships of the single disciplines might be derived by logical process.” from Einstein’s considerations concerning the foundation of theoretical physics. In a certain sense, ‘His thoughts’ is a place-holder for the ‘basic principles of physics’, and if He is reasonably well organized He created the World with, I suppose, less than two handfuls of thoughts.

Amazingly, the laws and principles that were found in the last century very often explained or even predicted the existence of new particles. Here, I am thinking of the neutrino, antiparticles, the charm, top and bottom quark, etc. In some cases, the theories allow for objects which have not (yet?) been observed. Where are for instance the other “beasts” from Wigner’s classification (e.g. the tachyons)? There must be another principle which excludes their existence. And in some cases, the principles allow a wider range of theories than those experimentally verified thus far. Which for instance is the principle by which Nature does not act according to the largest kinematical group, namely in the de Sitter spacetime?

Are Symmetries a Basic Principle?

Let me start with some quotations, mentioned already in the Introduction: “Today we realize that symmetry principles … dictate the form of the laws of nature.” (D. Gross, [247]). “To a remarkable degree, our present detailed theories of elementary particle interaction can be understood deductively, as consequence of symmetry principles.” (S. Weinberg, [534]). “…profound guiding principles are statements of symmetry.” (F. Wilczek, [559]). These well-respected physicists obviously believe in the ‘principle of symmetry’.

Pretty much as for any law/principle of Nature, we should have in mind three different notions of symmetries, and try to understand whether symmetries are ontological (part of Nature), epistomological (helping us understand Nature), or methodological (rendering certain calculations easier).

The weakest version is again the methodological one, and this is definitely true: Symmetries are welcome because they ease the description and analysis of physical systems. One of the typical abilities of a physicist is to abstract away from a problem all irrelevant details. In this process, one uses symmetry arguments (“Let’s assume that the cow is a sphere.”). In some cases only symmetric configurations are accessible for an analytic description (example: Schwarzschild black holes and isotropic static metrics, Friedmann cosmology and spatially isotropic and homogeneous spaces).

But in the previous quotations, it is not the methodological version of symmetry, but mainly the epistemological one as in "can be understood deductively, as consequence of symmetry principles." At another point, S. Weinberg states "Symmetry principles have moved to a new level of importance in this century and especially in the last few decades: there are symmetry principles that dictate the very existence of all the known forces of nature." Here I'm not sure who dictates, it could be Nature itself. And this would be the ontological aspect of symmetries. Weinberg outlines in Chap. VI of his "Dreams Of A Final Theory" that (1) the symmetry among different reference frames necessitates the existence of gravitation, (2) the symmetry between an electron, its neutrino, and an up- and down-quark requires the electroweak interaction and by this the existence of the weakons (Photon, W- and Z-bosons), (3) the symmetry among colored quarks enforces the existence of the strong interactions and the gluons, respectively.

Which are the symmetries of Nature (ontological notion of symmetries)–and not just symmetries in how we describe Nature in terms of actions and fields? Here are three different opinions, ranging somehow from ontological to methodological:
(from D. Gross [245]): "The primary lesson of physics in this century is that the secret of nature is symmetry. The most advanced form of symmetries we have understood are local symmetries–general coordinate invariance and gauge symmetry. In contrast we do not believe that global symmetries are fundamental. Most global symmetries are approximate and even those that, so far, have shown no sign of being broken, like baryon number and perhaps even CPT, are likely to be broken. They seem to be simply accidental features of low energy physics. Gauge symmetry, however is never really broken–it is only hidden by the asymmetric macroscopic state we live in. At high temperature or pressure gauge symmetry will always be restored."
(from D. Giulini [221]: "We need to distinguish between proper physical symmetries which map physical states or histories to other, physically distinguishable states or histories, and gauge symmetries which map one state or history to a physically indistinguishable one (redundancy of description). In field theories these two notions of "symmetry" often appear in a combined form: a proper normal subgroup Gau $\subset$ Sym represents gauge symmetries, whereas the quotient, Phys $=$ Sym/Gau corresponds to proper physical symmetries."
(from K. Hinterbichler [276]): "…gauge symmetry is a complete sham. It represents nothing more than a redundancy of description. We can take any theory and make it a gauge theory by introducing redundant variables. Conversely, given any gauge theory, we can always eliminate the gauge symmetry by eliminating the redundant degrees of freedom. The catch is that removing the redundancy is not always a smart thing to do. For example, in Maxwell electromagnetism it is impossible to remove the redundancy and at the same time preserve manifest Lorentz invariance and locality. Of course, electromagnetism with gauge redundancy removed is still electromagnetism, so it is still Lorentz invariant and local, it is just not manifestly so." (Here I only add, that "removing redundancy" is not only a clumsy thing to do, but that for instance for non-Abelian gauge theories and for generally covariant theories (in $D > 1$) no one succeeded so far in finding the non-redundant degrees of freedom.)

So whereas Gross ascribes gauge symmetries to Nature itself, Hinterbichler doubts the mere existence of symmetries. According to him the so-called local gauge symmetries relate to our inability to describe the theories in terms of the truly independent degrees of freedom. This sounds heretical, as it replaces symmetries by a redundancy in the mathematical description. Indeed it leads to the question of which entities in our theories correspond to reality. Maybe things are more subtle: C. Rovelli argues that gauge is more than mathematical redundancy [452], and less than observability, but the revelation of the "relational structure of our world". In order to understand this one needs to comprehend his idea of 'partial observables' (advocated for instance in [451])–quantities that can be measured but not predicted.

In trying to find out, in which sense symmetries may serve as principles of Nature, let me remind you, that we need to precisely state which symmetry transformations we are dealing with in fundamental physics: The symmetry transformations investigated in Yang-Mills type theories are not only restricted to point symmetries, but even more to those for which the new coordinates do not depend on the fields, namely $\hat{Q}^\alpha = Q^\alpha + \epsilon\eta^\alpha(x, Q)$, $\hat{x}^\mu = x^\mu + \epsilon\xi^\mu(x)$. But you agree that the claim "Fibre-preserving point symmetries dictate the laws of Nature" is not that exciting.

Are Gauge Fields Real?

What is real? This is of course another philosophical question. A possible answer is found in a joke relating to the people of East-Prussia[18]: Everything that can be fetched by your hands is real. By this, a cat and a hot oven are not real. This is easy, but what about the electromagnetic potential?

To become more serious: A "let-me-grasp"-attitude of reality is definitely to naive. By this none of the entities called "particles" in the Standard Model are real. And trying to relate reality to observables (used here in common sense) is not without problems: Observations are necessarily of macroscopic nature, and all observations relating to quantum entities refer for instance to energy spectra, decay rates, cross-sections, . . .

This subsection should be the place to embed the topic of symmetries into the context of the philosopher's scientific realism debate. A remarkable current trend in this debate is structural realism [330], [483]. By this one understands that scientific theories are not dealing with objects representing entities in the world, but by structures among objects. It is claimed that it is the structures/relations which are the stable anchor in the development of science whereas the meaning of the objects can change, or objects being *en vogue* for some time may cease away (take the "ether" as an example). This makes understandable the continuity in the transition from an old to a new theory in its mathematical structure, whereas the content changes. The structural realism approach serves as bridge between realists and anti-realists. If the entities postulated by successful theories did not exist, then the high predictive

[18] It really should be told in their dialect–as I experienced it from my high school teacher, Dr. Szillis, who is responsibly for my own fascination with mathematical physics.

power and explanatory success of modern science would be treated as a miracle; this is called the no-miracles argument by philosophers of science. On the other hand, there is still the ‘pessimistic induction’ opinion by which we have no reason to believe that our currently successful theories are true, given that past scientific theories which were successful at their time were refuted or abandoned afterwards. Thus structural realism weakens the latter anti-realism argument without claiming that all entities in physical theory corresponds to real entities in the world.

To me it seems that structural realism is a promising approach if it comes to symmetries,[19] and indeed is worked out in more depth by philosophers of science, for instance in [354]. However, “the devil is in the details”, and various manifestations of structural realism were thus proposed: Epistemic structural realism maintains that only relations between objects in the world, but not the objects themselves can be grasped as being real. Ontological structural realism claims that the reality of objects has no meaning independent of their interrelations. And further classes and subclasses are discussed.

The question whether an entity in a physical theory has a pendant in the world arises especially for those entities which are not observable in the sense of not being determinable by a measurement. Are these only instrumental without having the same ontological status as the observables?

Take electrodynamics as an example. The heading of this paragraph asks for the reality of the gauge fields, or–in the understanding of particle physics–the gauge bosons. In electrodynamics this is described by the four-potential A_μ. This is not identifiable with the photon, since the photon has only two degrees of freedom. Using the electromagnetic potential A_μ instead of the $\vec{E}$- and $\vec{B}$-fields has the advantage of making manifest the invariance of electrodynamics with respect to Poincaré and local $U(1)$-transformations. But does it make sense to assign to the electric and magnetic fields ’more’ reality than to the A_μ-field ? The latter is defined only up to gauge transformations and is thus not measurable. This conclusion is no longer valid if we go quantum and consider for instance the Aharonov-Bohm effect: an electrically-charged particle is influenced by an electromagnetic field, despite being confined to a region in which both the electric field $\vec{E}$ and the magnetic field $\vec{B}$ are zero. The interaction is through the coupling of the electromagnetic potential with the complex phase of a charged particle’s wave function. Thus in QED one needs for its description more than the $(\vec{E}, \vec{B})$ and less than the full A_μ, but something in between–called holonomies. In simplified terms these are loop-dependent quantities in which the electromagnetic potential is integrated along the loop. As especially emphasized by T. T. Wu and C. N. Yang [568], the minimal and complete set of variables in a gauge theory are the non-integrable phase factors

$$\mathcal{P}e^{\frac{ie}{\hbar c}\int_a^b A_\mu dx^\mu}$$

[19] It also blends in well with the relational interpretation of spacetime in the sense of Leibniz and Mach.

associated to a finite translation of a charged object from a to b (and where $\mathcal{P}$ denotes path ordering).

Be aware that strictly speaking also the matter fields are not observable. Quantum wave functions are defined only up to a phase. Only bi-linears (currents in general) built by a wave function together with its complex or Hermitean conjugate can be observed because the phase drops out. However, again there is an exception: M. Berry pointed out in 1984 that if a system undergoes an adiabatic change, the phase may have observable consequences. This is related to the Aharonov-Bohm effect. Both also go under the name 'geometric phase' and are now known to arise in various areas of physics. They can most appropriately be explained in terms of non-trivial fiber bundles, but even then various stances concerning "reality" are possible; see e.g. [35] (don't be surprised by the title: there are indeed analogies between falling cats and Yang-Mills potentials).

A question about the 'reality' of entities appearing in the formulation of a theory also arises for the electroweak sector of the standard model [355]. In the main text, the germ of the electroweak theory was explained by help of a tachyonic scalar field. After spontaneous symmetry breaking the only thing left from this scalar field is the Higgs boson. Is the original tachyon part of reality or only part of a concise description of the Higgs mechanism? Why this detour with the Higgs mechanism anyhow, and why not present and explore the electroweak theory in its ultimate form? Because the $\mathbf{SU(2)} \times \mathbf{U(1)}$ symmetry would be hidden? One should also understand that the Higgs mechanism is only a description of the electroweak spontaneous symmetry-breaking process. It does not explain the breaking itself, since there is no dynamics to explain the instability of the scalar potential at the origin.

Quantum theory is the most prominent example of an otherwise extremely successful physical theory for which the debate about 'realism' is still open. This is because of the measurement problem and the problem of interpretational underdetermination.

9.2.4 Origin of Symmetries

There is another question which is worth addressing: Can one think about mechanisms by which symmetries emerge from "something else"? Is it possible that symmetries are emergent and as such not a basic principle of Nature? Aside from forerunners in the early 1960's which aimed to derive QED from a field theory with four-fermion couplings (W. Heisenberg's nonlinear spinor theory from 1957), it was the 'random dynamics' program of H. B. Nielsen *et al.* which exposed and qualified the idea that none of the symmetries ascribed to our current understanding of Nature are (ontologically) actually present; for an outline of the program, see [391]. Its credo is expressed in [189] "... that symmetries and physical laws should arise naturally from some essentially random dynamics rather than being postulated to be exact or adjusted by hand" and that "Therefore we no longer have to believe that gauge theories occur in nature for some mysterious and deeply hidden reason."

There are other approaches, in which symmetry is a dynamical concept, derived instead of postulated from the beginning. For instance, the authors of [286] start from Lagrangian densities for scalar and vector fields which are not necessarily invariant under a symmetry group, and calculate their β functions. Then they show that in some of these theories, the fixed points are infrared attractors in case of symmetries. This means that a large class of possible–and even non-renormalizable Lagrangians–is attracted to a symmetric Lagrangian if one moves to lower energies and momenta. This was also exemplified for Lorentz invariance: In [82], the authors consider a non-covariant model of electrodynamics and show by calculating the relevant β functions that Lorentz invariance tends to becomes better and better at "low" energies, that is for energies which are currently attainable in what is called "high-energy" physics.

In another series of papers [91] it could be shown that the symmetries of Yang-Mills theories and of linearized general relativity can be due the spontaneous violation of Lorentz invariance.

9.3 Physics Beyond Symmetries

This book is motivated by the aim to explore how much of fundamental physics rests on symmetries and can even be derived from symmetry arguments. Nevertheless, my readers and I are aware of physics beyond symmetries.

9.3.1 Prominent Non-Symmetries

The most prominent example of a non-symmetry is time asymmetry. In classical mechanics, the equations of motion are invariant under time reversal, as long as there is no frictional force depending on the velocity. The theories of the fundamental interactions, all formulated on the microscopic level, are invariant under time reversal as well, except for a "tiny" corner of weak interaction. On a macroscopic scale, however, this invariance fades away. We all sense in our daily life how time passes by and cannot be reversed. The reasons for the time asymmetry of our (macroscopic) world are not completely understood thus far. There are different 'time arrows', for instance the thermodynamic one (related to the notion of entropy) and the cosmological one (related to the expansion of the universe). To interested readers, I warmly recommend [580].

Another non-symmetry is that under rescalings of objects. This was already known to Galilei. If you contemplate how to make a building higher than an existing one by rescaling all of its building blocks, you will find that each stone and steel beam upscaled in size will sooner or later break due to the gravitational forces exerted on it. Indeed, the gravitational attractive force will win over the electromagnetic force serving for stability on ordinary scales. Therefore the stories about Gulliver meeting the Lilliputanians and the Brobdingnagians are indeed pure fictions. The non-invariance of scale of the fundamental interactions regulates the size of living beings and artifacts on earth depending on their fabric.

Within non-relativistic mechanics, length, time, and energy all scale independently. Galilei symmetry allows us to use independent scale factors for each of these entities. Therefore, for any non-relativistic system, there exists another system which is for instance twice as large, has a period three times as long, and contains half the energy. There is no fundamental constant that can be built in terms of length, time, and energy. In relativistic mechanics however, there is such a constant, namely the vacuum velocity of light. By this, length and time need to be rescaled by the same factor λ. Due to $E^2 = m^2c^4 + p^2c^2$, masses, momenta and energy scale in the same way, by a factor λ', say. In a quantized theory we also have $E = h\nu$, and the presence of the Planck constant h restricts further scale transformations to $\lambda' = \lambda^{-1}$. Similarly, in classical electrodynamics with the expression for the potential energy $U = e^2/r$ (with the smallest electric charge e as a constant), one finds that the energy scales with the inverse length.

Invariance under rescaling (or dilations as it is called in the context of conformal symmetries) holds on the level of the microscopic fundamental theories only if there is no inherent mass (or length) scale. Scale invariance holds approximately, more so as the inherent masses become smaller. We saw this for high-energy QCD as an additional symmetry originating for vanishing quark masses.

On the other hand the scale dependence in quantum field theories related to the dependence of coupling parameters on energy is described by the renormalization group equation and encoded in the β function. The renormalization group equation is also an essential tool if one wishes to calculate the critical exponents of phase transitions. Indeed it can always be applied if the phenomenon in question shows features of self-similarity, that is, if the system looks qualitatively the same on different scales. For critical phenomena this is the case at the critical point because of the divergent correlation length of the statistical fluctuations; see [187, 571]. Critical points thus exhibit scale invariance. This reconstitution of scale invariance plays an essential, prominent role in "chaoplexitity", in which I include ideas about chaos (with fractals and bifurcations), complex systems (with nonlinearities and openness) and phenomena like self-organization, pattern formation, and phase transitions.

9.3.2 Other Notions of Fundamental Physics

Reductionism vs. Emergence

The development of the standard model followed a reductionist approach, which explains matter in terms of atoms, atoms in terms of nuclei and electrons, nuclei in terms of hadrons, hadrons in terms of quarks. Such a reductionist pattern holds true more general. One could have started a hierarchy of scientific disciplines at even higher levels, such as Level 8: Sociology/Economics/Politics, Level 7: Psychology, Level 6: Physiology, Level 5: Cell biology, Level 4: Biochemistry, Level 3: Chemistry, Level 2: Atomic physics, Level 1: Particle physics. This distinction of levels is taken from an article by G. Ellis [165].[20] On each level you find a specific language spoken by specifically educated people. On the bottom-level we have, what in this book is called fundamental physics. Although the next level could "in principle" be described in terms of the particle physics level, it is at least inappropriate to do so. Atomic physics, and if you opt for a Level 3': Many-body physics–make use of completely different concepts in their domains. For example, the macroscopic motion of a fluid is appropriately described by a wave equations. Even a description in terms of the Navier-Stokes equations is largely inadequate, let alone a description in terms of the interacting nuclei in the fluid molecules. This is somehow trivial, but it is mentioned here to reflect the difficulty in going bottom-up in the hierarchy of levels. In going up, arise or "emerge" patterns that do not have an obvious cause on the lower level. For example for a bulk of material one can measure a temperature; but the atoms and molecules of which the bulk is composed, do not have a temperature.

A reductionist necessarily looses sight for the whole, larger issue. Thus he fails to describe emergent phenomena. This prompted R. B. Laughlin to make the remark, that with a "Theory of Everything" not physics would come to an end, but only reductionistic physics [334].[21] It is claimed by many that "We cannot expect everything of a Theory of Everything" [31] (pg. 143). So the question is whether a TOE could in principle explain the working of a living cell or even of entities in the higher hierarchy levels. Is it only a matter of computational power, or are fundamentally new laws needed to explain the properties of complex systems with their capability of self-organization and emergence of collective properties?

In 1972, Nobel laureate P. W. Anderson wrote a short but influential article [8] "More Is Different", in which he gave examples from many-body physics that the hierarchical structure of science might be understood by breaking symmetries; symmetries that hold on a lower (more "fundamental") are broken on a higher level.

Symmetry Breaking

We would not be able to discover symmetries if they were not broken. For example the SU(2) isospin symmetry on the nuclear level between a proton and a neutron would not have been registered, if both had exactly the same mass and charge.

---

[20] G. Ellis is a highly respected relativist and cosmologist, who later in his career also worked on complex systems. I mention this because to me this gives him more reputation to discuss emergence and self-organization than many philosophers of science who deal with this topic.

[21] The German title of this book is "*Abschied von der Weltformel*"–something like "Farewell to the World Formula"; and this reveals the contrasting aspect of Laughlin's book to those bestsellers of B. Green and S. Hawking.

And although symmetry is the golden thread of fundamental physics, there are various phenomena of symmetry breaking. They arise with the Higgs mechanism, CP violations in weak interactions, matter-antimatter asymmetry, the occurrence of 'anomalies', and of course with violations of so far unjustified 'symmetries of everything'. The Higgs mechanism is an example of spontaneous symmetry breaking: The underlying action exhibits a symmetry, which however ceases to be a symmetry of the ground state. This is to be contrasted to explicit symmetry breaking: The dominant part of the action is symmetric, but there is another part which violates the symmetry. Without trying to quantify what "dominant" means in this context, we notice that symmetry itself loses its meaning if the symmetry non-conserving part becomes of the same order as the symmetry-conserving part.

Cosmology itself is nowadays understood as a history of symmetry breaking. It is assumed, that the universe had its greatest symmetry at the big bang. This is the symmetry of an envisaged TOE, let it be superstrings or a SO(32) gauge theory. In the expansion of the universe with concomitant cooling, the TOE splits into gravitation and a GUT (for example with a SO(10) symmetry) or a supersymmetric theory, which after further expansion spontaneously breaks into the strong interaction SU(3) symmetry and the electroweak theory with its $SU(2) \times U(1)$ symmetry. Again, further expansion with decreasing temperature results in another spontaneous symmetry breaking into the $U(1)$ symmetry of the electromagnetic theory. In the high-energy experiments today, one can see this last stage of cosmological history, in the sense that the energy/temperatures reached in the experiments allow one to verify the electroweak symmetry breaking.

It was mentioned that the very concepts of spontaneous symmetry breaking were developed by Y. Nambu and others with regard to phenomena in solid-state physics, examples being Euler's instability analysis of rods under longitudinal compressional forces, rotating self-gravitating equilibrium figures, aggregate states and phase transitions, magnetism. Also the renormalization-group procedure was refined by K. G. Wilson in order to calculate critical exponents of phase transitions.

Are There Other Basic Objects to Describe Physics?

In the introductory chapter, I motivated my choice of 'fundamental' physics: It is concerned with the four interactions known today, and the essential entity is the action functional. This in turn is formulated in a spacetime in terms of fields and their derivatives. But is it imaginable that in some far future time, physics will be based on completely different concepts? Already today there are ideas how to found physics on information, giving entropy a prominent role. Entropy is a concept that originated in thermodynamics, and thermodynamics makes strong use of emergent notions.

Is it conceivable that the 'fundamental' interactions can be described in terms of thermodynamics? This question might sound silly; but there is already a viable example: General relativity can be derived from black hole thermodynamics. This started with the work of T. Jacobson [299], is strongly promoted by T. Padmanabhan

[403], and recently was revived under the catchword 'entropic gravity' introduced in an article by E. Verlinde [522].

There are various investigations in which 'information' becomes the pivotal concept of thermodynamics and of quantum theory; it is even claimed that the universe is nothing but a computer [344]. Related to this is the question of how black holes destroy information, and whether there is a 'holographic principle' by which the universe has only limited capacity to store and process information. In [506], the classical least-action principle is derived from the action being a counter of computations.

There are other approaches by which even space and time are not fundamental ingredients to describe nature, but are derived from some deeper level structures like spin networks in loop quantum gravity.

Let me also mention a remarkable numerical quantum gravity calculation by J. Ambjorn, J. Jurkiewicz J. and R. Loll [6] starting from a functional integral in which each spacetime history appears with a weight given by the exponentiated Hilbert-Einstein action of the corresponding geometry. Computer simulations on this functional integral lead to the result that a 4D de Sitter universe emerges on large scales. The authors remark "This emergence is of an entropic, self-organizing nature, with the weight of the Einstein-Hilbert action playing a minor role. Also the quantum fluctuations around this de Sitter universe can be studied quantitatively and remain small until one gets close to the Planck scale. The structures found to describe Planck-scale gravity are reminiscent of certain aspects of condensed-matter systems." This latter facet is most excellently described in [346].

But even if, in the far future, a book on the foundations of physics will start off with the notions of information and entropy, it certainly must explain why 'in former times' physics could be understood from an action and its variational symmetries. That is, the action and its symmetries must be themselves emergent. Specifically the "random dynamics" program by H. B. Nielsen and his co-workers[22] addresses the question of how not only symmetries, but other features of the current world model can possibly be derived from some random and chaotic, more fundamental theory. Although not fully developed, and although in some of the publications disputable assumptions are made, within the program various interesting results could be derived: the origin of gauge symmetries, of the number of flavors, of the standard model gauge group, of the 3+1 dimensionality of our world, of Lorentz symmetry.

There are even conferences dedicated to the question, how quantum theory can be arrived at, starting from other basic notions, which are rooted for instance in thermodynamics, mathematical logic or systems dynamics [242]. For instance, G. 't Hooft proposes that on a deeper level there is a deterministic classical, chaotic dynamics, where a set of attractors determine the emergent effective quantum theory [501]. In [503], he claims: "… symmetries such as rotation symmetry, translation, Lorentz invariance, but also local gauge symmetries and coordinate reparametrization invariance, might all be emergent, together with quantum mechanics itself."

---

[22] There is a website: www.nbi.dk/ kleppe/random/qa/qa.html.

9.3.3 Are we Biased, or Haughty, or Simply in a Specific World?

Are we biased in our manner of doing physics? Despite the astounding progress which fundamental physics has seen in recent decades, and despite the phantasies of theoretical physicists, one might still be critically scrutinize prejudices of historically and culturally-established assumptions: We are still formulating our theories based on the mathematics of differentiable manifolds with their smoothness, in terms of commutative geometries, as local interactions. Or are we biased in not being able to look far enough? A. Zee: "Is the Ultimate Design truly symmetric, or is it that we can only understand the symmetric part of the Ultimate Design?"

Perhaps we are not only biased but haughty: Why can we dare to believe that there are only four interactions in nature and that we have found a "*Weltgesetz*"? If at all, we might have found the structure of the "*Weltgesetz*", as for example the Yang-Mills structure of the standard model. But we are unable to explain the origin of the gauge group, and the content in terms of the lepton families. And how can we believe to have found the laws of fundamental physics while remaining ignorant of 95 % of the matter-energy content of the universe.

And maybe we are not haughty but simply in a specific world: Perhaps our very existence might falsely convince us that symmetries are a basic constituent of the world: We could be only part of a multiverse [76] in which by a happenstance symmetries seem to play a prominent role. This leads immediately to the anthropomorphic controversy initiated by models of cosmological inflation ('bubble universes') and the 'string landscape' in the last decade.

Indeed it is astounding that only few numbers in the laws as we know them today need to be fine-tuned in order that life can exist.[23] According to M. Rees [436], these are the six numbers (1) N: ratio of electrical and gravitational forces between protons $= 10^{36}$, (2) E: nuclear binding energy as a fraction of rest mass energy $= 0.007$, (3) Ω: amount of matter in the universe in units of critical density $= 10^{-5}$, (4) Λ: cosmological constant in units of critical density $= 0.7$, (5) Q: amplitude of density fluctuations for cosmic structures, (6) D: number of spatial dimensions $= 3$.

The argument of the 'string landscape' as advocated by L. Susskind [494], is that the compatibility of string theory with its 10^{500} (or so) different solutions is not a bad feature, but should instead be taken seriously because it allows an anthropic argument for the small size of the cosmological constant. The landscape hypothesis is heavily disputed because it touches the basic pillar on which physics has rested since the times of Galileo Galilei, namely making predictions that possibly can be verified (or falsified). A book review by G. Ellis [164] starts with the words "Once upon a time, physics dealt with tangible objects–if you couldn't weigh them or smash them together, at least you could observe them. As times changed, physicists started to deal with more ethereal things: electromagnetic felds and space-time metrics, for example. You couldn't see them but you could measure their influence on particle

[23] In [65], the consequences of mostly slight changes of the parameters of the Standard Model on everyday life are elucidated.

trajectories and so justifiably claim evidence of their existence. Nowadays things have changed. A phalanx of heavyweight physicists and cosmologists are claiming to prove the existence of other expanding universe domains even though there is no chance of observing them, nor any possibility of testing their supposed nature except in the most tenuous, indirect way." and he criticizes that, "… science is venturing into areas where experimental verification simply isn't possible."

Appendix A
Group Theory

This appendix is a survey of only those topics in group theory that are needed to understand the composition of symmetry transformations and its consequences for fundamental physics. It is intended to be self-contained and covers those topics that are needed to follow the main text. Although in the end this appendix became quite long, a thorough understanding of group theory is possible only by consulting the appropriate literature in addition to this appendix. In order that this book not become too lengthy, proofs of theorems were largely omitted; again I refer to other monographs. From its very title, the book by H. Georgi [211] is the most appropriate if particle physics is the primary focus of interest. The book by G. Costa and G. Fogli [102] is written in the same spirit. Both books also cover the necessary group theory for grand unification ideas. A very comprehensive but also rather dense treatment is given by [428]. Still a classic is [254]; it contains more about the treatment of dynamical symmetries in quantum mechanics.

A.1 Basics

A.1.1 Definitions: Algebraic Structures

From the structureless notion of a set, one can successively generate more and more algebraic structures. Those that play a prominent role in physics are defined in the following.

Group

A *group* $\mathbf{G}$ is a set with elements g_i and an operation $\circ$ (called group multiplication) with the properties that (i) the operation is closed: $g_i \circ g_j \in \mathbf{G}$, (ii) a neutral element $g_0 \in \mathbf{G}$ exists such that $g_i \circ g_0 = g_0 \circ g_i = g_i$, (iii) for every g_i exists an inverse element $g_i^{-1} \in \mathbf{G}$ such that $g_i \circ g_i^{-1} = g_0 = g_i^{-1} \circ g_i$, (iv) the operation is associative: $g_i \circ (g_j \circ g_k) = (g_i \circ g_j) \circ g_k$. The closure property can be expressed

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

Table A.1 The group D_4

	1	i	−1	−i	C	Ci	−C	−Ci
1	1	i	−1	−i	C	Ci	−C	−Ci
i	i	−1	−i	1	−Ci	C	Ci	−C
−1	−1	−i	1	i	−C	−Ci	C	Ci
−i	−i	1	i	−1	Ci	−C	−Ci	C
C	C	Ci	−C	−Ci	1	i	−1	−i
Ci	Ci	−C	−Ci	C	−i	1	i	−1
−C	−C	−Ci	C	Ci	−1	−i	1	i
−Ci	−Ci	C	Ci	−C	i	−1	−i	1

by considering the operation as a mapping

$$\circ : G \times G \to G.$$

There is indeed only one neutral element and the inverse of a group element is unique, as easily can be shown.

A group is called *finite* if the number of its elements (also called the *order* of the group) is finite. An example of a finite group is the rotations of a square by 90° (in mathematics and crystallography called $\mathbf{C}_4$; a group of order 4). Examples of infinite groups are the integers with respect to addition (countably infinite) and the symmetry group of the circle, characterized by a parameter φ $[0 \leq \varphi < 2\pi]$ (continuous group).

The *center of a group* is defined as the set of those elements of **G** which commute with all other elements of the group:

$$\mathrm{Cent}(\mathbf{G}) = \{c \in \mathbf{G} \mid cg = gc \quad \text{for all} \quad g \in \mathbf{G}\}.$$

Example: From the multiplication table of $\mathbf{D}_4$ (see Table A.1; more about this group later), one can see that its center is given by the set $\{1, -1\}$.

A group is called *Abelian*[1] if the operation is commutative: $g_i \circ g_j = g_j \circ g_i$. Obviously the center of an Abelian group is the group itself.

The symmetry group $\mathbf{C_4}$ of the square is Abelian; the analogous group of the cube is an example of a non–Abelian group. In physics we often deal with groups (or subgroups of) $\mathbf{GL}(n, \mathbb{F})$. This is the set of non–singular $n \times n$ matrices with elements from the field $\mathbb{F}$ which are in the majority of cases real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$. $\mathbf{GL}(n, \mathbb{F})$ is in general a non–Abelian group with matrix multiplication as the group operation. The restriction to non–singular matrices is necessary because otherwise the inverse would not be defined.

A subset of **G**, which itself is a group, is called a *subgroup* of **G**. By this definition, every group has two trivial subgroups, namely itself ($\mathbf{G} \subset \mathbf{G}$) and the set consisting of the neutral element ($\{g_0\} \subset \mathbf{G}$).

[1] This term derives from the name of the Norwegian mathematician Niels Henrik Abel (1802–1829). He is treated with some disrespect by the fact that contemporary theoreticians use his name in a negated form, by talking for instance of "non–Abelian" gauge theories.

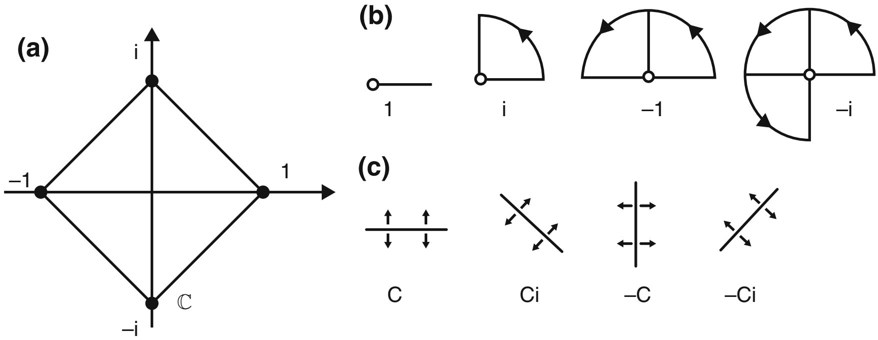


Fig. A.1 Symmetries of the square

Although in the context of symmetries in fundamental physics, continuous group are the most important ones, many notions from group theory will subsequently be explained on a finite group. This is the group $\mathbf{D}_4$, defined by its multiplication table above. The notation of its elements is taken from [410]. $\mathbf{D}_4$ is a group of order 8, its neutral element is (1), and it is non–Abelian (for instance $(Ci)C = -i$, but $C(Ci) = i$). This abstract group is realized by 90°-rotations of the square and reflections along its diagonals; see Fig. A.1. Here the vertices of the square are represented by the points 1, i, –1, –i. The group $\mathbf{D}_4$ has various nontrivial subgroups; these are

- order 2: rotations by 180° realized by $\mathbf{C_2} = \{1, -1\}$, and four reflections realized by $\{1, C\}$, $\{1, Ci\}$, $\{1, -C\}$, $\{1, -Ci\}$.
- order 4: rotations by 90° realized by $\mathbf{C_4} = \{1, i, -1, -i\}$, reflections along diagonals, that is $\{1, -1, C, -C\}$, and reflections along the two parallels to the edges of the square: $\{1, -1, Ci, -Ci\}$.

Field

A *field* is a set $\mathbb{F}$ with elements f_i together with two operations

$$+: \ \mathbb{F} \times \mathbb{F} \to \mathbb{F} \qquad\qquad * : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$$

with the properties that (i) $\mathbb{F}$ is an Abelian group under the operation $+$ with f_0 as the identity, (ii) $\mathbb{F} \setminus \{f_0\}$ is a group under $*$, and (iii) there is a distributive law

$$f_i * (f_j + f_k) = f_i * f_j + f_i * f_k, \qquad\qquad (f_i + f_j) * f_k = f_i * f_k + f_j * f_k.$$

Prominent examples of fields are the real and the complex numbers under the usual addition and multiplication operations. Further fields are known, as for instance quaternions (building a 4–dimensional number system and having some link to Dirac matrices), octonions, sedenions,…the latter bearing a relation to the exceptional Lie groups.

Vector Space

A (linear) *vector space* consists of a set V with elements $\mathbf{v}_i$ and a field $\mathbb{F}$ together with two operations

$$+: \qquad V \times V \to V \qquad\qquad \star : \mathbb{F} \times V \to V$$

with the properties that (i) $(V, +)$ is an Abelian group, (ii) the operation $\star$ fulfills

$$\begin{aligned}
&f_i \star (f_j \star \mathbf{v}_k) = (f_i * f_j) \star \mathbf{v}_k \\
&1 \star \mathbf{v}_i = \mathbf{v}_i = \mathbf{v}_i \star 1 \\
&f_i \star (\mathbf{v}_j + \mathbf{v}_k) = f_i \star \mathbf{v}_j + f_i \star \mathbf{v}_k, \qquad (f_i + f_j) \star \mathbf{v}_k = f_i \star \mathbf{v}_k + f_j \star \mathbf{v}_k.
\end{aligned}$$

Further definitions:

- A set of vectors $\mathbf{v}_k$ is called *linearly dependent* iff $f_j \in \mathbb{F}$ exist such that $\sum\limits_k f_k \, \mathbf{v}_k \equiv 0$.
- The *dimension* D of a vector space is defined as the maximal number of linearly independent vectors.
- In every vector space of dimension D one can find a set of basis vectors $\mathbf{e}_k$ which are linearly independent and which make it possible to expand any vector as $\mathbf{v} = \sum\limits^{D} f_j \mathbf{e}_j$.

Prominent examples of vector spaces are the Euclidean space, the solution space of linear differential and integral operators, the sets of $N \times M$–matrices, and Hilbert spaces.

Algebra

A (linear) *algebra* A consists of a set V with elements $\mathbf{v}_i$, a field $\mathbb{F}$ together with three operations $+, \star, \Box$. The algebra is a vector space with respect to $(V, \mathbb{F}, +, \star)$. The $\Box$ is a mapping $\Box : V \times V \to V$ with

$$\begin{aligned}
(\mathbf{v}_i + \mathbf{v}_j)\Box\mathbf{v}_k &= \mathbf{v}_i\Box\mathbf{v}_k + \mathbf{v}_j\Box\mathbf{v}_k \\
\mathbf{v}_i\Box(\mathbf{v}_j + \mathbf{v}_k) &= \mathbf{v}_i\Box\mathbf{v}_j + \mathbf{v}_i\Box\mathbf{v}_k.
\end{aligned}$$

Prominent examples are *Lie algebras*, which aside from the algebra properties additionally fulfill

$$\begin{aligned}
&\mathbf{v}_i\Box(a\mathbf{v}_j + b\mathbf{v}_k) = a\mathbf{v}_i\Box\mathbf{v}_j + b\mathbf{v}_i\Box\mathbf{v}_j, \qquad\qquad a, b \in \mathbb{F} \\
&\mathbf{v}_i\Box\mathbf{v}_j = -\mathbf{v}_j\Box\mathbf{v}_i \\
&\mathbf{v}_i\Box(\mathbf{v}_j\Box\mathbf{v}_k) = (\mathbf{v}_i\Box\mathbf{v}_j)\Box\mathbf{v}_k + \mathbf{v}_j\Box(\mathbf{v}_i\Box\mathbf{v}_k),
\end{aligned}$$

where the latter relation is the Jacobi identity, as it is known in physics for instance for the Poisson brackets or the generators of bosonic/Grassmann-even[2] symmetry transformations.

A.1.2 Mapping of Groups

Group Homomorphism

Let f be a group mapping $f : \mathbf{G} \to \hat{\mathbf{G}}$. If the mapping respects the group structure, that is if

$$f(g_1) \circ_{\hat{\mathbf{G}}} f(g_2) = f(g_1 \circ_{\mathbf{G}} g_2),$$

(where $\circ_{\mathbf{F}}$ denotes the operation in the respective group) it is called a (group) *homomorphism*. The definition of a homomorphism can be generalized to fields, vector spaces, and algebras as structure–preserving maps between two of these algebraic structures.

The *kernel* of the homomorphism f is defined as the subset of $\mathbf{G}$ for which each element is mapped to the neutral element in $\hat{\mathbf{G}}$:

$$\mathrm{Ker} f := \{g \in \mathbf{G} \mid f(g) = \hat{g}_0 \ \in \hat{\mathbf{G}}\}.$$

The definition of homomorphism does not exclude that $f(g_1) = f(g_2)$ for $g_1 \neq g_2$ (a so-called "many-to-one" relation). For instance, mapping all elements of an arbitrary group to the real number 1 is a homomorphism between the group and the group of real numbers. A "one-to-one" mapping is called an *isomorphism*. Isomorphic groups are not distinguishable in their mathematical structure; one denotes $\mathbf{G} \cong \hat{\mathbf{G}}$. If $\mathbf{G} = \hat{\mathbf{G}}$ the map f is called an *automorphism*. An *inner automorphism* is a mapping f_h such that $f_h \mapsto g^{-1}hg$.

For finite groups an isomorphism between two groups is easily detected in that both groups have identical multiplication tables. In the symmetry group of the square $\mathbf{D}_4$, for example, the group called $\mathbf{C_2}$ has the same multiplication table as all the other subgroups of order two (in the classification by crystallography these are called $\mathbf{D_1}$):

$$\mathbf{C}_2 \cong \mathbf{D}_1 \cong \{1, -1\} \cong \mathbf{Z}_2$$

which also defines the group $\mathbf{Z}_2$, as it occurs in the context of C-,P-, and T-transformations. In general, the groups $\mathbf{C_{2N}}$ are the cyclic groups: If their elements are written as $\{1, a, \ldots, a^{2N-1}\}$ their multiplication table can be generated compactly by the rule $a^n \circ a^m = a^{(n+m) \bmod 2N}$. Therefore, the group of rotations by 90° is isomorphic to $\mathbf{C_4}$ (take $a = i$). The two other order–4 subgroups are isomorphic

$$\mathbf{D}_2 \cong \{1, -1, C, -C\} \cong \{1, -1, Ci, -Ci\}.$$

Indeed $\mathbf{C_2}$ is the only abstract group of order two, and $\mathbf{C_4}$ and $\mathbf{D_2}$ are the only groups of order four.

---

[2] To be explained in B.2.1.

Representation of Groups

In fundamental physics, it is not the symmetry groups themselves that are of primary significance, but–for reasons arising from quantum theory–the "irreducible unitary representations of their central extensions". These terms are defined in Sect. 3 of this appendix, a section entirely devoted to representation theory. A (matrix)–*representation* D is a homomorphism

$$D : \mathbf{G} \to \mathbf{GL}(n, \mathbb{F}),$$

where n is called the *dimension* of the representation.

Example: Rotations of the square, now denoted by $\{r0, r90, r180, r270\}$

- 1-dimensional complex representation of $\mathbf{C}_4$:

$$r0 \mapsto 1, \qquad r90 \mapsto i, \qquad r180 \mapsto -1, \qquad r270 \mapsto -i.$$

- 2-dimensional real representation of $\mathbf{C}_4$: Since rotations in a plane can be written as $\vec{x}' = \mathrm{R}\,\vec{x}$ with a matrix

$$\mathrm{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

we get immediately the representation

$$r0 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad r90 \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$
$$r180 \mapsto \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad r270 \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

A.1.3 Simple Groups

It is obvious to ask how many different (that is, not mutually isomorphic) groups there are. A first step towards an answer is the investigation of *simple* groups, where "simple" is to be understood in the sense of "prime": The idea is to factorize a group into its simple subgroups. Before a definition can be given, we need some further group–theoretical terms and definitions.

Equivalence Classes

An *equivalence relation* $\sim$ on a set $M = \{a, b, \ldots\}$ is defined by its properties: (i) $a \sim a$, (ii) $a \sim b \Leftrightarrow b \sim a$, (iii) $a \sim b, b \sim c \Rightarrow a \sim c$. Every equivalence relation on M allows to subdivide the set into *equivalence classes*

$$[a] := \{a \sim b_i \mid a, b_i \in M\},$$

i.e. the class $[a]$ contains all those elements, which are equivalent to a. Thus M is divided in disjunct classes:

$$M = \cup_i \ [a_i] \quad \text{with} \quad [a_i] \cap [a_j] = \emptyset \quad \text{if} \quad a_i \nsim a_j.$$

Cosets

If $\mathbf{H}$ is a subgroup of $\mathbf{G}$, the set

$$g\mathrm{H} := \{g \circ h_i \mid g \in \mathbf{G},\ h_i \in \mathbf{H}\}$$

is called *left coset*. (A right coset $\mathrm{H}g$ can be defined analogously.) This gives rise to the introduction of an equivalence relation

$$g_i \sim g \ \Leftrightarrow g_i \ \in g\mathrm{H}.$$

Thus the group G can (as a set) be subdivided in disjunct cosets

$$\mathrm{G} = \cup_1^p \ g_i\mathrm{H} \qquad \text{with} \qquad g_i\mathrm{H} \cap g_j\mathrm{H} = \emptyset \quad \text{for} \quad g_i \neq g_j.$$

Example: Take $\mathbf{H} = \{1, Ci\}$ to be a subgroup of $\mathbf{D}_4$. Because of

$$\begin{aligned} 1\{1, Ci\} = \{1, Ci\}, \quad i\{1, Ci\} = \{i, C\}, \quad -1\{1, Ci\} \\ = \{-1, -Ci\}, \quad -i\{1, Ci\} = \{-i, -C\} \end{aligned}$$

the full set G is representable as $\mathrm{G} = 1\mathrm{H} \ \cup i\mathrm{H} \ \cup (-1)\mathrm{H} \ \cup (-i)\mathrm{H}$.

The *quotient* of a group w.r.t. to one of its left cosets is defined as

$$\mathbf{G}/\mathbf{H} = \{g_1 H, g_2 H, \ldots, g_p H\}.$$

For finite groups, one can show, that if $\mathbf{H}$ is of order q, $\mathbf{G}$ is of order $p * q$. (In the example, the subgroup is of order 2, the number of independent cosets is 4, the full group is of order $8 = 2 * 4$.) Thus the order of a subgroup must be a divisor of the order of the full group, and therefore groups of prime order (2, 3, 5, 7, 11,...) do not have non–trivial subgroups.

Conjugate Classes and Conjugate Groups

An element $\bar{g} \in \mathbf{G}$ is said to be conjugate to the element g if one can find an element $\tilde{g} \in \mathbf{G}$ such that

$$\tilde{g} g \tilde{g}^{-1} = \bar{g}.$$

This gives rise to another equivalence relation and to a separation of the group in a set of *conjugacy classes*. If the group is Abelian, each element forms a class by itself. The *conjugate group* $\mathbf{H}_{\tilde{g}}$ with respect to a subgroup $\mathbf{H} \subset \mathbf{G}$ and to a group element $\tilde{g} \in \mathbf{G}$ is formed by all elements $\tilde{g} H \tilde{g}^{-1}$.

In the example with $\mathbf{H} = \{1, Ci\}$, we find for instance $\mathbf{H}_{(i)} = \{1, -Ci\}$.

Invariant/Normal Subgroups

If it happens that the conjugate group with respect to **H** and all elements in **G** are identically the same, **H** is called an *invariant* or *self-conjugate* or *normal* subgroup.

Definition: $\mathbf{N} \subseteq \mathbf{G}$ is an *invariant* subgroup iff $gN = Ng$ for all $g \in \mathbf{G}$. This can be interpreted as if **N** commutes with all $g \in \mathbf{G}$. By this definition (1) the trivial subgroups of **G** are invariant subgroups, (2) all subgroups of an Abelian group are invariant subgroups.

If **N** is an invariant subgroup, the quotient **G/N** is itself a group, as can be seen from

$$(g_1\mathrm{N}) \circ (g_2\mathrm{N}) = g_1\mathrm{N} \circ \mathrm{N}g_2 = g_1 \circ \mathrm{N}g_2 = (g1 \circ g_2)\mathrm{N}.$$

This group is called *factor group*.

Let us come back to the $\mathbf{D}_4$ example with the subgroup $\mathbf{H} = \{1, Ci\}$. This is not an invariant subgroup, since for instance

$$iH = \{i, iCi\} = \{i, -Ci^2\} = \{i, C\} \quad \text{but} \quad Hi = \{i, Ci^2\} = \{i, -C\}.$$

On the other hand $\mathbf{D}_2 = \{1, -1, C, -C\}$ is an invariant subgroup of $\mathbf{D}_4$. The full group can be written in terms of this subgroup as $\mathbf{D}_4 = 1\mathbf{D}_2 \cup i\mathbf{D}_2$. The factor group $\mathbf{D}_4/\mathbf{D}_2 = \{1\mathbf{D}_2, i\mathbf{D}_2\}$ is isomorphic to $\mathbf{Z}_2$ by the mappings $1\mathbf{D}_2 \mapsto 1, i\mathbf{D}_2 \mapsto -1$.

Simple and Semi-Simple Groups

With the previous introduction to subgroup structures we are finally able to define "simple":

(1) A group is *simple* if it has no non-trivial invariant subgroup.
(2) A group is *semi-simple* if it has no Abelian invariant subgroup. In other words: A semi-simple group may have a nontrivial invariant subgroup, but this is not allowed to be Abelian.

All simple finite groups have been known since 1982. All other finite groups are "products" of these in the following sense: A group **G** is a *direct product* of its subgroups **A**, **B** written as $\mathbf{G} = \mathbf{A} \times \mathbf{B}$ if

(1) for all $a_i \in \mathbf{A}, b_j \in \mathbf{B}$, it holds that $a_i \circ b_j = b_j \circ a_i$
(2) every element $g \in \mathrm{G}$ can be uniquely expressed as $g = a \circ b$.

Some consequences of this definition are

- $g_1 \circ g_2 = (a_1 \circ b_1) \circ (a_2 \circ b_2) = (a_1 \circ a_2) \circ (b_1 \circ b_2)$.
- Both **A** and **B** are invariant subgroups of **G**, e.g. for **A**

$$g_i A = (a_i \circ b_i)A = (b_i \circ a_i)A = b_i A \quad \text{and} \quad Ag_i = A(a_i \circ b_i) = Ab_i = b_i A.$$

- **G** can be written in terms of the subgroups **A** and **B** as

$$\mathbf{G} = \cup b_i A = \cup a_i B \qquad \text{such that} \qquad \mathbf{G/A} \cong \mathbf{B}, \qquad \mathbf{G/B} \cong \mathbf{A}.$$

- $\mathbf{G}$ is not simple.

For example $\mathbf{D}_2$ can be written as $\mathbf{D_2} = 1\mathbf{C_2} \cup C\ \mathbf{C_2}$, i.e. $\mathbf{D}_2/\mathbf{C}_2 = \mathbf{C}_2$, and also $\mathbf{D}_2 \cong \mathbf{C}_2 \times \mathbf{C}_2$.
Observe that in general a relation $\mathbf{G}_1 = \mathbf{G}_2/\mathbf{G}_3$ does not allow to conclude that $\mathbf{G}_2$ is a direct product of $\mathbf{G}_1$ and $\mathbf{G}_3$.

A.2 Lie Groups

A.2.1 Definitions and Examples

An *r-parameter Lie group*[3] $\mathbf{G}$ is a group which at the same time is an r–dimensional differentiable manifold[4]. This means that the group operations

$$\begin{aligned} C &: \mathbf{G} \times \mathbf{G} \to \mathbf{G}, & C(g, g') &= g \circ g' \\ I &: \mathbf{G} \to \mathbf{G} & I(g) &= g^{-1} \end{aligned}$$

are differentiable maps. The group elements of a Lie-group have the generic form

$$\mathbf{G} \ni g(\xi^1, \ldots \xi^r) =: g(\xi),$$

where the $\{\xi^j\}$ are (local) coordinates in the group manifold. If the group multiplication is written as $g(\xi) \circ g(\xi') = g(\xi''(\xi, \xi'))$, the Lie group property implicates that the functions $\xi''(\xi, \xi')$ are differentiable[5] in the parameters/coordinates.

All notions and results of abstract group theory as derived in the previous section do hold for Lie groups except those in which reference is made to finiteness and to the order of the group. This is especially true for the notions subgroups, cosets, conjugacy classes, invariant subgroups, quotient groups, etc.

In the following, several examples of Lie groups will be given, many of them of utmost importance for describing symmetries in fundamental physics. Most of these groups are not just abstract structures, but can be interpreted/realized as transformations in geometric spaces; therefore being called transformation groups.

Transformations of the Straight Line

The *translation group* is realized by the set of all displacements

$$T_a: \quad x' = x + a \qquad a \in \mathbb{R}.$$

---

[3] named after Sophus Lie (1842–1899), a Norwegian mathematician. Interesting enough, he arrived at these notions by investigating solutions of differential equations and using symmetry arguments in this context. These techniques are mentioned in Sect. 2.2.4.

[4] A Lie group is a subcategory of topological groups, namely as a topological space for which the group operation is a continuous mapping.

[5] Remarkably, due to the group structure the manifold is even analytic.

This is obviously an Abelian group with $T_b \circ T_a = T_{a+b}$. The *dilation group* is defined by its elements

$$D_A: \quad x' = Ax \qquad A \in \mathbb{R}, \qquad A \neq 0;$$

again an Abelian group. The condition $A \neq 0$ is imposed to guarantee the invertibility. The group is a local symmetry group if $A \in \mathbb{R}^+$. Both the translations and the dilations may be combined to the one-dimensional coordinate transformations $x' = Ax + a$ constituting a two-parameter continuous group. In writing its elements as (A, a), the group multiplication becomes $(A, a) \circ (A', a') = (AA', a + Aa')$. The neutral element is $(1, 0)$, and the inverse is $(A, a)^{-1} := A^{-1}(1, -a)$. This group is a subgroup of the *linear fractional transformations*

$$x' = \frac{a_{11}x + a_{12}}{a_{21}x + a_{22}} \qquad a_{11}\, a_{22} - a_{12}\, a_{21} \neq 0.$$

The identity transformation is recovered for $a_{12} = 0 = a_{21}$, $a_{11} = a_{22}$ and we can write this three-parameter subgroup as

$$x' = \frac{a_1 x + a_2}{a_3 x + 1} \qquad a_1 \neq a_2 a_3.$$

This is the *projective group* of the straight line.

Groups in the Plane

The most general transformations mapping any straight line in a plane again into a straight line are

$$x' = \frac{Ax + By + C}{Lx + My + N} \qquad y' = \frac{Dx + Ey + F}{Lx + My + N}.$$

These transformations realize the projective group in the plane. The parameter are not completely independent, because one insists on invertibility. Further, the subgroup containing the identity transformation has $N \neq 0$, and thus there are eight essential parameter. This projective group has as important subgroups, for instance, translations $x' = x + a$, $y' = y + b$, dilations $x' = e^{\kappa}x$, $y' = e^{\lambda}y$, general linear transformations $x' = a_{11}x + a_{12}y$, $y' = a_{21}x + a_{22}y$, and among these rotations $x' = x\cos\alpha + y\sin\alpha$, $y' = y\cos\alpha - x\sin\alpha$, and special conformal transformations $x' = x(1 - ax)^{-1}$, $y' = y(1 - ax)^{-1}$. A further prominent group of transformation (providing mappings between circles and straight lines) is given by the inversion with respect to a circle of radius R centered for example in the origin;

$$I_R: \qquad x' = R^2 \frac{x}{r^2} \qquad y' = R^2 \frac{y}{r^2} \qquad \text{with} \qquad r^2 = x^2 + y^2.$$

From inversions I_1 and translations T_a (along the x-direction, say) one can build conformal transformations:

$$C_a = I_1 T_a I_1: \qquad x' = \frac{x + ar^2}{1 + 2ax + a^2r^2} \qquad y' = \frac{y}{1 + 2ax + a^2r^2}.$$

Matrix Groups

The Lie group $\mathbf{GL(n}, \mathbb{F})$ is given by the set of all nonsingular $(n \times n)$ matrices with values in the field $\mathbb{F}$

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \cdot & & \cdot \\ a_{n1} & \dots & a_{n1} \end{pmatrix}, \qquad a_{ik} \in \mathbb{F}, \qquad \det A \neq 0,$$

and the group operation realized as matrix multiplication. If there are no further restrictions on the matrices, $\mathbf{GL}(n, \mathbb{R})$ is an n^2 parameter and $\mathbf{GL}(n, \mathbb{C})$ a $2n^2$ parameter group. These matrices describe linear transformations in an n-dimensional real or complex vector space as: $z'_i = \sum_k a_{ik} z_k$. The matrices with positive determinant constitute the Lie subgroup $\mathbf{SL(n}, \mathbb{F})$.

In fundamental physics, especially regarding symmetries, the matrix groups given below are of major relevance:

- *orthogonal groups*
  The group $\mathbf{O(n)} \subset \mathbf{GL(n}, \mathbb{R})$ contains all matrices for which

  $$AA^T = \mathbf{1}, \quad \text{i.e.} \quad A_{ik}A_{kj} = \delta_{ij}$$

  where $(A^T)_{ik} = A_{ki}$ is the transposed matrix of A. Because of this restriction, $\mathbf{O(n)}$ is an $\frac{n(n-1)}{2}$ -parameter group. $\mathbf{O(n)}$ characterizes the symmetry of the sphere, with n real coordinates, in the sense that the "radius"

  $$x_1^2 + x_2^2 + \ldots + x_n^2$$

  is invariant with respect to linear transformations with matrices from $\mathbf{O(n)}$.
  A special case is the subgroup $\mathbf{SO(n)}$, which contains all matrices of $\mathbf{O(n)}$ with $\det A = 1$. Since as a set, $\mathbf{O(n)}$ has elements with either $\det A = 1$ or $\det A = -1$ (this follows from the defining condition) it can be understood as the quotient

  $$\mathbf{O(n)} = \mathbf{SO(n)}/\mathbf{Z}_2.$$

  As will be seen later, from the classification of groups (or rather, their associated algebras) one needs to distinguish even and odd n, because these differ in their spinor representations.
- *unitary groups*
  The group $\mathbf{U(n)} \subset \mathbf{GL(n}, \mathbb{C})$ contains all matrices for which

  $$AA^\dagger = \mathbf{1} \qquad \text{with} \qquad A^\dagger := (A^T)^*.$$

  $\mathbf{U(n)}$ has n^2 parameter. It is the invariance group of the sphere with n complex coordinates. A special case is the $(n^2 - 1)$-parameter group $\mathbf{SU(n)}$, in which all matrices have determinants equal to 1. It leaves invariant the real quadratic form

  $$z_1 z_1^* + z_2 z_2^* + \ldots + z_n z_n^*$$

  where the z_i are complex variables.

- *pseudo-orthogonal groups*
  $\mathbf{O(n, m)}$ leaves invariant $\sum_{ij} \eta_{ij} x^i x^j$, with

$$\eta_{ij} = \delta_{ij} \quad \text{for } i = 1, \ldots, n \qquad\qquad \eta_{ij} = -\delta_{ij} \quad \text{for } i = n+1, \ldots, m.$$

  The further restriction on uni-modularity ($\det A = 1$), leads to the $\mathbf{SO(n, m)}$ groups. Prominent examples in fundamental physics are the Lorentz group $\mathbf{SO(3, 1)}$, the de Sitter group $\mathbf{SO(4, 1)}$ and the conformal group $\mathbf{SO(4, 2)}$.
- *symplectic groups*
  Matrices of the group $\mathbf{Sp(2\,m)}$ are those elements A of the general linear groups which obey the condition

$$A^T \Gamma A = \Gamma, \qquad \Gamma := \begin{pmatrix} 0 & 1_m \\ -1_m & 0 \end{pmatrix}.$$

  The transformations generated by $\mathbf{Sp(2\,m)}$ leave the real antisymmetric quadratic form $\sum x_i \Gamma_{ij} y_j$ invariant. These groups typically arise in the phase-space formulation of physical systems, for instance in defining canonical transformations.

Inhomogeneous Linear Transformations

The transformation $x' = Ax + a$ is a mixture of a linear transformation (mediated by a real number A) and a translation (described by the parameter a) in one dimension. This can be generalized to higher dimensions: Let $\mathbf{G} = \mathbf{GL(n}, \mathbb{R})\ni A$, and $\mathbb{R}^n \ni a$ be the n-dimensional real space with addition of vectors. Define

$$\mathbf{IG} := \{(A, a) |\ A \in \mathbf{G},\ \ a \in \mathbb{R}^n\}$$

with the group operation

$$(A, a) \circ (A', a') := (AA', a + Aa').$$

One discovers $\mathbf{IG} \subset \mathbf{GL(n+1}, \mathbb{R})$ by the identification

$$(A, a) \mapsto \begin{pmatrix} A & a \\ 0^T & 1 \end{pmatrix}.$$

Another notation is $\mathbf{IG} = \mathbf{G} \ltimes \mathbb{R}^n$ in terms of the *semi-direct product* $\ltimes$.

Groups of Motion

The *group of motion* of a geometry, also called *isometry group* is a transformation group that leaves distances invariant. In one dimension this is just the translation group. For the Euclidean plane it is the group of inhomogeneous linear transformations composed out of rotations and translations. It is the largest group that changes only the location and orientation of a figure (triangle, circle, etc.) but not its magnitude and shape. For $\mathbb{E}^3$ the transformations of its group of motion have the form $\vec{x}' = O(\vec{x}) + \vec{a}$ with an $\mathbf{SO(3)}$ matrix O. For Minkowski space it is the Poincaré group as the semi-direct product of the Lorentz group with the group of spacetime translations.

A.2.2 Generators of a Lie Group

The generators of a Lie group may be defined through the tangent space of the group manifold near the neutral element g_0. One can always choose a chart with coordinates $\{\xi^i\}$ such that $g(\xi = 0) = g_0$. A Taylor expansion for the group elements near the identity is of the form

$$g(\xi) = g_0 + i\xi^a X^a + \mathcal{O}(\xi^2) \qquad \xi \in \mathbb{R}.$$

The X^a are called the infinitesimal *generators* of the Lie group[6]. Clearly,

$$X^a = -i \frac{\partial g}{\partial \xi^a}\bigg|_{\xi=0}. \tag{A.1}$$

In some cases, it is possible to construct the complete group from its infinitesimal generators. Consider for instance a one-parameter Lie group. Because of the analyticity of $g(\xi)$, it is possible to choose a parametrization in which $\xi''(\xi, \xi') = \xi + \xi'$. We may now attempt to construct finite group elements by making an infinite series of infinitesimally small steps away from the group identity:

$$g(\xi) = g(\xi/n)^n = \lim_{n\to\infty}\left(1 + \frac{i\xi X}{n}\right)^n = \exp(i\xi X).$$

This can be extended to an r-parameter Lie group if it is Abelian

$$g(\xi^1, \ldots, \xi^r) = \exp(i\xi^a X^a). \tag{A.2}$$

In case of a non-Abelian Lie group the expression (A.2) is generically not valid, but is only valid near the identity; otherwise one needs to invoke the Campbell-Baker-Hausdorff formula, see (A.6) below.

As an example consider again the 2-parameter group of coordinate transformations $x \to (1 + \xi^1)x + \xi^2$. A function $f(x)$ changes into

$$f(x) \to f((1 + \xi^1)x + \xi^2) = f(x) + \xi^1 x \frac{d}{dx} f(x) + \xi^2 \frac{d}{dx} f(x) + \mathcal{O}(\xi^2)$$

---

[6] Singling out a factor i here is merely a convention. This convention is chosen throughout this book, since from the point of view of quantum physics the generators X are to be interpreted as representations of Hermitean operators.

from which we read the two infinitesimal generators

$$X^1 = -ix\frac{d}{dx}, \;\; X^2 = -i\frac{d}{dx}.$$

Taking the commutator of these generators, we calculate

$$(X^1X^2 - X^2X^1)\; f(x) = -(x\frac{d}{dx}\frac{d}{dx} - \frac{d}{dx}x\frac{d}{dx})f(x) = \frac{d}{dx}f(x) = iX^2\; f(x).$$

The commutator of the two infinitesimal generators is again an infinitesimal generator.

As a second example, consider rotations in a plane, characterized by an angle θ. In polar coordinates, a clockwise rotation amounts to

$$\begin{aligned} x = r\cos\theta \rightarrow \qquad & r\cos(\theta-\xi) = r\cos\theta + \xi r\sin\theta + \mathcal{O}(\xi^2) = x + \xi y \\ y = r\sin\theta \rightarrow \qquad & r\cos(\theta-\xi) = r\sin\theta - \xi r\cos\theta + \mathcal{O}(\xi^2) = y - \xi x. \end{aligned}$$

Thus the generator can be written as a 2×2 matrix

$$X = -i\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{A.3}$$

with $X^2 = I$. The exponential map (A.2) gives

$$\begin{aligned} g(\xi) &= \sum_{n=0}^{\infty}\frac{1}{n!}\xi^n(iX)^n = \sum_{n=0}^{\infty}\left\{\frac{1}{2n!}(-\xi^2)^n I + \frac{1}{(2n+1)!}(-1)^n\xi^{2n+1}iX\right\} \\ &= \cos\xi I + i\sin\xi X = \begin{pmatrix} \cos\xi & \sin\xi \\ -\sin\xi & \cos\xi \end{pmatrix} \end{aligned}$$

which indeed is simply a two-dimensional rotation.

This can be extended to rotations in three dimensions: Think of the infinitesimal rotation (A.3) as being a rotation in the x-y plane around the z-axis with an infinitesimal angle $\xi^3 = \xi$. Any rotation in three dimensions can be constructed from rotations around the x-, y-, and z-axis with angles (ξ^1, ξ^2, ξ^3), according to

$$\begin{aligned} &\vec{r} \rightarrow \vec{r} + \vec{r}\times\vec{\xi} + \mathcal{O}(|\xi|^2) = \vec{r} + i\xi^a X^a\vec{r} \\ &\qquad X^1 = -i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} X^2 = -i\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} X^3 = -i\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{A.4}$$

A.2.3 Lie Algebra Associated to a Lie Group

The infinitesimal generators of a Lie group are local representatives of the group elements. So far, we have not investigated how the defining attributes of a group become manifest in features of the generators. Firstly, because of the closure property of a group, there must exist a relation

$$g(\xi^1, \ldots, \xi^r) \circ g(\bar{\xi}^1, \ldots, \bar{\xi}^r) = \exp(i\xi^a X^a) \circ \exp(i\bar{\xi}^a X^a)$$
$$= \exp(i\bar{\bar{\xi}}^a X^a) = g(\bar{\bar{\xi}}^1, \ldots, \bar{\bar{\xi}}^r). \quad \text{(A.5)}$$

Now the product of the two exponentials can be expressed by the *Campbell-Baker-Hausdorff* formula:

$$e^A e^B = e^{A+B+C(A,B)}$$
$$C(A, B) = \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B]] + \frac{1}{12}[[A, B], B] + \ldots.., \quad \text{(A.6)}$$

where the dots indicate the possible presence of higher-order commutators of A and B. Therefore, group closure (A.5) demands that the commutator of two matrices/generators is itself a generator:

$$[X^a, X^b] = X^a X^b - X^b X^a = i f^{abc} X^c \qquad f^{abc} \in \mathbb{R}. \quad \text{(A.7)}$$

The constants f^{abc} are called the *structure constants* of the group. Because of the antisymmetry of the commutator these are of course antisymmetric in their first two indices:

$$f^{abc} = -f^{bac}.$$

A further identity among the structure constants originates from the commutator identity

$$[[A, B], C] + [[B, C], A] + [[C, A], B] \equiv 0$$

resulting in the *Jacobi identity*

$$\sum_e (f^{abe} f^{ecd} + f^{bce} f^{ead} + f^{cae} f^{ebd}) \equiv 0. \quad \text{(A.8)}$$

Given these relations, one can easily see that the infinitesimal generators of a Lie group constitute a Lie algebra (in the sense introduced in Sect. A.1.1) if the $\Box$ operator is realized by the commutator, that is if $v_1 \Box v_2 := [v_1, v_2]$.

For an Abelian (sub)group all the structure coefficients vanish. Since for an invariant subgroup with generators Y^b, it holds that

$$\exp(i\xi^a X^a) \circ \exp(i\xi^b Y^b) = \exp(i\xi^{b'} Y'^{b'}) \circ \exp(i\xi^a X^a)$$

the Campbell–Baker–Hausdorff formula shows that

$$[X^a, Y^b] = f^{abb'} Y^{b'}. \quad \text{(A.9)}$$

The Lie algebra associated to a Lie group **G** is by convention denoted by **g**. Lie groups with the "same" Lie algebra are locally isomorphic, that is isomorphic near the neutral element:

$$\mathbf{g} = \hat{\mathbf{g}} \qquad \Leftrightarrow \qquad \mathbf{G} \cong \hat{\mathbf{G}}\mid_{g_0} .$$

Be aware that in giving meaning to "same", linear combinations $X'^a = \Gamma^{ab} X^b$ of generators are again generators. The algebra now reads

$$[X'^a, X'^b] = f'^{abc} X'^c,$$

where the new structure constants f'^{abc} are linear combinations of the old ones:

$$f'^{abc} = \Gamma^{ad}\Gamma^{be} \; f^{deg} \; (\Gamma^{-1})^{gc}. \tag{A.10}$$

This might mimic another algebra. The techniques to identify all non–isomorphic Lie algebras make use of 'root spaces' and 'Dynkin diagrams', techniques that will not be covered here.

Examples

- Lie algebra of $\mathbf{SO(3)}$
  The Lie algebra is directly derived from the explicit representation of the generators as given by (A.4):

$$[X^a, X^b] = i\epsilon^{abc} X^c. \tag{A.11}$$

  This is the same as the algebra of the three angular momenta $\{J^1, J^2, J^3\}$, as we expect them to arise from rotational symmetry.
- Lie algebra of $\mathbf{SU(2)}$
  $\mathbf{SU(2)}$ is a three-dimensional group. The requirement that the group elements $g(\xi) = \exp(i\xi^a X^a)$ are unitary matrices with determinant $+1$ demands that

$$X^a = (X^a)^\dagger \qquad \text{and} \qquad \mathrm{tr}(X^a) = 0.$$

  In the physical literature, it is common to choose $X^a = \frac{1}{2}\sigma^a$ with the *Pauli matrices*

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A.12}$$

  Since the Pauli matrices do have the property $\sigma^a\sigma^b = i\epsilon^{abc}\sigma^c + \delta^{ab} I$, the $\mathbf{su(2)}$ commutators are

$$[X^a, X^b] = i\epsilon^{abc} X^c.$$

This is identical to the $\mathfrak{so}(3)$ algebra (A.11). Therefore the Lie groups $\mathbf{SU(2)}$ and $\mathbf{SO(3)}$ are locally isomorphic. Globally, however, both groups differ, and in Sect. A.3.5 $\mathbf{SU(2)}$ will be identified as the 'universal covering group' of $\mathbf{SO(3)}$. This mathematical subtlety becomes physically manifest in the existence of half-integer spin particles, as explained in Sect. 5.3. Here, we keep in mind the representation of $\mathbf{SU(2)}$ group elements in terms of the Pauli matrices as

$$g(\vec{\alpha}) = \exp\left(\frac{i}{2}\vec{\sigma}\cdot\vec{\alpha}\right). \tag{A.13}$$

- Lie algebra of $\mathbf{SO(2,1)}$
  If one determines the $\mathfrak{so}(2,1)$ according to the previously specified recipe, one finds

$$[X^1, X^2] = iX^3, \qquad [X^2, X^3] = -iX^1, \qquad [X^3, X^1] = iX^2.$$

  There is no choice of linear combinations of these X^a by which one regains the algebra of the rotation group. Thus $\mathbf{SO(2,1)}$ and $\mathbf{SO(3)}$ are different as groups. But one would also find that $\mathfrak{so}(2,1) \equiv \mathfrak{so}(1,2)$ and thus–at least locally–$\mathbf{SO(2,1)} \cong \mathbf{SO(1,2)}$.
- Lie algebra of the 2-dimensional Poincaré group
  The group elements of the Poincaré group in one spatial and one temporal coordinate are parametrized through the rapidity η (characterizing the Lorentz boost, see Sect. 3.1.2) and the constant shifts (a,τ) in spacetime:

$$g(\eta, a, \tau) = \begin{pmatrix} \cosh\eta & \sinh\eta & a \\ \sinh\eta & \cosh\eta & \tau \\ 0 & 0 & 1 \end{pmatrix} \tag{A.14}$$

  from which by (A.1) we obtain the generators

$$X^\eta = -i\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X^a = -i\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X^\tau = -i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

  These obey the $\mathfrak{iso}(2,1)$ algebra

$$[X^\eta, X^a] = iX^\tau, \qquad [X^\eta, X^\tau] = iX^a, \qquad [X^a, X^\tau] = -iX^a.$$

- $\mathfrak{so}(\mathfrak{n},\mathfrak{m})$
  The algebra of the pseudo-orthogonal group $\mathbf{SO(n,m)}$ with $(n+m=D)$ is spanned by $\frac{D(D-1)}{2}$ generators $M^{AB} = -M^{BA}$, where the indices A, B range from 1 to D. The Lie algebra $\mathfrak{so}(\mathfrak{n},\mathfrak{m})$ is

$$[M^{AB}, M^{CD}] = i(\eta^{BC}M^{AD} - \eta^{BD}M^{AC} + \eta^{AD}M^{BC} - \eta^{AC}M^{BD}). \tag{A.15}$$

Here $\eta^{AB} = \delta^{AB}$ for $(A = 1, \ldots, n)$, $\eta^{AB} = -\delta^{AB}$ for $(A = n+1, \ldots, m)$ (all others vanishing). In coordinate space (that is as pseudo-rotations) the generators can be realized in the form

$$M^{AB} = i(x^A \partial^B - x^B \partial^A). \tag{A.16}$$

Important examples (in D=4) are the Lorentz algebra $\mathfrak{so}(3,1)$, the de Sitter algebra $\mathfrak{so}(4,1)$, the anti-de Sitter algebra $\mathfrak{so}(3,2)$, and the conformal algebra $\mathfrak{so}(4,2)$.

A.2.4 Inönü–Wigner Contraction of Lie Groups

"The process of group contraction is a method for obtaining from a given Lie group another, non-isomorphic Lie group through a limit operation." (from [288]). This was originally investigated by E. Inönü and E. Wigner [289], motivated by the desire to establish a mathematically sound derivation of the Galilei algebra from the Poincaré algebra (see Sect. 3.4). They found that every Lie group can be contracted with respect to any of its Lie subgroups–and only with respect to these. Technically one can describe the construction of the contracted group in terms of local algebras.

Consider a Lie algebra L with a Lie subalgebra L'. Assume that $\{X_1, \ldots, X_n\}$ is a basis in L, and $\{X'_1, \ldots, X'_{n'}\}$ is a basis in L'. The Lie algebra is explicitly of the form[7]

$$\begin{aligned}
[X_a, X_b] &= i f_{ab}{}^c X_c \; + \; i f_{ab}{}^{c'} X'_{c'} \\
[X_a, X'_{b'}] &= i f_{ab'}{}^c X_c \; + \; i f_{ab'}{}^{c'} X'_{c'} \\
[X'_{a'}, X'_{b'}] &= i f_{a'b'}{}^{c'} X'_{c'}.
\end{aligned}$$

where the latter reflects that the X' span a subalgebra. Define rescaled generators as

$$Y_a := \epsilon X_a, \qquad Y'_{a'} := X'_{a'}$$

which fulfill the algebra

$$\begin{aligned}
[Y_a, Y_b] &= \epsilon^{-1} i f_{ab}{}^c Y_c \; + \; \epsilon^{-2}\, i f_{ab}{}^{c'} Y'_{c'} \\
[Y_a, Y'_{b'}] &= i f_{ab'}{}^c Y_c \; + \; \epsilon^{-1} i f_{ab'}{}^{c'} Y'_{c'} \\
[Y'_{a'}, Y'_{b'}] &= i f_{a'b'}{}^{c'} Y_{c'}.
\end{aligned}$$

Now the limit $\epsilon \to 0$ can be taken only if all $f_{ab}{}^c, f_{ab}{}^{c'}, f_{ab'}{}^{c'}$ vanish identically. This results in the contracted algebra L_c:

$$[Y_a, Y_b] = 0 \tag{A.17a}$$

$$[Y_a, Y'_{b'}] = i f_{ab'}{}^c Y_c \tag{A.17b}$$

$$[Y'_{a'}, Y'_{b'}] = i f_{a'b'}{}^{c'} Y'_{c'}. \tag{A.17c}$$

---

[7] Here, I prefer to write the infinitesimal generators with lower indices and the structure constants with mixed indices. This is only to make the expressions more comprehensible.

Looking at the associated Lie groups, we see that if a Lie group $\mathbf{G}$ is contracted with respect to one of its Lie subgroups $\mathbf{G}'$, one arrives at a contracted Lie group $\mathbf{G_c}$. This group has the same dimension as its "mother" group. $\mathbf{G_c}$ contains a subgroup which is isomorphic to $\mathbf{G}'$ and another Abelian subgroup $\bar{\mathbf{G}}$ spanned by the contracted generators X_a. $\bar{\mathbf{G}}$ is an invariant subgroup of $\mathbf{G}'$, and thus $\mathbf{G}' \sim \mathbf{G_c}/\bar{\mathbf{G}}$. Inönü and Wigner also showed that the necessary condition for a group $\mathbf{G_c}$ to be derivable from another group $\mathbf{G}$ by contraction is the existence of an Abelian invariant subgroup in $\mathbf{G_c}$ and the possibility of choosing from each of its cosets an element so that these form a subgroup.

Following [288], the technique and the consequences of group contractions shall be made explicit with two examples:

1. contraction of the 3-dim rotation group $\mathbf{SO(3)}$
   We start from the algebra L of the generators X^a which generate rotations around the x-, y-, and z-axis according to (A.11): $[X^a, X^b] = i\epsilon^{abc}X^c$. Take as L' the generator X^3 for the one-parameter subgroup of rotations around the z-axis. Then with the procedure explained before, that is defining $Y^3 = X^3$, $Y^1 = \epsilon X^1$, $Y^2 = \epsilon X^2$ and taking the limit $\epsilon \to 0$ we obtain the algebra L_c
$$[Y^1, Y^2] = 0 \tag{A.18a}$$
$$[Y^2, Y^3] = iY^1 \tag{A.18b}$$
$$[Y^3, Y^1] = iY^2. \tag{A.18c}$$
   This is isomorphic to the algebra of $\mathbf{E_2}$, the Euclidean group in two dimensions consisting of translations (generators Y^1 and Y^2) and rotations (generator Y^3) in the plane.
2. contraction of the 3-dim Lorentz-group
   In splitting the set of generators in $\mathfrak{so}(3,1)$ as M^{ij} and M^{0i} $(i = 1, 2, 3)$ one rewrites the algebra (A.15)
$$\begin{aligned}
[M^{ij}, M^{kl}] &= i(\eta^{jk}M^{il} - \eta^{jl}M^{ik} + \eta^{il}M^{jk} - \eta^{ik}M^{jl}) \\
[M^{0j}, M^{kl}] &= i(\eta^{jk}M^{0l} - \eta^{jl}M^{0k}) \\
[M^{0i}, M^{0j}] &= iM^{ij}.
\end{aligned}$$
   Taking for L' the subalgebra generated by the M^{ij} and defining
$$J^i = \frac{1}{2}\epsilon^{ikl}M^{kl}, \qquad G^i = \epsilon M^{0i}$$
   results in the algebra
$$\begin{aligned}
[J^i, J^j] &= i\epsilon^{ijk}J^k \\
[J^i, G^j] &= i\epsilon^{ijk}G^k \\
[G^i, G^j] &= 0.
\end{aligned}$$
   This is the algebra of the homogeneous Galilei group made up of three-dimensional rotations and Galilei boosts; compare (2.75).

Table A.2 Cartan classification

Class	Range	Dimension	Compact group
$\mathbf{A_n}$	$n \geqslant 1$	n(n+2)	$\mathbf{SU(n+1)}$
$\mathbf{B_n}$	$n \geqslant 2$	n(2n+1)	$\mathbf{SO(2n+1)}$
$\mathbf{C_n}$	$n \geqslant 3$	n(2n-1)	$\mathbf{Sp(2n)}$
$\mathbf{D_n}$	$n \geqslant 4$	n(2n-1)	$\mathbf{SO(2n)}$
$\mathbf{E_n}$	$n = 6, 7, 8$	78, 133, 248	$\mathbf{E_n}$
$\mathbf{F_4}$	$n = 4$	52	$\mathbf{F_4}$
$\mathbf{G_2}$	$n = 2$	14	$\mathbf{G_2}$

A.2.5 Classification of Lie Groups

Strangely enough the classification of Lie groups (i.e. the reduction of an arbitrary Lie-group to its "prime" factors) was completed already in 1894 by Élie Cartan, and this was nearly a century earlier then the analogous classification for finite groups, which as mentioned above was completed only in 1982.

Cartan Classification

The classification of all simple Lie-groups by Cartan[8] is as follows: There are four infinite series of groups and five exceptional groups. The series are $\mathbf{A_n}$, $\mathbf{B_n}$, $\mathbf{C_n}$, $\mathbf{D_n}$. The subscript n denotes the *rank* of the group. The groups in the Cartan series are isomorphic to the groups introduced in Sect. A.2.1 as shown in Table A.2. The value n may take on smaller values than the limits indicated in the table, but because of mutual (local) isomorphisms these are not independent of the others given. For example $\mathbf{B_1} \cong \mathbf{A_1}$, $\mathbf{D_2} \cong \mathbf{B_1} \oplus \mathbf{B_1}$, $\mathbf{D_3} \cong \mathbf{A_3}$. The exceptional groups are called $\mathbf{G_2}$, $\mathbf{F_4}$, $\mathbf{E_6}$, $\mathbf{E_7}$, $\mathbf{E_8}$. The transformations generated by these groups typically leave higher than quadratic polynomials invariant. The exceptional groups are indeed every now and then considered in fundamental physics (for instance in "Grand Unified Theories" or in string models). We note that

$$\mathbf{E}_8 \supset \mathbf{E}_7 \supset \mathbf{E}_6 \supset \mathbf{SO(10)}.$$

The classification of Cartan includes both compact and non-compact groups, a topic to be clarified next. Thus, for instance, $\mathbf{A_{n-1}}$ stands for the series of groups $\mathbf{SU(p, n-p)}$.

[8] Actually, the classification was already published by W. Killing some years before. He however had a further exceptional group, which Cartan showed to be isomorphic to $\mathbf{G_2}$. Wilhelm Killing (1847–1923) was a professor of mathematics in Münster, a town in Lower Saxony of Germany. His name also continues to be used in the term 'Killing symmetries', which gives a precise meaning to symmetries of metrics independent of their coordinate expressions.

The problem of finding all possible Lie groups can be transferred to finding the classification of Lie algebras since all group elements that are continuosly connected to the identity can be generated by exponentiation.

Compact Lie Groups

The classification program of Lie groups is largely similar to the classification of finite groups, and it is just the same if the group in question is compact. A Lie group is *compact* if the space of parameters ξ^a is bounded ($u^a \leq \xi^a < o^a$) and closed, that is, the limit of every convergent sequence of parameters is itself a parameter value.

- The rotation group $\mathbf{SO(n)}$ is compact.
- The translation group $\mathbb{R}^n$ is not compact, since it is not bounded.
- The Lorentz group is not compact, since it is not closed ($v = c$ is not a valid parameter).

Cartan-Killing Metric

An extremely useful concept in the classification and representation theory of Lie algebras is the Cartan-Killing metric, which can be expressed directly by the structure constants as

$$g^{ab} = \sum_{c,d} f^{acd} f^{bdc}. \tag{A.19}$$

The Cartan-Killing metric provides criteria for compactness and semi–simplicity of a Lie algebra:

- A Lie algebra is semi-simple if the metric is nonsingular ($\det g \neq 0$); otherwise the Lie algebra is not semi-simple.
- A Lie algebra is compact if the metric is negative definite; that is if for every $A = \alpha_a X^a$, we have $|A, A| = g^{ab}\, \alpha_a \alpha_b < 0$.

Let us understand these notions by resorting to examples of Lie groups and their Lie algebras which were already discussed before. These are the three-dimensional groups $\mathbf{SO(3)}$, $\mathbf{SO(2,1)}$, $\mathbf{E_2}$. The first two are simple, while $\mathbf{E_2}$ is not semi-simple. This can be seen from their respective Lie algebras. For an invariant subalgebra with generators X' (A.9) requires that its commutator with any other generator is contained in the subalgebra; in short $[X, X'] \Rightarrow X'$. There are Abelian one-parameter subgroups in $\mathbf{SO(3)}$ and in $\mathbf{SO(2,1)}$ generated by any of the X^a, but their commutators with any of the other generators are not elements of the respective subalgebras. Hence these two groups are simple. The situation is different for $\mathbf{E_2}$. Its Abelian subalgebra, generating the translations Y^1, Y^2 (see (A.18)), is an invariant algebra; thus this Lie group is not semi-simple. The non-vanishing structure coefficients for the three groups are aggregated in Table A.3.

Table A.3 Groups and (nonvanishing) structure constants

Group	f^{123}	f^{231}	f^{312}
SO(3)	1	1	1
SO(2, 1)	1	-1	1
$\mathbf{E_2}$	0	1	1

For the respective Cartan-Killing metric one finds

$$G_{\mathbf{SO(3)}} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad G_{\mathbf{SO(2,1)}} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad G_{\mathbf{E_2}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Thus indeed the metrics are regular for the two simple groups, and singular for the non-semi-simple one. Furthermore, the metric is negative definite for the compact group **SO(3)**.

For compact semi-simple Lie groups, one can always find a basis X^a in the algebra such that the structure constants in $[X^a, X^b] = i\ f^{abc} X^c$ are totally antisymmetric in their indices.

Casimir Operators

In quantum mechanics, use is made of the fact that $J^2 = (J_1)^2 + (J_2)^2 + (J_3)^2$ commutes with each of the $(J_i)^2$. In general terms, J^2 is a Casimir operator of the group **SO(3)**. A Casimir operator generically is a polynomial in the generators that commutes with all generators. For a semi-simple algebra the number of independent Casimir operators is equal to the rank of the algebra. Notice that the Casimir operators are not part of the Lie algebra–in mathematical terms they belong to the enveloping algebra.

For semi-simple Lie groups, the Cartan-Killing metric can be used to calculate the quadratic Casimir operator of the group by

$$C := g_{ab} X^a X^b$$

since the Cartan-Killing metric has an inverse g_{ab}. Because this is a real symmetric matrix, one can always bring it to the form $C = X^a X^a$ by an orthonormal transformation of the generators.

A.2.6 Infinite-Dimensional Lie Groups

Kac-Moody Algebra

Kac–Moody algebras are a generalization of finite-dimensional semi-simple Lie algebras defined by

$$\left[X_m^a, X_n^b\right] = i f^{abc} X_{m+n}^c + c\delta^{ab}\delta_{m,-n}.$$

Here the $f's$ are the structure constants of a finite-dimensional Lie-algebra and the constant c is a central term. Kac-Moody algebras typically arise in conformal field theories and in dimensional reductions of gravity and supergravity theories.

A related algebra is the Virasoro algebra

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}m(m^2-1)\delta_{m,-n}.$$

This algebra generates the diffeomorphisms of the circle. It turns up in string theory together with further generators for fermionic degrees of freedom.

Diffeomorphism Group

The diffeomorphism group arises in all theories which are reparametrization invariant, or generally covariant. Examples are the relativistic point particle, the relativistic string and general relativity.

Diffeomorphisms are smooth, invertible mappings of a D-dimensional manifold M onto itself: $d : M \to M$. These form a group **Diff**(**M**). In a coordinate patch, the mapping is

$$d : x \to d(x) = x'(x) \qquad \text{or locally} \qquad x^\mu \to x^\mu + \xi^\mu.$$

The Lie algebra associated to **Diff**(**M**) is spanned by generators

$$L_\xi(x) = \xi^\mu(x)\partial_\mu$$

that is by the set of all vector fields on the manifold M. On any tensor T, a diffeomorphism acts as the Lie derivative $\delta T = -\pounds_\xi T$. The "structure constants" of the diffeomorphism group are derived from the commutator of two Lie derivatives

$$[\pounds_X, \pounds_Y]T = \pounds_{[X,Y]}T$$

where $[X, Y] = \pounds_X Y = -\pounds_Y X$. Then (in the condensed DeWitt notation)

$$\int d^D x' \int d^D x'' C^\mu{}_{\nu'\rho''} X^{\nu'} Y^{\rho''} = -[X, Y]^\mu = X^\mu_{,\nu} Y^\nu - Y^\mu_{,\nu} X^\nu$$

from which

$$C^\mu{}_{\nu'\rho''} = \delta^\mu_{\nu',\sigma}\delta^\sigma_{\rho''} - \delta^\mu_{\rho'',\sigma}\delta^\sigma_{\nu'} = \delta^\mu_{\nu';\sigma}\delta^\sigma_{\rho''} - \delta^\mu_{\rho'';\sigma}\delta^\sigma_{\nu'} \tag{A.20}$$

with

$$\delta^\sigma_{\rho'} = \delta^\sigma_\rho \delta^D(x, x'), \qquad \delta^\mu_{\nu',\sigma} = \delta^\mu_\nu \partial_\sigma \delta^D(x, x').$$

A.3 Representation of Groups

As elaborated in the main text, quantum physics shifts the attention from groups to their representations (or even more precisely, to ray representations; see Sect. A.3.5).

A.3.1 Definitions and Examples

The n-dimensional *representation* $D^{(n)}$ of a group $\mathbf{G}$ is a homomorphism

$$D^{(n)} : \mathbf{G} \to \mathbf{L}(V^n)$$

where $\mathbf{L}(V^n)$ is a group of linear operators that act on an n-dimensional vector space V^n. In physical applications the vector space is mainly a Hilbert space. If the operators are realized as matrices the representation is a *matrix representation*. The mapping D obeys the postulates of a group homomorphism: If $D(g_i) \in \mathbf{GL}(\mathbf{n}, \mathbb{F})$ it fulfills

$$D(g_i) \star D(g_k) = D(g_i \circ g_k), \qquad D(g_i) \star D(g_i{}^{-1}) = D(g_0) = I.$$

Here $\circ$ denotes the group multiplication in $\mathbf{G}$, and $\star$ is the group multiplication in $D(\mathbf{G}) \subset \mathbf{L}(V^n)$.

In order to motivate and illustrate notions introduced later, let us consider some real representations of the most rudimentary non-trivial group, written as $\mathbf{Z}_2 = \{e, a\}$:

$$e \circ a = a \circ e = a \qquad a \circ a^{-1} = e \qquad a \circ a = e \qquad \text{i.e.} \qquad a = a^{-1}.$$

- 1-dimensional representations:

$$D_1^{(1)}(e) = 1 \qquad D_1^{(1)}(a) = -1$$
$$D_2^{(1)}(e) = 1 \qquad D_2^{(1)}(a) = 1.$$

  The representation $D_1^{(1)}$ is an isomorphism (in the context of representations called a *faithful representation*).
- 2-dimensional representations:

$$D_1^{(2)}(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad D_1^{(2)}(a) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

  is the *trivial* representation,

$$D_2^{(2)}(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad D_2^{(2)}(a) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

  is a faithful representation.

$$D_3^{(2)}(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad D_3^{(2)}(a) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This representation is related to the representation $D_2^{(2)}$ by a *similarity transformation*

$$D_3^{(2)}(Z_2) = S \star D^{(2)}(Z_2) \star S^{-1} \qquad \text{with} \qquad S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

- 4-dimensional representations, for instance

$$D^{(4)}(e) = \begin{pmatrix} D_1^{(2)}(e) & 0 \\ 0 & D_2^{(2)}(e) \end{pmatrix} \qquad D^{(4)}(a) = \begin{pmatrix} D_1^{(2)}(a) & 0 \\ 0 & D_2^{(2)}(a) \end{pmatrix}.$$

This representation is–by construction–completely reducible into $D_1^{(2)}(Z_2)$ and $D_2^{(2)}(Z_2)$; one writes $D^{(4)} = D_1^{(2)} \oplus D_2^{(2)}$. Obviously, reducibility is equivalent to the fact that the action of $D^{(4)}$ on the 4-dimensional vector space leaves two 2-dimensional subspaces invariant.

A.3.2 Representations of Finite Groups

On the evidence of the previous example, there are many representations of a group. Some of them are related by similarity transformations, others are reducible to (dimensionally) smaller representations. Of course one is interested in the most basic (or "elementary", or "prime") ones. The theory of (finite) group representations was essentially developed by F.G. Frobenius and I. Schur in 1906. (Further refinements are due to W. Burnside and R. Brauer.) They found (1) a criterion for a representation to be irreducible, (2) the number and dimensions of inequivalent representations, (3) the important role played by the so-called character of the representation, and (4) how to construct the representation using the character table. These findings shall be outlined in the following

Some More Definitions…

- If $D(G)$ is a representation and if a matrix S exists such that $D'(G) = SD(G)S^{-1}$ is a representation, both representations are called *equivalent with similarity transformation* S. The further investigation can then be restricted to the equivalence classes of equivalent (sic!) representations.
- If a representation is equivalent to

$$D(g) = \begin{pmatrix} D_1(g) & A(g) \\ 0 & D_2(g) \end{pmatrix} \tag{A.21}$$

it is called *reducible*. Otherwise, if this triangle form cannot be attained, the representation is *irreducible*. If D_2 has dimension m, the representation $D(g)$ leaves an m-dimensional subspace invariant.

- If for a representation a similarity transformation exists such that in (A.21) $A(g) = 0$, the representation is *completely reducible*; one writes in this situation
$$D^{(n)} = \oplus\, m_j\; D^{(j)}; \tag{A.22}$$
the integers m_j indicating how often the respective representation appears in this decomposition.

…and Some Theorems

- Every representation of a finite group is equivalent to a unitary representation, i.e. $D^{(n)}(g) \in U(n)$.
- A unitary representation of a finite group is always completely reducible, and thus every representation of a finite group is completely reducible.
- Abelian finite groups have only 1-dimensional irreducible representations.
- Schur's two lemmata:
  (1) If $D(G)$ is a finite dimensional irreducible representation of a group, and if $D(g)A = AD(g)\ \forall g \in G$ then A is a multiple of the unit matrix.
  (2) If D_1 and D_2 are two different irreducible representations and if for a matrix $A,\ D_1(g)A = AD_2(g)\ \forall g \in G$ holds, then $A = 0$.

Character of a Representation

With the aid of the two lemmata named after I. Schur one can prove orthogonality properties of irreducible representations and of the character of a representation. The notion of the character of a representation has its origin in the fact that similarity equivalent representations of a group element, that is $D'(g) = SD(g)S^{-1}$, although realized by different matrices, do have a common index: Because of the cyclic property of the trace, tr (ABC) = tr (BCA) = tr (CAB), we immediately see that the traces of $D(g)$ and $D'(g)$ are the same. And this is the character of a group element with respect to the representation D. The *character* of a representation (also called the character table) is defined as the set

$$\chi(D) := \{\chi_D(g_i)\} \qquad \text{where} \qquad \chi_D(g_i) := \text{tr}\ D(g_i).$$

In the following, in line with the established group representation literature, the dependence of the characters on the representation will be denoted as $\chi^{(\mu)}$. It can be shown that

- For finite groups of order $[g]$ the characters of two irreducible representations $\chi^{(\mu)}$ und $\chi^{(\nu)}$ fulfill
$$\sum_g^{[g]} \chi^{(\mu)}(g)\chi^{*(\nu)}(g) = [g]\delta^{\mu\nu}. \tag{A.23}$$

- The characters are constant on conjugacy classes: Conjugate elements in the group do have the same character. If **G** has k conjugacy classes $C_1, \ldots, C_k$, each containing $h_1, \ldots, h_k$ elements, and denoting by $\chi_r(r = 1, \ldots, k)$ the character of the conjugacy classes, one can write (A.23) in the form

$$\sum_{r=1}^{k} h_r \chi_r^{(\mu)} \chi_r^{*(\nu)} = [g]\delta^{\mu\nu}.$$

  This entails specifically that the number of irreducible representations of a group **G** is the same as the number of its conjugacy classes.
- Any representation can be expanded as a direct sum of irreducible representations:

$$D(g_i) = \oplus_\nu \, m_\nu^D D^{(\nu)}(g_i) \qquad \text{with} \qquad m_\nu^D = \frac{1}{[g]} \sum_r h_r \; \chi_r^{*(\nu)} \chi_r^D .$$

- The criterion for irreducibility of a representation is $\sum_r h_r |\chi_r|^2 = [g]$.

Product Representations

Multiplying two representations results in another representation: The *direct product* of an $(n \times n)$-matrix A_{ij} and an $(m \times m)$ -matrix $B_{\alpha\beta}$ is an $(n \cdot m) \times (n \cdot m)$-matrix $(A \times B)$ defined by

$$(A \times B)_{i\alpha, j\beta} := A_{ij} B_{\alpha\beta}.$$

The representation matrices of a product representation are then given by the product of the single representations:

$$D^{(n \times m)}_{i\alpha, j\beta}(g) = D^{(n)}_{ij}(g) D^{(m)}_{\alpha\beta}(g).$$

The result can be decomposed into irreducible components. This process is familiar from quantum mechanics as the addition of angular momenta and the subsequent determination of the Clebsch-Gordan series.

$$D^{(n)} \otimes D^{(m)} = \oplus \, \Gamma^{nmk} \quad D^{(k)}. \tag{A.24}$$

The coefficients Γ^{nmk} can be derived from the characters of the groups in question as

$$\Gamma^{nmk} = \frac{1}{[g]} \sum_g \chi^{(n)}(g) \; \chi^{(m)}(g) \; \chi^{(k)}(g^{-1}). \tag{A.25}$$

A.3.3 Representation of Continuous Groups

Many notions and results for the representation of finite groups are directly transferable to Lie groups. You saw that a central technical means was the character for distinguishing inequivalent representations. Together with a "group mean value"

$$M(f) = \frac{1}{[g]} \sum_g f(g)$$

of a function $f : G \to \mathbb{C}$ this gave clues on type and number of irreducible representations.

Invariant Integration

Only for those continuous groups for which a substitute of the group mean value with its specific properties can be defined do all results from the representation theory of finite groups hold. The specific properties of $M(f)$ are (1) linearity: $M(\alpha f_1 + \beta f_2) = \alpha M(f_1) + \beta M(f_2)$ (2) positiveness: from $f \geqslant 0$ follows $M(f) \geqslant 0$, and $M(f) = 0$ only for $f = 0$ (3) normalization: if $f(g) = 1$ for all $g \in G$, then $M(f) = 1$ (4) left and right invariance, e.g. left invariance:

$$M_{l_a}(f) = \frac{1}{[g]} \sum_g f(ag) = \frac{1}{[g]} \sum_g f(g) = M(f)$$

since for any fixed $a \in G$, we have $\{ag|g \in G\} = \{g|g \in G\}$. The same applies to right invariance.

For a continuous group, the mean group value defined in terms of a summation needs to be replaced by an integral. Then the question arises of whether one can find an (integration) measure. The answer is positive if the Lie group is compact. In this case

$$M(f) = \int_G d\mu(g) f(g)$$

where $d\mu$ is called the *Haar measure*. I refrain from further details and from deriving the Haar measure for those compact groups most important in the Standard Model, namely $\mathbf{U(1)}$, $\mathbf{SU(2)}$, and $\mathbf{SU(3)}$, because this would require several additional pages.

Some Results on Representations

Here I state without proof that

- All representations of a semi-simple group are fully reducible.
- Unitary representations are fully reducible.

- The finite-dimensional representations of compact Lie groups are all unitary (and the finite-dimensional representations of compact Lie-algebras are all Hermitean).
- Every representation of a compact group is equivalent to a unitary representation.
- A compact group has a countably infinite number of inequivalent, unitary, irreducible representations which are all finite-dimensional.
- Non-compact Lie groups have uncountable numbers of inequivalent, unitary irreducible representations which are all infinite dimensional. Their finite-dimensional representations are not unitary.

Adjoint Representation

For every Lie algebra, there exists a distinguished representation. This is the adjoint[9] representation defined by matrices $\hat{T}^a$:

$$(\hat{T}^a)^{bc} := -if^{bca}. \tag{A.26}$$

(Observe the different positions of the indices on both sides, and remember that the structure constants are antisymmetric in their first two indices.) By inserting these into the Jacobi–identity (A.8) one indeed finds $([\hat{T}^a, \hat{T}^b])^{de} = if^{abc}(\hat{T}^c)^{de}$ which assures that the matrices $(\hat{T}^a)$ are a representation of the algebra. The adjoint representation has–by construction–the same dimension as the group itself.

We saw that a linear transformation on the Lie group generators induces a linear transformation of the structure constants in the form (A.10). Now one can easily show that taking the transformed structure functions to define the adjoint representation is nothing but a similarity transformation on the matrices $(\hat{T}^a)$:

$$(\hat{T}^a) \quad \to \quad (\hat{T}'^a) = \Gamma^{ad}\ \Gamma(\hat{T}^d)\Gamma^{-1}.$$

This similarity transformation induces a change

$$\mathrm{tr}\ (\hat{T}^a\hat{T}^b) \quad \to \quad \mathrm{tr}\ (\hat{T}'^a\hat{T}'^b) = \Gamma^{ac}\Gamma^{bd}\ \mathrm{tr}\ (\hat{T}^c\hat{T}^d).$$

The trace can be diagonalized by an appropriate choice of the Γ-matrix, and thus we can assume that

$$\mathrm{tr}\ (\hat{T}^a\hat{T}^b) = C^a\delta^{ab} \qquad \text{(no sum)}.$$

By a suitable rescaling of the generators the C^a can be assumed to have values $\{+1, 0, -1\}$. As it turns out, the case with all C^a positive is characteristic for a compact Lie algebra: In a suitable basis $(\hat{T}^a\hat{T}^b) = \delta^{ab}$, multiply the commutator of the representation matrices with $\hat{T}^d$ and take the trace:

$$\mathrm{tr}\ ([\hat{T}^a, \hat{T}^b]\hat{T}^d) = \mathrm{tr}\ (\hat{T}^a\hat{T}^b\hat{T}^d) - \mathrm{tr}\ (\hat{T}^b\hat{T}^a\hat{T}^d) = \mathrm{tr}(f^{abc}\hat{T}^c\hat{T}^d) = if^{abd}.$$

Because of the cyclic property of the trace, the previous relation reveals that the structure functions are completely antisymmetric in all indices. Therefore the $(\hat{T}^a)$ are Hermitean matrices: $((\hat{T}^a)^{bc})^\dagger = ((\hat{T}^a)^{cb})^* = if^{cba} = -if^{bca} = (\hat{T}^a)^{bc}$.

[9] A similar notion exists for finite groups where it is called the regular representation.

Defining/Fundamental Representation

If the group in question is already a matrix group $\mathbf{GL(n, F)}$, there obviously is a *defining representation* $D[\mathbf{GL}(\mathbf{n}, \mathbb{F})] = \mathbf{GL}(\mathbf{n}, \mathbb{F})$, that is a $\mathbb{F}$-valued n-representation. This is sometimes also called the *fundamental* representation. Other representations may be obtained by taking the repeated direct product of this representation (and/or its complex conjugate).

Complex and Real Representations

Let T^a be a representation of the generators X^a associated to the Lie group $\mathbf{G}$: $\left[T^a, T^b\right] = if^{abc}T^c$. Taking the complex conjugate of this expression, we see that $(-T^{a*})$ also forms a representation:

$$\left[(-T^{a*}), (-T^{b*})\right] = if^{abc}(-T^{c*}).$$

Thus every representation T has an associated *conjugate representation* $\bar{T}$ with $\bar{T}^a = -T^{a*}$. If a representation is equivalent to its conjugate, there is an unitary similarity transformation such that $\bar{T}^a = UT^aU^{-1}$. In this case the representation is real. Otherwise the representation is complex.

Casimir's and Dynkin's

Denote by T^a_ρ the matrices of an irreducible representation ρ with dimension $d(\rho)$. For every representation one can define characteristic indices: Assume that the representation matrices satisfy the normalization condition

$$\mathrm{Tr}\ (T^a_\rho T^b_\rho) = C(\rho)\delta^{ab}$$

(it can be shown that this is always possible to arrange for semi-simple algebras). The constant $C(\rho)$ is called the Dynkin index of the representation.
A further index relates to the quadratic Casimir operator $g_{ab}X^aX^b$ where g_{ab} is the Cartan–Killing metric. For any simple Lie algebra, one can choose a basis in which it has the form $X^2 = X^aX^a$. By definition, it commutes with all generators in the algebra. Thus, according to the Schur lemma, its representation takes on a constant value in each irreducible representation:

$$T^a_\rho T^a_\rho = C_2(\rho)I.$$

The constant $C_2(\rho)$ is called the quadratic Casimir index. In the adjoint representation it becomes

$$f^{cba}f^{cda} = C_2(\mathrm{adj})\delta^{bd}.$$

The quadratic Casimir index is related to the Dynkin index by

$$d(\rho)C_2(\rho) = d(\text{adj})C(\rho). \tag{A.27}$$

These indices appear for instance in the beta functions of field theories.
Another index arises repeatedly in expressions for the chiral anomaly. Take

$$\text{Tr}\,\left[T^a_\rho \{T^b_\rho T^c_\rho\}\right] = \mathcal{A}^{abc}(\rho).$$

Specifically for $\mathbf{SU(N)}$ one defines the anomaly coefficient $A(\rho)$ by

$$\mathcal{A}^{abc}(\rho) = \frac{1}{2}A(\rho)d^{abc}$$

where d^{abc} is defined from the anti-commutator of the representation matrices in the fundamental representation

$$\{T^a_f T^b_f\} = \frac{1}{N}\delta^{ab} + d^{abc}T^c_f.$$

A.3.4 Examples: Representations of SO(2), SO(3), SU(3)

SO(2)

The group $\mathbf{SO}(2)$ describes the 1-parameter Lie-group of rotations in the plane, parameterized by an angle $0 \le \varphi < 2\pi$. The defining/fundamental representation is given by matrices

$$R = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}. \tag{A.28}$$

This group is Abelian, therefore all irreducible representations are one dimensional. However, the defining representation is obviously not reducible in the domain of the real numbers but only in the domain of complex numbers:

$$\begin{pmatrix} x' + iy' \\ x' - iy' \end{pmatrix} = \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix} \begin{pmatrix} x + iy \\ x - iy \end{pmatrix}$$

and $e^{i\varphi}$ is indeed a $\mathbf{U(1)}$-matrix. Hence $\mathbf{SO(2)}$ and $\mathbf{U(1)}$ are locally isomorphic. (As will be explained in Sect. A.3.5, $\mathbf{U(1)}$ is the universal covering group of $\mathbf{SO(2)}$.) All representations are classified by the integer (quantum) number m as

$$D^{[m]} = e^{im\varphi}.$$

SO(3)

This is the rotation group in three dimensions. The defining representation is given in form of real-valued (3×3)-matrices $R_n(\varphi)$ which, referring to Cartesian coordinates, implement the effect of a rotation around the axis $\vec{n}$ by an angle φ. If one denotes by $R_i(\varphi)$ the rotation around the coordinate axis x_i, one simply can enlarge the $SO(2)$ matrix (A.28) for each of the three axis, i.e.

$$R_3 = \begin{pmatrix} \cos\varphi_3 & -\sin\varphi_3 & 0 \\ \sin\varphi_3 & \cos\varphi_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_1 & -\sin\varphi_1 \\ 0 & \sin\varphi_1 & \cos\varphi_1 \end{pmatrix}$$
$$R_2 = \begin{pmatrix} \cos\varphi_2 & 0 & -\sin\varphi_2 \\ 0 & 1 & 0 \\ \sin\varphi_2 & 0 & \cos\varphi_2 \end{pmatrix}.$$

The generators $X_k = -i\frac{dR_k}{d\varphi}|_{\varphi=0}$ were already calculated as the matrices in (A.4).

In order to exemplify how to construct irreducible representations of a group, we will recapitulate how one proceeds in the case of quantum mechanics. The generators X_k correspond up to a factor $\hbar$ (which in the subsequent calculation is set to 1 in any case) to the angular momentum operators J_k with

$$[J_x, J_y] = i\hbar J_z \qquad [J_y, J_z] = i\hbar J_x \qquad [J_z, J_x] = i\hbar J_y.$$

In the subsequent considerations, a prominent role is played by the quadratic Casimir operator

$$J^2 = J_x^2 + J_y^2 + J_z^2.$$

It follows indeed from the Cartan-Killing metric of $\mathfrak{so}(3)$ according to (A.19). Therefore–complying with the quantum mechanical interpretation– J^2 can be measured simultaneously with any one of the angular momentum operators. Customarily one chooses the z-component, for which the eigenvalues are already known from the **SO(2)** consideration in the previous subsection. Therefore we have the two eigenvalue equations

$$J^2|\alpha, m\rangle = \alpha|\alpha, m\rangle \qquad J_z|\alpha, m\rangle = m|\alpha, m\rangle$$

with real-valued α and integer m (and $\hbar = 1$). Because of

$$(J_x^2 + J_y^2)|\alpha, m\rangle = (J^2 - J_z^2)|\alpha, m\rangle = (\alpha - m^2)|\alpha, m\rangle$$

the eigenvalues α must obey the inequalities

$$\alpha \geq 0, \qquad \alpha - m^2 \geq 0. \tag{A.29}$$

Now introduce the ladder operators

$$J_\pm := J_x \pm iJ_y \qquad \text{with} \qquad [J_z, J_\pm] = \pm J_\pm.$$

Their name 'ladder operators' (or 'raising and lowering operators') is due to their property of changing the eigenvalue of J_z by ± 1:

$$J_{\pm}|\alpha, m\rangle = \sqrt{\alpha - m^2 \mp m}\; |\alpha, m \pm 1\rangle. \tag{A.30}$$

As a consequence of this and of the inequalities (A.29), one argues that the eigenvalues can be found only in the spectrum

$$\alpha = j(j+1); \qquad j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots; \qquad m = -j, -j+1, \ldots., j-1, j.$$

Therefore the eigenvalues of J_z are restricted to $(2j+1)$ values. The $(2j+1)$-dimensional irreducible representations can be enumerated by the quantum numbers (j, m) and the representation of the generators in terms of matrices are

$$\langle j, m'|J_z|j, m\rangle = m\delta_{m'm} \tag{A.31a}$$

$$\langle j, m'|J_+|j, m\rangle = [(j-m)(j+m+1)]^{1/2}\delta_{m'm+1} \tag{A.31b}$$

$$\langle j, m'|J_-|j, m\rangle = [(j+m)(j-m+1)]^{1/2}\delta_{m'm-1}, \tag{A.31c}$$

where (A.31b, A.31c) follow from (A.30). The first three lowest dimensional representations are therefore:

- one-dimensional representation $(j = 0)$
  $J_z = J_{\pm} = 0$ with the consequence that $R_k = \exp(i\varphi J_k) = 1$, and thus–as to be expected–this is simply the trivial representation.
- two-dimensional representation $(j = \frac{1}{2})$
  The representation matrices (A.31) are

  $$J_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad J_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad J_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

  which amounts to $J_i = \frac{1}{2}\sigma_i$ with the Pauli matrices (A.12). This representation is called the spinor representation of the rotation group. Like all other half-integral representations, it cannot be derived or expressed in terms of tensor representations. The half-integral representations of $\mathbf{SO}(n, \mathbb{C})$ were discovered by E. Cartan in 1913. Later in the 20th they were rediscovered independently (by e.g. Dirac and Pauli) in quantum physics in the context of describing the spin of the electron.
  As shown in Sect. A.2.4 the groups $\mathbf{SO(3)}$ and $\mathbf{SU(2)}$ have isomorphic Lie algebras. This comes about again since the Pauli matrices arising in the $j = 1/2$ representation of $\mathbf{SO(3)}$ constitute the defining representation of $\mathbf{SU(2)}$ in the form $U(\xi) = \exp(\frac{i}{2}\,\xi^a\sigma^a)$; see (A.13). In the next section, $\mathbf{SU(2)}$ is identified as the universal covering group of $\mathbf{SO(3)}$.
- three-dimensional representation $(j = 1)$
  For $j = 1$ the matrices (A.31) combine to

  $$J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad J_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_2 = \frac{1}{i\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

At first sight they seem to be unrelated to the three-dimensional matrices (A.4) from the defining representation of **SO**(**3**). But remember that irreducible representations are defined up to similarity transformations. In this case the unitary matrix

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 0 & 1 \\ -i & 0 & -i \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

mediates the transformation to $X_k = S J_k S^\dagger$.

The coefficients of the Clebsch-Gordan series (A.24) for **SO**(**3**) can be determined from the character of the group and the orthogonality of its elements. According to (A.31a), the generator J_3 has in the $(2j+1)$-dimensional representation always the form

$$J_3 = \mathrm{diag}(j, j-1, \ldots, -j)$$

and thus

$$R_3(\varphi) = e^{iJ_3\varphi} = \mathrm{diag}(e^{ij\varphi}, e^{i(j-1)\varphi}, \ldots, e^{-ij\varphi}),$$

and the trace of this representation is calculated to be

$$\chi^{[j]} = e^{-ij\varphi} + e^{-i(j-1)\varphi} + \ldots + e^{ij\varphi} = \frac{\sin(j+\frac{1}{2})\varphi}{\sin\frac{1}{2}\varphi}.$$

In the case of **SO**(**3**) the "group mean value" derived from the Haar measure is

$$\frac{1}{[g]} \sum_g \Rightarrow \int_0^{2\pi} \frac{d\varphi}{2\pi}(1 - \cos\varphi),$$

and it can be deduced that the coefficients in (A.24) calculated from the character elements as in (A.25) yield

$$D^{[j_1]} \otimes D^{[j_2]} = \sum_{|j_1-j_2|}^{|j_1+j_2|} \oplus D^{[j]}.$$

You can find this result in books on quantum mechanics from the addition of angular momenta and its subsequent reduction in terms of multiplets. In introducing the notation **1** for a singlet ($j = 0$), **2** for a doublet ($j = 1/2$), **3** for a triplet ($j = 1$), etc. one arrives at expressions such as

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{3} \oplus \mathbf{5}$$
$$\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3}$$

from which the dimension of the product representation and its reduction can be immediately recognized.

SU(3)

Historically, the group $\mathbf{SU}(\mathbf{3})$ arose in physics in the original (up, down, strange) quark model in explaining the composition of hadrons from constituent quarks. Today this is recognized as an approximately valid "flavor" $\mathbf{SU_F}(\mathbf{3})$), and it is the $\mathbf{SU_C}(\mathbf{3})$ with respect to the three color degrees of freedom that is considered to be the exact symmetry of quantum chromodynamics; see Sect. 6.2.

$\mathbf{SU}(\mathbf{3})$ is an eight-dimensional compact semi-simple group. Thus it is possible to find a basis F_a in its associated Lie algebra such that the structure constants in

$$[F_a, F_b] = i f_{abc} F_c \qquad (a, b, c = 1, \ldots 8)$$

are real and totally antisymmetric. The convention of choosing the non-vanishing structure constants as

$$f_{123} = 1 \qquad\qquad f_{147} = f_{246} = f_{257} = f_{345} = \frac{1}{2}$$

$$f_{156} = f_{367} = -\frac{1}{2} \qquad\qquad f_{458} = f_{678} = \frac{1}{2}\sqrt{3}.$$

is due to Gell-Mann. The defining three-dimensional matrix representation of the generators is provided by $F_a = \frac{1}{2}\lambda_a$, with the *Gell-Mann matrices*

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \qquad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Observe that λ_1, λ_4 and λ_6 look like the Pauli matrix σ_1, aside from having an additional column and row with zeros. In the same manner, λ_2, λ_5 and λ_7 are related to σ_2, and λ_3 with σ_3[10].

Other representations of $\mathbf{SU}(\mathbf{3})$ can be obtained in a similar manner to the procedure for $\mathbf{SO}(\mathbf{3})$: One has to identify a set of operators which can be diagonalized simultaneously. Again, this task is supported by means of Casimir operators. The quadratic Casimir operator derived from the Cartan-Killing metric is

$$F^2 = \sum_{a=1}^{8} F_i^2.$$

[10] This observation gives a hint for how to construct the defining representations of $\mathbf{SU}(\mathbf{n})$ for $n > 3$.

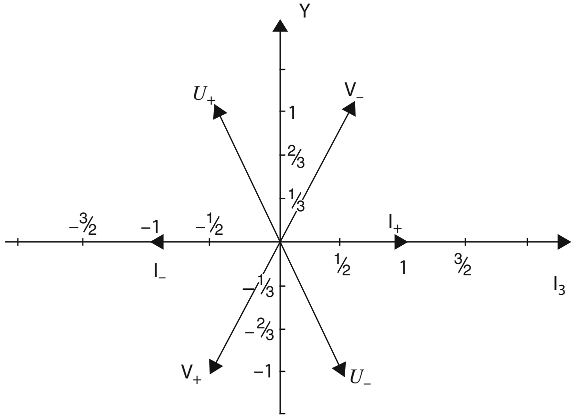


Fig. A.2 Ladder operations for SU(3)

There are two further commuting generators, namely F_3 and F_8, or–following the denotation of quantum numbers in particle physics–the isospin and the hypercharge operator

$$I_3 = F_3, \qquad Y = \frac{2}{\sqrt{3}} F_8.$$

Quite in analogy with the further procedure in case of the rotation group, one defines linear combinations with the remaining operators:

$$I_\pm := F_1 \pm i F_2, \qquad U_\pm := F_6 \pm i F_7, \qquad V_\pm := F_4 \pm i F_5$$

which act as ladder operators with respect to the isospin (I-spin), the so-called U-spin and the V-spin. Indeed, accordingly there are three $\mathfrak{su}(2)$ -subalgebras of $\mathfrak{su}(3)$, spanned by the generators

$$\{I_\pm, I_3\}, \qquad \{U_\pm, U_3 = \frac{3}{4}Y - \frac{1}{2}I_3\}, \qquad \{V_\pm, V_3 = -\frac{3}{4}Y - \frac{1}{2}I_3\}.$$

Again in analogy with the ladder operators of the rotation group, whose action can be visualized on a line (stretched along the eigenvalue of I_3), it is immediately clear that in the present case, one can argue within a two-dimensional visualization in the (Y, I_3)-plane. The eigenvalues of I_3 are $\{0, \pm{}^1\!/_2, \pm 1, \pm{}^3\!/_2, \ldots\}$, and the eigenvalues of $Y = {}^2\!/_3(U_3\text{-}V_3)$ are $\{0, \pm{}^1\!/_3, \pm{}^2\!/_3, \pm 1, \pm{}^4\!/_3, \ldots\}$. Thus, all the states are localized on a hexagonal lattice in which, starting from an arbitrary point, all displaced states can be reached by ladder operations I_+, U_+, V_+ with $(I_3+{}^1\!/_2, Y)$, $(I_3\text{-}{}^1\!/_2, Y+{}^1\!/_3)$, and $(I_3\text{-}\frac{1}{2}, Y\text{-}{}^1\!/_3)$, respectively; see Fig. A.2.
The representations of $\mathbf{SU(3)}$ are therefore characterized by two numbers α and β which indicate by how many units the ladder operations act. A standard has been established (see [163], Chap. 11) by which parameters α and β are chosen such that

$$\Delta I_3 = \alpha \qquad \Delta Y = \frac{1}{3}(\alpha + 2\beta).$$

The dimension d of the representation $D^{[\alpha\beta]}$ is then

$$d(\alpha,\beta)=\frac{1}{2}(\alpha+1)(\beta+1)(\alpha+\beta+2).$$

Accordingly $D^{[00]}$ is the singlet, $D^{[10]}$ the triplet, $D^{[01]}$ the anti-triplet, $D^{[20]}$ the sextet, $D^{[11]}$ the octet, and $D^{[30]}$ the decuplet representation.

The smallest non-trivial representations are the triplet and the anti-triplet. These are the defining representations of **SU(3)** in terms of the Gell-Mann λ-matrices or their Hermitean conjugates. The objects which transform under these representations are three-component $\mathbf{SU}(\mathbf{3})$-vectors ψ, i.e. $\psi'=e^{i(\alpha^a\lambda_a)}\psi$.

Some of the lower-dimensional product representations of $\mathbf{SU}(\mathbf{3})$ are

$$\begin{aligned}\mathbf{3}\otimes\overline{\mathbf{3}}&=\mathbf{8}\oplus\mathbf{1},\\ \mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}&=\mathbf{10}\oplus\mathbf{8}\oplus\mathbf{8}\oplus\mathbf{1},\\ \mathbf{3}\otimes\mathbf{3}&=\overline{\mathbf{3}}\oplus\mathbf{6}.\end{aligned}$$

If one performs the reduction of $\mathbf{3}\otimes\overline{\mathbf{3}}$ with the $\mathbf{SU}(3)$-vectors ψ, one can by use of the identity

$$\psi_\alpha\bar{\psi}^\beta=\left(\psi_\alpha\bar{\psi}^\beta-\frac{1}{3}\delta_\alpha^\beta\psi_\gamma\bar{\psi}^\gamma\right)+\frac{1}{3}\delta_\alpha^\beta\psi_\gamma\bar{\psi}^\gamma, \tag{A.32}$$

directly read of the octet- and singlet-parts.

Finally, some remarks on the Dynkin and Casimir indices of $\mathbf{SU}(\mathbf{N})$ are appropriate. The fundamental representation of $\mathbf{SU}(\mathbf{2})$ is given in terms of Pauli matrices $T_2^a=\sigma^a/2$, which satisfy $\mathrm{Tr}\left[T_2^aT_2^b\right]=\frac{1}{2}\delta^{ab}$. As we saw for $\mathbf{SU}(\mathbf{3})$, three of the matrices of its fundamental representation contain the Pauli matrices, which here act on the first two components of an $\mathbf{SU}(\mathbf{3})$-vector. A quite similar choice of generators can be made for larger $\mathbf{SU}(\mathbf{N})$ algebras. Therefore

$$\mathrm{Tr}\left[T_N^aT_N^b\right]=\frac{1}{2}\delta^{ab}$$

and thus the Dynkin index is $C(N)=1/2$ and the quadratic Casimir is found from (A.27). We note down that in the fundamental representation of SU(N)

$$C(N)=\frac{1}{2}\qquad C_2(N)=\frac{N^2-1}{2N}. \tag{A.33}$$

At various places (for instance in the generic expression for the β function), one also needs the Dynkin and Casimir in the adjoint representation of $\mathbf{SU}(\mathbf{N})$. This may be derived from the product of fundamental representations

$$\mathbf{N}\otimes\overline{\mathbf{N}}=(\mathbf{N^2}-\mathbf{1})\oplus\mathbf{1}. \tag{A.34}$$

For a product representation $r_1 \times r_2 = \Gamma^i\ r_i$, it holds that

$$\big(C_2(r_1) + C_2(r_2)\big)\ d(r_1)d(r_2) = \sum \Gamma^i\ C_2(r_i)d(r_i).$$

In applying this to (A.34) by using (A.33), we derive from

$$\big(2 \cdot \frac{N^2-1}{2N}\big)N^2 = C_2(adj) \cdot (N^2-1)$$

the $\mathbf{SU(N)}$ quadratic Casimir and Dynkin indices for the adjoint representation:

$$C_2(adj) = C(adj) = N. \tag{A.35}$$

A.3.5 Projective Representations and Central Charges

As explained in Sect. 4.2., the fact that quantum physical states are not vectors but "rays" in a Hilbert space makes it necessary to consider ray representations (or projective representations) of symmetry groups.

Ray Representations

A *ray representation* of a group $\mathbf{G} = \{g, g', \ldots\}$ by unitary operators $\mathbf{U}$ in a Hilbert space is a map $\mathbf{G} \to \mathbf{U}$ i.e. $g \mapsto U_g$ obeying

$$U_{g'} \cdot U_g = e^{i\phi(g',g)} U_{g' \circ g} \tag{A.36}$$

with a phase $\phi(g', g)$. The phases in (A.36) are constrained because of the associativity of the group operation: From $U_{g''} \cdot (U_{g'} \cdot U_g) \stackrel{!}{=} (U_{g''} \cdot U_{g'}) \cdot U_g$, one immediately derives the condition

$$\phi(g', g) + \phi(g'', g' \circ g) \stackrel{!}{=} \phi(g'', g') + \phi(g'' \circ g', g) \tag{A.37a}$$

$$\phi(g', g_0) = \phi(g_0, g) = 1. \tag{A.37b}$$

The second line is due to the generic property of a phase if either g or g' is the neutral element g_0 in the group. We are still allowed to redefine the unitary operators as $\tilde{U}_g = U_g \exp(i\alpha(g))$. Then a simple calculation shows that the phase $\tilde{\phi}(g', g)$ in $\tilde{U}_{g'} \cdot \tilde{U}_g = e^{i\tilde{\phi}(g',g)}\ U_{g' \circ g}$ is related to $\phi(g', g)$ by

$$\tilde{\phi}(g', g) = \phi(g', g) - \alpha(g' \circ g) + \alpha(g') + \alpha(g). \tag{A.38}$$

and also obeys the conditions (A.37). The ray representations U and $\tilde{U}$ are said to be equivalent.

Curiosity provokes the question as to whether and when it is possible to find representations for which the phase $\phi(g', g)$ can be made to vanish. The answer was given by V. Bargmann in 1954 [27]:

(1) All central charges of the Lie algebra associated to the symmetry group vanish or can be transformed away.

(2) The group is simply connected.

Interestingly enough, the first condition refers to properties of the group on a small scale (near the identity) and the second one is related to properties on a large scale (i.e. the topology of the group manifold). These two points are explained below.

Central Charges

There is a relation between the phases of a projective representation of a Lie group and central charges in its Lie algebra. This will be derived next by expanding the group elements and their representation operators (near the respective identities) in terms of the coordinates $\{\xi^a\}$. At first observe that if we write $g' \circ g = g(\xi')g(\xi) = g(\xi''(\xi', \xi))$, the dependence of the "composed" parameter set ξ'' takes on the form

$$\xi''^a = \xi'^a + \xi^a + \gamma^{abc}\xi'^b\xi^c + \dots$$

(with real coefficients γ^{abc}) because $\xi''^a(0, \xi) = \xi^a$ and $\xi''^a(\xi', 0) = \xi'^a$. Expanding the phase $\phi(g', g)$, we observe that because of (A.37)

$$\phi(g', g) = \phi(\xi', \xi) = \gamma^{bc}\, \xi'^b\xi^c + \dots \tag{A.39}$$

Finally, the representation matrices near the identity can be expanded as

$$U_{g(\xi)} = 1 + i\xi^a T^a + \frac{1}{2}\xi^a\xi^b T^{ab} + \dots$$

with Hermitean matrices T^a and with $T^{ab} = T^{ba}$. Now expand (A.36) as

$$\begin{aligned}\Big(1 + i\xi'^a T^a + \frac{1}{2}\xi'^a\xi'^b T^{ab} + \dots\Big)\Big(1 + i\xi^a T^a + \frac{1}{2}\xi^a\xi^b T^{ab} + \dots\Big)& \\ = \Big(1 + i\gamma^{bc}\xi'^b\xi^c + \dots\Big)\Big(1 + i(\xi'^a + \xi^a + \gamma^{abc}\xi'^b\xi^c + \dots)T^a& \\ + \frac{1}{2}(\xi'^a + \xi^a + \dots)(\xi'^b + \xi^b + \dots)T^{ab} + \dots\Big).&\end{aligned}$$

Comparing both sides of this expression order by order in the ξ and the ξ', we observe that the terms of order $1, \xi, \xi', \xi^2, \xi'^2$ match. The terms proportional to $\xi'^c\xi^b$ disappear if

$$-T^c T^b = i\gamma^{cb}1 + i\gamma^{acb}T^a + T^{cb}.$$

By defining

$$f^{abc} := \gamma^{acb} - \gamma^{abc} \qquad f^{bc} := \gamma^{cb} - \gamma^{bc}$$

the previous condition becomes

$$[T^b, T^c] = i f^{abc} T^a + i f^{bc} 1. \tag{A.40}$$

Thus we recover in the first term on the right-hand side the structure coefficients f^{abc} of the Lie algebra. The further terms f^{bc} are the *central charges*. Due to (A.39) the central charges can be calculated from the phase as

$$f^{bc} = \left(\frac{\partial^2 \phi(\xi', \xi)}{\partial \xi^b \partial \xi'^c} - \frac{\partial^2 \phi(\xi', \xi)}{\partial \xi^c \partial \xi'^b} \right)_{|\xi=0=\xi'}.$$

Now the Jacobi identities in the Lie algebra impose conditions on the central charges in the form

$$f^{abd}\ f^{cd} + f^{bcd}\ f^{ad} + f^{cad}\ f^{bd} = 0. \tag{A.41}$$

In some cases, these conditions are so stringent that the only solution is $f^{ab} = 0$. In other cases one may be able to eliminate the central charges by redefining the Lie algebra generators as $X^a \mapsto \tilde{X}^a = X^a + C^a(f^{bc})$. This is possible specifically for solutions of (A.41) in the form $f^{ab} = f^{abd} C_d$ with an arbitrary set of real constants C_d. However, there are Lie algebras for which these are not the only solutions. In this case, not all central charges can be transformed away and one must concentrate one's considerations onto an algebra which is extended by these central charges. This extension amounts to augmenting the set of generators by additional ones being proportional to the non-vanishing central charges. A prominent example is given in Sect. 3.4 by the Galilei algebra centrally extended to the Bargmann algebra.

Global Properties of Lie Groups

As before, let $\mathbf{G}$ be a Lie group with elements $g, g', \ldots$. Two group elements g and g' are called *connected* if a continuous curve $c(\lambda)$ $(0 \leq \lambda \leq 1)$ exists in $\mathbf{G}$ with $c(0) = g$ and $c(1) = g'$. The set of all elements which are connected to the unit element is an invariant subgroup $\mathbf{G}_0$ of $\mathbf{G}$. A connected group $\mathbf{G}$ is called *simply-connected* if every loop in the group can be shrunk continuously to an element in $\mathbf{G}$.

Each connected Lie group $\mathbf{G}$ has a unique simply-connected *universal covering group* $\mathbf{G}^*$ with the properties (i) There is a homomorphic mapping ϱ: $\mathbf{G}^* \to \mathbf{G}$ (ii) The kernel of the homomorphism ϱ is an invariant discrete subgroup $\mathbf{N}$ of the center of $\mathbf{G}^*$.

Examples for physically important universal covering groups are

- **SU(2)** as universal cover of **SO(3)**
  The defining representation of **SU(2)** is provided by (A.13) as

$$U(\vec{\alpha}) = \exp(\frac{i}{2}\vec{\sigma}\cdot\vec{\alpha})$$

in terms of Pauli matrices σ. Introducing a normal vector $\vec{n}$ by $\vec{\alpha} = \alpha\vec{n}$ with $(\vec{\sigma}\cdot\vec{n})^2 = I$, one can write

$$U = \cos\frac{1}{2}\alpha + i\vec{\sigma}\cdot\vec{n}\sin\frac{1}{2}\theta.$$

Comparing this with the generic form of an **SO(3)** matrix

$$R(\theta) = \exp(i\vec{n}\cdot\vec{J}\theta),$$

where the J_k are the defining generators, we identify a homomorphism $\varrho : \mathbf{SU(2)} \to \mathbf{SO(3)}$. Since $U(2\pi) = -1$ and $R(2\pi) = +1 = R(0)$, two elements from **SU(2)**, namely $U(0)$ and $U(2\pi)$, are mapped onto the neutral element of **SO(3)**, the kernel of the homomorphism is

$$\text{Ker}\ \varrho = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.$$

This is the center of **SU(2)** and is non other than $\mathbf{Z_2}$. Therefore

$$\mathbf{SO(3)} \cong \mathbf{SU(2)}/\mathbf{Z_2},$$

and **SU(2)** is the (twofold) universal covering group of **SO(3)**.

- $\mathbf{SL(2,\mathbb{C})}$ as universal cover of **SO(3, 1)**
  First define from the Pauli matrices the four-vectors $\sigma^\mu := (1, \vec{\sigma})$, $\overline{\sigma}^\mu := (1, -\vec{\sigma})$. These obey

$$\text{Tr}(\sigma^\mu\overline{\sigma}^\nu) = 2\eta^{\mu\nu} \qquad \sigma^\mu\overline{\sigma}^\nu + \sigma^\nu\overline{\sigma}^\mu = 2\eta^{\mu\nu}.$$

Now construct a mapping from Minkowski space to the set of Hermitean complex 2×2 matrices by

$$x^\mu \mapsto X = x^\mu\sigma_\mu = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix}$$

with an arbitrary Lorentz vector x^μ. Observe that $\det X = x^2$. The action of $\mathbf{SL(2, C)}$ on X, that is $X' = AXA^\dagger$ preserves $\det X = x^2$ and can be identified as a Lorentz transformation. Explicitly: To each transformation $A \in \mathbf{SL(2, C)}$ there corresponds a Lorentz transformation $\Lambda \in \mathbf{SO(3, 1)}$ of x^μ:

$$x^{'\mu} = \Lambda^\mu{}_\nu x^\mu \qquad \Lambda^\mu{}_\nu(A) = \frac{1}{2}\text{Tr}(\overline{\sigma}^\mu A\sigma_\nu A^\dagger).$$

The mapping $A \mapsto \Lambda(A)$ is a homomorphism of $\mathbf{SL}(\mathbf{2}, \mathbf{C})$ to $\mathbf{SO}(\mathbf{3}, \mathbf{1})$ as

$$\Lambda^{\mu}{}_{\nu}(A) \in \mathbf{SO}(\mathbf{3}, \mathbf{1})$$
$$\Lambda^{\mu}{}_{\nu}(A_1 A_2) = \Lambda^{\mu}{}_{\varrho}(A_1)\Lambda^{\varrho}{}_{\nu}(A_2).$$

But since $\Lambda(-A) = \Lambda(A)$, the inverse mapping is defined only up to a sign. Or–put another way–the elements $A = \pm 1$ of $\mathbf{SL}(\mathbf{2}, \mathbf{C})$ are mapped onto the identity element of $\mathbf{SO}(\mathbf{3}, \mathbf{1})$. Thus, by definition, the kernel of the mapping is the discrete invariant subgroup $\mathbf{Z_2}$. In conclusion, $\mathbf{SL}(\mathbf{2}, \mathbf{C})$ is the universal covering group of the Lorentz group

$$\mathbf{SO}(\mathbf{3}, \mathbf{1}) \cong \mathbf{SL}(\mathbf{2}, \mathbf{C})/\mathbf{Z_2}.$$

A proof that $\mathbf{SL}(\mathbf{2}, \mathbf{C})$ is simply connected can be found in [536], Sect. 2.7.

Representation Procedure

Based on the concepts and findings of the previous two sections, in this concluding section the generic procedure to find the irreducible (ray) representations of a symmetry group **S** in a Hilbert space is summarized:

1. The first question refers to the connection components of **S**. If the symmetry group is not simply connected one must call on its universal covering group $\mathbf{S}'$.
2. Secondly, there is the question about the possible appearance of central charges in the algebra of generators, i.e. whether there are non-trivial solutions of (A.41).
3. If non-trivial solutions exist, one investigates whether by a re-definition of the generators one can transform the central charges away.
4. For the remaining central charges one enlarges the set of generators (and the group) and one applies the procedure on the central extension of **S**.

Prominent examples in fundamental physics in which this procedure is brought to bear are the representation theory of the Galilei group (in which just one central charge, namely the mass, remains), and the representation theory of the Poincaré group, in which all central charges can be transformed away, but where one needs to regard $\mathbf{SL}(\mathbf{2}, \mathbf{C})$ as the universal covering group of the Lorentz group.

Appendix B
Spinors, $\mathbb{Z}_2$-gradings, and Supergeometry

In this appendix I have collected all those formal techniques which are relevant to the fact that to the best of our knowledge, nature seems to prefer two rather distinct variants of fields, namely

- bosons $\triangleq$ commuting fields $\triangleq$ of integer spin
- fermions $\triangleq$ anti-commuting fields $\triangleq$ of half-integer spin.

B.1 Spinors

The following citation is from the preface of [411]: "To a very high degree of accuracy, the space-time we inhabit can be taken to be a smooth four-dimensional manifold, endowed with the smooth Lorentzian metric of Einstein's special or general relativity. The formalism most commonly used for the mathematical treatment of manifolds and their metric is, of course, the tensor calculus (…[1]). But in the specific case of four dimensions and Lorentzian metric there happens to exist–by accidence or providence–another formalism which is in many ways more appropriate, and that is the formalism of 2-spinors. Yet 2-spinor calculus is still comparatively unfamiliar even now–some seventy years[2] after Cartan first introduced the general spinor concept, and over fifty years since Dirac, in his equation for the electron, revealed a fundamentally important role for spinors in relativistic physics and van der Waerden provided the basic 2-spinor algebra and notation."

This already contains the central message about spinors. They are fundamental objects which can describe four-dimensional Minkowski spacetime. Their geometric origin was revealed in 1913 by E. Cartan. They were so to speak rediscovered in 1928 by Dirac in physics (however as 4-spinors), and gave rise to the concepts of particle spin and antiparticles. And the so-called spinor calculus and the "dot"-notation,

[1] Here, I have deleted a side remark from the original text.

[2] We have to add another 30 years considering the year of publication of the book by R. Penrose and W. Rindler.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

which allows one to construct tensors and spinors of higher weight from the basic 2-spinors, is due to B. van der Waerden. The term "spinor" was coined by P. Ehrenfest, who is said to have encouraged van der Waerden to develop his calculus.

Interestingly enough, nature seems to prefer to be described by 2-spinors: For one thing, as outlined in some detail in Chap. 6, due to parity violation the Lagrangian of the Standard Model is formulated in terms of left-handed leptons and quarks. For another thing, in supersymmetric models the building blocks of matter are chiral supermultiplets, each of which contains a single Weyl fermion.

First of all, we need to fix some conventions on signs and factors i and (1/2), since–mostly for historical reasons–you find various slightly different notations in textbooks and in the research literature. Here, I follow essentially [47] and [137]. The latter authors state explicitly how things change in case of "the other" metric convention and how their definitions relate to those of the influential textbook [545].

B.1.1 Pauli and Dirac Matrices

Pauli Matrices and their Relatives

The three Pauli matrices are introduced in Appendix A.2.3 as the defining representation of $\mathfrak{su}(2)$. A widely used explicit representation is

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{B.1}$$

They have the property

$$\sigma^a \sigma^b = i\epsilon^{abc}\sigma^c + \delta^{ab} I. \tag{B.2}$$

The Pauli matrices are used to define the "four-vectors"

$$\sigma^\mu := (1, \vec{\sigma}), \qquad \overline{\sigma}^\mu := (1, -\vec{\sigma}). \tag{B.3}$$

From (B.2) one finds

$$\sigma^\mu \overline{\sigma}^\nu + \sigma^\nu \overline{\sigma}^\mu = 2\eta^{\mu\nu}.$$

Further definitions are

$$\sigma^{\mu\nu} := \frac{1}{4}\left(\sigma^\mu \bar{\sigma}^\nu - \sigma^\nu \bar{\sigma}^\mu\right) \qquad \bar{\sigma}^{\mu\nu} := \frac{1}{4}\left(\bar{\sigma}^\mu \sigma^\nu - \bar{\sigma}^\nu \sigma^\mu\right). \tag{B.4}$$

These can be expressed in terms of the Pauli matrices themselves:

$$\sigma^{0i} = -\bar{\sigma}^{0i} = -\frac{1}{2}\sigma^i \qquad \sigma^{jk} = \bar{\sigma}^{jk} = -\frac{i}{2}\epsilon^{jkl}\sigma^l.$$

Furthermore, because of $\sigma_a^\dagger = \sigma_a$, one has $(\sigma^{\mu\nu})^\dagger = -\bar{\sigma}^{\mu\nu}$.

Dirac Matrices and their Relatives

The 4×4 Dirac matrices are defined by their algebraic relation

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu}. \tag{B.5}$$

In the metric convention chosen in this book $(\gamma^0)^2 = +1$ and $(\gamma^k)^2 = -1$. Since the γ^μ anti-commute with each other, any product of gamma matrices can be reduced to products which contain at most four factors of gamma matrices. This gives rise to the definition of the matrix

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = -\frac{i}{4!}\epsilon_{\mu\nu\rho\sigma}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma \tag{B.6}$$

for which $(\gamma^5)^2 = 1$. The matrix γ^5 anti-commutes with all the other γ^μ:

$$\{\gamma^5, \gamma^\nu\} = 0.$$

Products of three different gamma matrices can be written in terms of $\gamma^\mu\gamma^5$. (For instance $\gamma^1\gamma^2\gamma^3 = -i\gamma^0\gamma^5$.) And finally, products of two different gamma matrices can be directly expressed via (B.5) as $\gamma^\mu\gamma^\nu = \eta^{\mu\nu} - i\gamma^{\mu\nu}$ where

$$\gamma^{\mu\nu} := \frac{i}{2}[\gamma^\mu, \gamma^\nu]. \tag{B.7}$$

The set of the 16 matrices

$$\Gamma = \{1, \gamma^5, \gamma^\mu, \gamma^\mu\gamma^5, \gamma^{\mu\nu}\} = \{\Gamma^A\} \tag{B.8}$$

constitutes a basis of $\boldsymbol{GL}(4, \mathbb{C})$, that is every complex 4×4 matrix can be written as a linear combination in this basis.

The Clifford condition (B.5) does not have a unique solution. If a certain set of matrices γ^μ is a solution, so is $\gamma'^\mu = S^{-1}\gamma^\mu S$. This means that the two sets of matrices are connected by a similarity transformation. Specifically by taking the Hermitean conjugate and the transpose of (B.5), we notice that $\gamma^{\mu\dagger}$ and -$\gamma^{\mu T}$ are representations of the Clifford algebra. This gives rise to the specific similarity transformations with matrices A and C:

$$A\gamma^\mu A^{-1} = \gamma^{\mu\dagger}, \qquad C^{-1}\gamma^\mu C = -\gamma^{\mu T}$$

from which

$$(CA^T)^{-1}\gamma^\mu(CA^T) = -\gamma^{\mu *}. \tag{B.9}$$

Simply by algebraic manipulations, one can derive various properties of A and C, such as

$$A = A^\dagger, \qquad C = -C^T, \qquad C\gamma^5 = -\gamma^5\, C.$$

The previous relations are valid independently of the representation of the gamma matrices. They hold generally as long as the algebraic relation (B.5) is fulfilled. In terms of specific representations, there are two privileged ones. Their choice depends of whether one wants to work with four-component Dirac spinors (as appropriate to all parity–conserving elementary processes) or with two-component Weyl spinors (as appropriate to the left-right asymmetry in nature). In the *Dirac representation* the gamma matrices are represented as

$$\gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \qquad \gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \qquad \gamma^5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \tag{B.10}$$

where of course the explicit form of γ^5 follows from its definition (B.6). In the *Weyl representation* (also called the *chiral representation* or the *spinor representation*),

$$\gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \qquad \gamma^0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \qquad \gamma^5 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}. \tag{B.11}$$

Thus in both the Dirac and the Weyl representations, the γ^i assume the same form. Furthermore, γ^0 and γ^5 change their role (up to a sign). In the Weyl basis one can write compactly

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}. \tag{B.12}$$

by using the definition (B.3). Furthermore, (B.7) becomes

$$\gamma^{\mu\nu} = 2i \begin{pmatrix} \sigma^{\mu\nu} & 0 \\ 0 & \bar{\sigma}^{\mu\nu} \end{pmatrix}. \tag{B.13}$$

By the way, the terminology in this context is a little confusing. Here we referred to the Dirac and Weyl representations. In fact, there is another privileged one, namely the Majorana representation. On the other hand, in the following we will investigate Weyl, Dirac, and Majorana spinors. Their properties are completely unrelated to the representation issues of the Clifford algebra. And to make it even worse: Majorana spinors are specific Dirac spinors, but Dirac and Majorana fermions obey different field equations, characterized by Dirac and Majorana mass terms.

B.1.2 Weyl Spinors

Weyl spinors do have a genuine geometric and group-theoretical origin, like vectors. They are directly related to the smallest non-trivial representations $(\mathbf{2}) = (\frac{1}{2}, 0)$ and $(\mathbf{2}^*) = (0, \frac{1}{2})$ of the Lorentz group $\mathbf{SO(3, 1)}$, or rather its double covering group $\mathbf{SL(2, C)}$. A left-handed Weyl spinor is a two-component complex objects transforming under $\mathbf{SL(2, C)}$ as

$$\xi'_\alpha = M_\alpha{}^\beta \xi_\beta \tag{B.14}$$

with a matrix M from $\mathbf{SL(2, C)}$. This Weyl spinor belongs to the representation $(\mathbf{2})$, as deduced in Sect. 5.5.3.

Introduce the matrix

$$(\varepsilon^{\alpha\beta}) := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i\sigma^2.$$

For an arbitrary 2×2 matrix A, it holds that $A^T \varepsilon A = (\det A)\, \varepsilon$. Now consider the expression bilinear in two left-handed Weyl spinors ξ and η

$$\varepsilon^{\alpha\beta}\xi_\alpha\eta_\beta = \xi_1\eta_2 - \xi_2\eta_1.$$

This expression is invariant, since

$$\varepsilon^{\alpha\beta}\xi'_\alpha\eta'_\beta = \varepsilon^{\alpha\beta}M_\alpha{}^\gamma M_\beta{}^\delta\xi_\gamma\eta_\delta = (\det M)\; \varepsilon^{\gamma\delta}\xi_\gamma\eta_\delta = \varepsilon^{\gamma\delta}\xi_\gamma\eta_\delta.$$

Therefore $\varepsilon^{\alpha\beta}$ can be considered as a metric on $\mathbf{SL(2, C)}$. Its inverse $\varepsilon_{\alpha\beta}$, defined of course by $\varepsilon^{\alpha\beta}\varepsilon_{\beta\gamma} = \delta^\alpha_\beta$, becomes in components: $(\varepsilon_{\alpha\beta}) = -(\varepsilon^{\beta\alpha})$. Making use of the ε-matrix, define the dual of a left-handed spinor ξ_α as

$$\xi^\alpha := \varepsilon^{\alpha\beta}\xi_\beta.$$

(This construction resembles how one switches from covariant to contravariant indices for Lorentz vectors.) Component-wise, this is $\xi^1 = \xi_2$ and $\xi^2 = -\xi_1$, and–again in analogy with Lorentz vectors–the product of a lower-index spinor with an upper-index spinor transforms as a scalar. However there is a subtlety: Unlike in the case of Lorentz vectors, one must pay attention to the position of indices in the Einstein summation convention:

$$\xi\eta := \xi^\alpha\eta_\alpha = \varepsilon^{\alpha\beta}\xi_\beta\eta_\alpha = -\varepsilon^{\beta\alpha}\xi_\beta\eta_\alpha = -\xi_\beta\eta^\beta = \eta^\beta\xi_\beta = \eta\xi, \tag{B.15}$$

assuming that spinors anti-commute. Since for a matrix M with unit determinant, $\varepsilon M \varepsilon^{-1} = (M^{-1})^T$ holds, the transformation behavior of an upper-index spinor under Lorentz transformations is calculated as

$$\xi'^\alpha = \varepsilon^{\alpha\beta}\xi'_\beta = \varepsilon^{\alpha\beta}M_\beta{}^\gamma\xi_\gamma = \varepsilon^{\alpha\beta}M_\beta{}^\gamma\varepsilon_{\gamma\delta}\xi^\delta = (M^{-1})_\gamma{}^\alpha\xi^\gamma. \tag{B.16}$$

The representations defined by (B.14) and (B.16) are equivalent due to the similarity transformation $M = \varepsilon^{-1}(M^{-1})^T\varepsilon$.
The complex conjugate spinor ξ^*_α transforms as

$$\xi'^*_\alpha = M^*_\alpha{}^\beta\xi^*_\beta.$$

The notation with dotted indices, where by definition[3]

$$\bar{\xi}_{\dot\alpha} := \xi^*_\alpha.$$

[3] This is only a notational sophistry: complex conjugation puts on a dot and a bar. The bar notation stems from [545]; others, like [137] use a † instead.

is due to B.L. van der Waerden. Thus

$$\bar{\xi}'_{\dot{\alpha}} = M^{*\ \dot{\beta}}_{\dot{\alpha}} \bar{\xi}_{\dot{\beta}}. \tag{B.17}$$

This representation is not equivalent to (B.14) since there is in general no matrix S which provides a similarity transformation $M = SM^*S^{-1}$. Spinors which transform as in (B.17) are called right-handed Weyl spinors. As shown in the main text, they constitute the $(\mathbf{2}^*) = (0, \frac{1}{2})$ representation of $\mathbf{SL}(\mathbf{2}, \mathbf{C})$. Of course, one can also define the dual partners to right-handed Weyl spinors by

$$\bar{\xi}^{\dot{\alpha}} := \varepsilon^{\dot{\alpha}\dot{\beta}} \bar{\xi}_{\dot{\beta}} \qquad (\varepsilon^{\dot{\alpha}\dot{\beta}}) \equiv (\varepsilon^{\alpha\beta})$$

which transforms with $(M^{*-1\ T})$. And again, the bilinear form

$$\bar{\xi}\bar{\eta} := \bar{\xi}_{\dot{\alpha}} \bar{\eta}^{\dot{\alpha}} = -\bar{\xi}^{\dot{\alpha}} \bar{\eta}_{\dot{\alpha}} = \bar{\eta}_{\dot{\alpha}} \bar{\xi}^{\dot{\alpha}} = \bar{\eta}\bar{\xi} \tag{B.18}$$

is a Lorentz scalar. If the dotted and the undotted indices are not denoted explicitly, it is always understood that the contraction of right-handed spinors follows this "down-up rule", whereas the contraction of two left-handed spinors happens with the "up-down rule" (B.15). With this convention, we employ the notation

$$\xi^2 = \xi\xi = \xi^\alpha \xi_\alpha \qquad \bar{\xi}^2 = \bar{\xi}\bar{\xi} = \bar{\xi}_{\dot{\alpha}} \bar{\xi}^{\dot{\alpha}}.$$

Furthermore,

$$(\xi\eta)^\dagger = (\xi^\alpha \eta_\alpha)^\dagger = \bar{\eta}_{\dot{\alpha}} \bar{\xi}^{\dot{\alpha}} = \bar{\eta}\bar{\xi} = \bar{\xi}\bar{\eta}. \tag{B.19}$$

Two-Component Fierz Identities

In supersymmetry calculations, one frequently needs the Fierz identities[4]. To formulate them, first adapt the matrices σ^μ and $\bar{\sigma}^\mu$ to the van der Waerden calculus. In the dot-notation σ^μ has the index structure $\sigma^\mu_{\alpha\dot{\alpha}}$ with

$$(\sigma^\mu_{\alpha\dot{\alpha}})^* = (\sigma^{\mu *})_{\dot{\alpha}\alpha} = (\sigma^{\mu\dagger})_{\alpha\dot{\alpha}} = (\sigma^\mu)_{\alpha\dot{\alpha}}.$$

The ε-tensors relate this to $\bar{\sigma}^\mu$ via

$$(\bar{\sigma}^\mu)^{\dot{\alpha}\alpha} = \epsilon^{\dot{\alpha}\dot{\beta}} \epsilon^{\alpha\beta} (\sigma^\mu)_{\beta\dot{\beta}}.$$

The following relations hold:

$$\sigma^\mu_{\alpha\dot{\alpha}} \overline{\sigma}^{\nu\dot{\alpha}\beta} + \sigma^\nu_{\alpha\dot{\alpha}} \overline{\sigma}^{\mu\dot{\alpha}\beta} = 2\eta_{\mu\nu} \delta_\alpha^{\ \beta} \tag{B.20a}$$

$$\overline{\sigma}^{\mu\dot{\alpha}\alpha} \sigma^\nu_{\alpha\dot{\beta}} + \overline{\sigma}^{\nu\dot{\alpha}\alpha} \sigma^\mu_{\alpha\dot{\beta}} = 2\eta_{\mu\nu} \delta^{\dot{\alpha}}_{\ \dot{\beta}}, \tag{B.20b}$$

[4] named after the Swiss physicist Markus Fierz (1912–2006) who derived them as identities for Dirac spinors.

from which

$$\mathrm{Tr}(\sigma_\mu \overline{\sigma}_\nu) = 2\eta_{\mu\nu}.$$

Furthermore,

$$\sigma^\mu_{\alpha\dot{\alpha}}\sigma_{\mu\beta\dot{\beta}} = 2\varepsilon_{\alpha\beta}\varepsilon_{\dot{\alpha}\dot{\beta}} \qquad \bar{\sigma}^{\mu\dot{\alpha}\alpha}\overline{\sigma}^{\dot{\beta}\beta}_\mu = 2\varepsilon^{\alpha\beta}\varepsilon^{\dot{\alpha}\dot{\beta}} \qquad \sigma^\mu_{\alpha\dot{\alpha}}\overline{\sigma}^{\dot{\beta}\beta}_\mu = 2\delta_\alpha{}^\beta\delta^{\dot{\beta}}{}_{\dot{\alpha}}.$$

The generators of the Lorentz group are in the dot notation given by

$$\begin{aligned}(\sigma^{\mu\nu})_\alpha{}^\beta &= \frac{1}{4}\Big((\sigma^\mu)_{\alpha\dot{\alpha}}(\bar{\sigma}^\nu)^{\dot{\alpha}\beta} - (\sigma^\nu)_{\alpha\dot{\alpha}}(\bar{\sigma}^\mu)^{\dot{\alpha}\beta}\Big)\\ (\bar{\sigma}^{\mu\nu})^{\dot{\alpha}}{}_{\dot{\beta}} &= \frac{1}{4}\Big((\bar{\sigma}^\mu)^{\dot{\alpha}\alpha}(\sigma^\nu)_{\alpha\dot{\beta}} - (\bar{\sigma}^\nu)^{\dot{\alpha}\alpha}(\sigma^\mu)_{\alpha\dot{\beta}}\Big).\end{aligned} \tag{B.21}$$

For two anti-commuting spinors, there are the following Fierz identities

$$\xi_\alpha\eta_\beta = -\frac{1}{2}\xi\eta\epsilon_{\alpha\beta} - \frac{1}{8}\xi\sigma^{\mu\nu}\eta(\sigma_{\mu\nu})_{\alpha\beta} \tag{B.22a}$$

$$\bar{\xi}_{\dot{\alpha}}\bar{\eta}_{\dot{\beta}} = -\frac{1}{2}\bar{\xi}\bar{\eta}\bar{\epsilon}_{\dot{\alpha}\dot{\beta}} - \frac{1}{8}\bar{\xi}\bar{\sigma}^{\mu\nu}\bar{\eta}(\bar{\sigma}_{\mu\nu})_{\dot{\alpha}\dot{\beta}} \tag{B.22b}$$

$$\xi_\alpha\bar{\eta}_{\dot{\beta}} = \frac{1}{2}\xi\sigma^\mu\bar{\eta}(\bar{\sigma}_\mu)_{\dot{\beta}\alpha}. \tag{B.22c}$$

These identities allow one to derive further specific identities by contracting indices and making use of the relations (B.20) among the Pauli matrices. Examples are

$$(\xi\sigma^\mu\bar{\xi})(\xi\sigma^\nu\bar{\xi}) = \frac{1}{2}\xi^2\bar{\xi}^2\eta^{\mu\nu} \tag{B.23a}$$

$$\xi\psi\xi\sigma^\mu\bar{\eta} = -\frac{1}{2}\xi^2\psi\sigma^\mu\bar{\eta} \tag{B.23b}$$

$$\bar{\xi}\bar{\psi}\bar{\xi}\bar{\sigma}^\mu\eta = -\frac{1}{2}\bar{\xi}^2\bar{\psi}\bar{\sigma}^\mu\eta. \tag{B.23c}$$

B.1.3 Spinors and Tensors

We saw in the previous subsection that it is possible to form a Lorentz scalar $\xi\eta$ from two Weyl spinors. Lorentz vectors and higher-order tensors can also be built explicitly from these spinors. This is again due to the fact that $\mathbf{SL(2, C)}$ is the double universal covering group of $\mathbf{SO(3, 1)}$. The notations and techniques for the explicit construction were already prepared in Appendix A.3.4.

The σ^μ and $\bar{\sigma}^\mu$ matrices can be used to establish a one-to-one correspondence between any bi-spinor $V_{\alpha\dot{\alpha}}$ respectively $W_{\dot{\alpha}\alpha}$ and the Lorentz four vector V^μ or W^μ, respectively, by

$$V^\mu = \bar{\sigma}^{\mu\dot{\alpha}\alpha}V_{\alpha\dot{\alpha}} \leftrightarrow V_{\alpha\dot{\alpha}} = \frac{1}{2}V^\mu\sigma_{\mu\alpha\dot{\alpha}} \quad W^\mu = \frac{1}{2}\bar{\sigma}^{\mu\alpha\dot{\alpha}}W_{\dot{\alpha}\alpha} \leftrightarrow W_{\dot{\alpha}\alpha} = \frac{1}{2}W^\mu\sigma_{\mu\dot{\alpha}\alpha}. \tag{B.24}$$

One easily verifies that V^2 and W^2 are Lorentz scalars and that both V^μ and W^μ transform as Lorentz vectors, e.g.

$$V'^{\mu} = \bar{\sigma}^{\mu\dot{\alpha}\alpha} M_\alpha{}^\beta M^{*\dot{\gamma}}_{\dot{\alpha}} V_{\beta\dot{\gamma}} = (M^\dagger \bar{\sigma}^\mu M)^{\dot{\gamma}\beta} V_{\beta\dot{\gamma}} = \Lambda^\mu{}_\nu \bar{\sigma}^{\mu\dot{\gamma}\beta} V_{\beta\dot{\gamma}} = \Lambda^\mu{}_\nu V^\nu .$$

This construction of a vector from Weyl spinors as

$$V^\mu = \xi^\alpha \sigma^\mu_{\alpha\dot{\beta}} \bar{\eta}^{\dot{\beta}}$$

shows that one is dealing with a $(\frac{1}{2}, \frac{1}{2})$ representation.

Two spinors may also serve to construct objects

$$F_{\alpha\beta} = \frac{1}{2}(\xi_\alpha \chi_\beta + \xi_\beta \chi_\alpha), \qquad \bar{F}_{\dot{\alpha}\dot{\beta}} = \frac{1}{2}(\bar{\zeta}_{\dot{\alpha}} \bar{\eta}_{\dot{\beta}} + \bar{\zeta}_{\dot{\beta}} \bar{\eta}_{\dot{\alpha}})$$

and from these

$$F^+_{\mu\nu} = \frac{i}{4}(\epsilon \sigma_{\mu\nu})^{\alpha\beta} F_{\alpha\beta}, \qquad F^-_{\mu\nu} = -\frac{i}{4}(\bar{\sigma}_{\mu\nu}\epsilon^T)^{\dot{\alpha}\dot{\beta}} \bar{F}_{\dot{\alpha}\dot{\beta}}$$

which obey the duality relations

$$F^{\pm\mu\nu} = \pm \epsilon^{\mu\nu\rho\sigma} F^\pm_{\rho\sigma}.$$

Finally $F_{\mu\nu} = F^+_{\mu\nu} + F^-_{\mu\nu}$ is a real antisymmetric Lorentz tensor. Since $F^+_{\mu\nu}$ belongs to the $(1, 0)$-representation, and $F^-_{\mu\nu}$ to $(0, 1)$, $F_{\mu\nu}$ is associated to the representation $(1, 0) \oplus (0, 1)$.

B.1.4 Dirac and Majorana Spinors

A Dirac spinor can be built from two Weyl spinors in the form

$$\psi(x) := \begin{pmatrix} \xi_\alpha \\ \bar{\eta}^{\dot{\alpha}} \end{pmatrix} = \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix}. \tag{B.25}$$

It transforms under $\mathbf{SL(2, C)}$ as

$$\psi' = S(M)\psi = \begin{pmatrix} M\xi \\ M^{-1\dagger}\bar{\eta} \end{pmatrix}.$$

Explicitly (and infinitesimally) this is

$$\psi' = \begin{pmatrix} \xi'_\alpha \\ \bar{\eta}'^{\dot{\alpha}} \end{pmatrix} = \begin{pmatrix} \left(1 + \frac{1}{2}\omega_{\mu\nu}\sigma^{\mu\nu}\right)_\alpha{}^\beta & 0 \\ 0 & \left(1 + \frac{1}{2}\omega_{\mu\nu}\bar{\sigma}^{\mu\nu}\right)^{\dot{\alpha}}{}_{\dot{\beta}} \end{pmatrix} \begin{pmatrix} \xi_\beta \\ \bar{\eta}^{\dot{\beta}} \end{pmatrix}$$
$$= \left(I + \frac{1}{2}\omega_{\mu\nu} \begin{pmatrix} \sigma^{\mu\nu} & 0 \\ 0 & \bar{\sigma}^{\mu\nu} \end{pmatrix}\right) \begin{pmatrix} \xi \\ \bar{\eta} \end{pmatrix} = \left(I - \frac{i}{4}\omega_{\mu\nu}\gamma^{\mu\nu}\right)\psi. \tag{B.26}$$

where the expression for $\gamma^{\mu\nu}$ according to (B.13) was used. Therefore, the spin matrix for a Dirac field is

$$\Sigma_D^{\mu\nu} = \frac{i}{2}\gamma^{\mu\nu}.$$

The Weyl components–also called chirality eigenstates–are recovered from the Dirac spinor by projection operators

$$P_L = \frac{1 - \gamma^5}{2} \qquad P_R = \frac{1 + \gamma^5}{2} \tag{B.27}$$

as

$$\psi_L = P_L\psi = \begin{pmatrix} \xi_\alpha \\ 0 \end{pmatrix} \qquad \psi_R = P_R\psi = \begin{pmatrix} 0 \\ \bar{\eta}^{\dot{\alpha}} \end{pmatrix}.$$

The matrices A and C, implicitly given by (B.9), serve to define the adjoint and the charge conjugate spinor as

$$\overline{\psi} := \psi^\dagger A \qquad \psi^C := C\overline{\psi}^T = CA^T\psi^*. \tag{B.28}$$

The Dirac bilinears built from the basis (B.8) as $\overline{\psi}_1\Gamma^A\psi_2$ can be shown to obey

$$(\overline{\psi}_1\Gamma^\alpha\psi_2)^\dagger = +\overline{\psi}_2\Gamma^\alpha\psi_1 \quad \text{for} \quad \Gamma^\alpha = (1, \gamma^\mu, \gamma^5) \tag{B.29a}$$
$$(\overline{\psi}_1\Gamma^\alpha\psi_2)^\dagger = -\overline{\psi}_2\Gamma^\alpha\psi_1 \quad \text{for} \quad \Gamma^\alpha = (\gamma^\mu\gamma^5, \gamma^{\mu\nu}). \tag{B.29b}$$

In terms of the Weyl components and in the chiral basis, the adjoint becomes explicitly

$$\overline{\psi} = (\bar{\xi}_{\dot{\alpha}}, \ \ \eta^\alpha) \begin{pmatrix} 0 & \delta^{\dot{\alpha}}{}_{\dot{\beta}} \\ \delta_\alpha{}^\beta & 0 \end{pmatrix} = (\eta^\beta, \ \ \bar{\xi}_{\dot{\beta}}). \tag{B.30}$$

According to (B.11), numerically $A = \gamma^0$, but using the gamma matrix here would spoil the spinor index structure. One verifies that

$$\overline{\psi}_1\psi_2 = \eta_1\xi_2 + \bar{\xi}_1\bar{\eta}_2,$$

revealing that $\bar{\psi}_1\psi_2$ is a Lorentz scalar. The bilinears

$$\overline{\psi}_1\gamma^5\psi_2, \qquad \overline{\psi}_1\gamma^\mu\psi_2, \qquad \overline{\psi}_1\gamma^\mu\gamma^5\psi_2, \qquad \overline{\psi}_1\gamma^{\mu\nu}\psi_2$$

transform as pseudo-scalars, vectors, axial-vectors and tensors with respect to parity transformations.

Charge conjugation is in the chiral representation mediated by

$$C = \begin{pmatrix} \varepsilon_{\alpha\beta} & 0 \\ 0 & \varepsilon^{\dot\alpha\dot\beta} \end{pmatrix} = i\gamma^2\gamma^0 \tag{B.31}$$

that is,

$$\psi^C := C\bar\psi^T = \begin{pmatrix} \eta_\alpha \\ \bar\xi^{\dot\alpha} \end{pmatrix}. \tag{B.32}$$

As a consequence we derive the identity

$$\overline{\psi}^C\psi = \xi^2 + \bar\eta^2 = \overline{\psi}^C_R\psi_L + \overline{\psi}^C_L\psi_R. \tag{B.33}$$

The charge conjugate of a left-handed field is right-handed and vice versa. This is obvious in the Weyl notation: If for example $\psi = \begin{pmatrix} \xi_\alpha \\ 0 \end{pmatrix}$ then $\psi^C = \begin{pmatrix} 0 \\ \bar\xi^{\dot\alpha} \end{pmatrix}$. Furthermore,

$$\begin{aligned} \psi^C_L &= P_L\psi^C = P_L C\overline{\psi}^T = C(\overline{\psi}P_L)^T = C\overline{\psi}^T_R = C\gamma^0\psi^*_R \\ \overline{\psi}^C_L &= \psi^{C\dagger}_L\gamma^0 = \psi^{*\dagger}_R\gamma^{0\ T\dagger}C^\dagger\gamma^0 = -\psi^T_R C^{-1} = \psi^T_R C. \end{aligned}$$

Be aware that $(\psi^C)_L \neq (\psi_L)^C$ but $(\psi^C)_L = (\psi_R)^C$.

A *Majorana spinors* ψ_M is a Dirac spinor which is identical to its charge conjugate: $\psi^C_M = \psi_M$. Thus a Majorana spinor can be expressed in terms of a single Weyl spinor

$$\psi_M = \begin{pmatrix} \xi_\alpha \\ \bar\xi^{\dot\alpha} \end{pmatrix}. \tag{B.34}$$

It is possible to define two Majorana spinors from a Dirac spinor $\psi = \begin{pmatrix} \xi_\alpha \\ \bar\eta^{\dot\alpha} \end{pmatrix}$ as

$$\chi_L = \psi_L + (\psi^C)_R = \begin{pmatrix} \xi_\alpha \\ \bar\xi^{\dot\alpha} \end{pmatrix} \qquad \chi_R = \psi_R + (\psi^C)_L = \begin{pmatrix} \eta_\alpha \\ \bar\eta^{\dot\alpha} \end{pmatrix}. \tag{B.35}$$

Majorana spinors satisfy the identities

$$(\overline{\psi}_1\Gamma\psi_2) = +\overline{\psi}_2\Gamma\psi_1 \qquad \text{for} \qquad \Gamma = (1, \gamma^5, \gamma^\mu\gamma^5) \tag{B.36a}$$

$$(\overline{\psi}_1\Gamma\psi_2) = -\overline{\psi}_2\Gamma\psi_1 \qquad \text{for} \qquad \Gamma = (\gamma^\mu, \gamma^{\mu\nu}). \tag{B.36b}$$

B.2 $\mathbb{Z}_2$-Gradings

By the term $\mathbb{Z}_2$-gradings I denote in the following those mathematical techniques which allow one to speak of bosonic and fermionic degrees of freedom even in an non-quantized theory.

B.2.1 Definitions

A $\mathbb{Z}_2$-graded vector space is a vector space which admits the direct sum decomposition

$$V = V^{(0)} \bigoplus V^{(1)}.$$

Hereafter, I will leave off the $\mathbb{Z}_2$, and also refer to $V^{(0)}$ and $V^{(1)}$ as the even and odd subspaces. An element $X \in V$ is called 'homogeneous' if it belongs either to $V^{(0)}$ or to $V^{(1)}$. Homogeneous elements are assigned a *Grassmann parity* $|X|$ by the convention that $|X| = i$ for $X \in V^{(i)}$. Thus, homogeneous elements are either even/bosonic or odd/fermionic.

A graded algebra is a graded vector space $A = A^{(0)} \bigoplus A^{(1)}$ which is also an algebra, that is a structure with a linear and associative mapping $\Box : A \times A \to A$; see Appendix A.1.1. A graded algebra is called supercommutative if all homogeneous elements $X, Y \in A$ satisfy $X \Box Y = (-1)^{|X||Y|} Y \Box X$.

The elementary supercommutative graded algebra is the *Grassmann algebra*[5]. It consists of the set of generators $\{\psi_a\}$ $(a = 1, \ldots, N)$ with $\psi_a \psi_b + \psi_b \psi_a = 0$. An arbitrary element in this algebra (also called a 'supernumber') has the form

$$\Omega(\psi) = \omega_0 + \omega^a \psi_a + \frac{1}{2} \omega^{ab} \psi_a \psi_b + \ldots + \frac{1}{N!} \omega^{1 \ldots N} \psi_1 \psi_2 \ldots \psi_N \tag{B.37}$$

with real or complex coefficients $\omega^{a_1, \ldots, a_k}$ which can be considered completely antisymmetric in their indices.

B.2.2 Supertrace and Superdeterminant

In the space of supernumbers, one can establish the notions of supermatrices: The object

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

---

5 named after the German mathematician H.G. Grassmann (1809–1877).

built from submatrices A and D with even, and B and C with odd supernumbers constitutes a supermatrix. The standard matrix multiplication of two supermatrices is again a supermatrix. The supertrace of the supermatrix A is defined by

$$\text{str } M := \text{tr } A - \text{tr } D.$$

It obeys the usual relation for a trace of a matrix:

$$\text{str } (M_1 + M_2) = \text{str } M_1 + \text{str } M_2 \qquad \text{str } (M_1 M_2) = \text{str } (M_2 M_1).$$

The superdeterminant (sometimes called the *Berezinian* Ber) of a supermatrix is defined by

$$\text{Ber} = \text{sdet } M := \exp \text{str } \ln M$$

with

$$\ln M = \ln(I + M - I) = \sum_{i \geqslant 1} (-1)^{i-1} \frac{1}{i} (M - I)^i.$$

For matrices M which differ infinitesimally from the unit matrix ($M = I + \epsilon$):

$$\text{sdet } M \approx 1 + \text{str } \epsilon.$$

Furthermore,

$$\text{sdet } M = \det A[\det(D - CA^{-1}B)]^{-1} = (\det D)^{-1}[\det(A - BD^{-1}C)],$$

from which one can see that M is non-invertible if either A or D are non-invertible. Therefore, sdet M is defined only for invertible supermatrices M. If it exists, the superdeterminant fulfills

$$\text{sdet } (M_1 M_2) = (\text{sdet } M_1)(\text{sdet } M_2).$$

The (super)transpose of a supermatrix M is

$$M^T = \begin{pmatrix} A^T & C^T \\ -B^T & D^T \end{pmatrix}$$

with

$$(M_1 M_2)^T = M_2^T M_1^T \qquad (M^{-1})^T = (M^T)^{-1}$$
$$\text{str } M^T = \text{str } M \qquad \text{sdet } M^T = \text{sdet } M.$$

B.2.3 Differentiation and Integration

It is possible to define derivatives of the multinomials (B.37). Consider the variations

$$\delta\Omega(\psi) = \delta\psi_a \frac{\partial^R \Omega}{\partial\psi_a} = \frac{\partial^L \Omega}{\partial\psi_a}_L \delta\psi_a$$

in which a R(ight)- and a L(eft)-derivative is defined. Only for even elements in the Grassmann algebra are they identical. In this text, if not stated otherwise, R-derivatives are always used. For these,

$$\frac{\partial}{\partial\psi_a}(\psi_{a_1}\ldots\psi_{a_k}) = \delta_{aa_1}\psi_{a_2}\ldots\psi_{a_k} - \delta_{aa_2}\psi_{a_1}\psi_{a_3}\ldots\psi_{a_k} + \ldots(-1)^{k-1}\delta_{aa_k}\psi_{a_1}\ldots\psi_{a_{k-1}}. \tag{B.38}$$

As with ordinary differentiation, there is a product rule

$$\frac{\partial}{\partial\psi_\alpha}(\Omega_1\Omega_2) = \frac{\partial\Omega_1}{\partial\psi_\alpha}\Omega_2 + (-1)^{|\Omega_1|}\Omega_1\frac{\partial\Omega_2}{\partial\psi_\alpha}.$$

For $\psi'_a = A_{ab}\psi_b$, there also is a chain rule

$$\frac{\partial}{\partial\psi_a}\Omega(\psi'(\psi)) = \frac{\partial\Omega}{\partial\psi'_b}A_{ba}.$$

A more or less heuristic, but nevertheless very successful concept of integration within a Grassmann algebra[6] is due to F.A. Berezin. Berezin's recipe is motivated by requiring the definite integral over a Grassmann function to have properties of an ordinary integral in the limits $-\infty$ and $+\infty$, namely linearity and shift invariance. For instance, for one variable

$$\int_{-\infty}^{+\infty} dx\; cF(x) = c\int_{-\infty}^{+\infty} dx\; F(x) \qquad \int_{-\infty}^{+\infty} dx\; F(x+a) = \int_{-\infty}^{+\infty} dx\; F(x).$$

A Grassmann function in one odd variable necessarily has the form $f(\psi) = \alpha + \theta\beta$. Demanding that linearity and shift invariance hold for $\int d\psi\; f(\psi)$ leads to

$$\int d\psi\; f(\psi) = \beta \quad \text{or} \quad \int d\psi = 0, \qquad \int d\psi\;\psi = 1.$$

This is immediately extended to more than one Grassmann variable as

$$\int d\psi_a = 0 \qquad \int \psi_b d\psi_a = \delta_{ab} \tag{B.39}$$

[6] A sophisticated differential geometric meaning is elaborated in [414].

together with the linearity relation

$$\int (c_1\Omega_1 + c_2\Omega_2) d\psi_\alpha = c_1 \int \Omega_1 d\psi_\alpha + c_2 \int \Omega_2 d\psi_\alpha .$$

Observe a peculiarity arising in the change of variables: Take first the simple example of changing only one variable as in $\psi'_a = \lambda\psi_a$. Since

$$\int \psi'_a d\psi_a = \lambda \int \psi_a d\psi_a = \lambda,$$

consistency requires $d\psi'_a = \lambda^{-1} d\psi_a$. In general, for $\psi'_a = \lambda_{ab}\psi_b$:

$$\int \Omega\big(\psi(\psi')\big) d\psi'_N \ldots d\psi'_1 = (\det \lambda)^{-1} \int \Omega(\psi) d\psi_N \ldots d\psi_1 .$$

Thus changes in the integration of Grassmann odd variables transform with the inverse of the Jacobian, contrary to the case for "ordinary" even variables. If there are both odd and even variables $z^A = (x_k, \psi_a)$ one can show that

$$\int_{\mathcal{V}} F(z) dz = \int_{\mathcal{V}} F(x(x', \psi'), \psi(x', \psi')) \; \mathrm{sdet}\Big(\frac{\partial(x, \psi)}{\partial(x', \psi')}\Big).$$

This is immediately obvious for the case that even variables are transformed to even ones and odd variables to odd ones, where the super(Jacobian) factorizes as

$$\mathrm{sdet} \frac{\partial(x, \psi)}{\partial(x', \psi')} = \det\Big(\frac{\partial x}{\partial x'}\Big) \det{}^{-1}\Big(\frac{\partial \psi}{\partial \psi'}\Big).$$

And it can easily be proven for infinitesimal transformations.

The following calculation with Grassmann numbers seems to be a little preposterous, but the result is beneficial for formulating the functional-integral version of gauge theories:

$$I(A) = \int \prod_{k=1}^{N} d\psi_k d\tilde{\psi}_k \exp\Big(\sum_{i,j=1}^{N} \tilde{\psi}_i A_{ij} \psi_j\Big) = \det A. \tag{B.40}$$

In this expression $\psi_j, \tilde{\psi}_i$ are two distinct sets of Grassmann variables, and A is some real or complex matrix. This identity offers a seemingly strange way to write a determinant. The proof is straightforward: Think of expanding the exponential and integrating each summand. Due to the rules of Berezin integration (B.39) the only term that survives in this step is the Nth order term:

$$I(A) = \int \prod_{k=1}^{N} d\psi_k d\tilde{\psi}_k \frac{1}{N!}\Big(\sum_{i,j=1}^{N} \tilde{\psi}_i A_{ij} \psi_j\Big)^N .$$

But still by far not all terms contribute; instead,

$$I(A) = \Big(\int \prod_{k=1}^{N} d\psi_k d\tilde{\psi}_k \tilde{\psi}_k \psi_k\Big)\Big(\sum_{\mathrm{perm}} \epsilon^{i_1 i_2 \ldots i_N} A_{1i_1} A_{2i_2} \ldots A_{Ni_N}\Big) = \det A .$$

By the shift invariance of the Berezin integration, we can immediately generalize (B.40) to

$$I(A,\eta,\tilde{\eta}) = \int \prod_{k=1}^{N} d\psi_k d\tilde{\psi}_k \ \exp(\tilde{\psi}A\psi + \tilde{\eta}\psi + \tilde{\psi}\eta) = \det A \ \exp(-\tilde{\eta}A^{-1}\eta). \tag{B.41}$$

Another generalization can be obtained by additionally considering the Gaussian integration (D.4) with respect to even variables:

$$\int d^N z \exp\Big(-\frac{1}{2}z^A M_{AB} z^B\Big) = (2\pi)^{N/2}(\mathrm{sdet}M)^{-1/2},$$

where $M = M^T$ is an $N \times N$ supermatrix.

B.2.4 Pseudo-Classical Mechanics

Based upon the notion of Grassmann variables it is possible to formulate dynamical theories with bosonic and fermionic degrees of freedom; see e.g. [80]. Assume an action

$$S = \int dt \ L(q,\dot{q},\psi,\dot{\psi})$$

where the Lagrangian depends on even coordinates q_i and odd coordinates ψ_α. Define the momenta

$$p^i = \frac{\partial L}{\partial q_i} \qquad \pi^\alpha = \frac{\partial L}{\partial \psi_\alpha}.$$

Then from the Hamiltonian

$$H = \dot{q}_i p^i + \dot{\psi}_\alpha \pi^\alpha - L$$

the equations of motion are derived as

$$\dot{p}^i = -\frac{\partial H}{\partial q_i} \qquad \dot{q}_i = \frac{\partial H}{\partial p^i} \tag{B.42a}$$

$$\dot{\pi}^\alpha = -\frac{\partial H}{\partial \psi_\alpha} \qquad \dot{\psi}_\alpha = -\frac{\partial H}{\partial \pi^\alpha} \tag{B.42b}$$

if the derivatives from the right are used. (The minus sign in the last equation is not a mistake!) The Hamilton equations can be written in terms of generalized Poisson brackets or *Bose-Fermi brackets* $\{X,Y\}_{\mathrm{BF}}$:

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \{A,H\}_{\mathrm{BF}}.$$

The Bose-Fermi brackets are most compactly defined in the following way–motivated by the symplectic structure of “ordinary” mechanics: Call $(Z^A) = \{q_i, \psi_\alpha, p^i, \pi^\alpha\}$, then

$$\{F, G\}_{\mathrm{BF}} := \frac{\partial^R F}{\partial Z^A} C^{AB} \frac{\partial^L G}{\partial Z^B} \tag{B.43}$$

with $C^{AB} = \{Z^A, Z^B\}_{BF}$, or explicitly

$$\{q_i, p^j\}_{BF} = -\{p^j, q_i\}_{BF} = \delta_i^j, \quad \{\psi_\alpha, \pi^\beta\}_{BF} = \{\pi^\beta, \psi_\alpha\}_{BF} = -\delta_\alpha^\beta, \tag{B.44}$$

all others vanishing. Thus we not only get back the Poisson brackets for bosonic phase-space variables, but also brackets for the fermionic phase-space variables that prepare the commutator relations for fermionic quantum operators or–conversely–are the classical limit for a fermionic system. This of course is taken to be only formal, but it is a successful calculational device. In the following–and in all those cases where use is made of the Bose-Fermi bracket–the index BF on the bracket will be omitted.

The Bose-Fermi brackets obey $\{F, G\} = -(-1)^{|F||G|}\{F, G\}$ and therefore they qualify as a supercommutative graded algebra with the identification $F \,\square\, G = \{F, G\}$. Furthermore

$$\begin{aligned}
\{A, B + C\} &= \{A, B\} + \{A, C\} \\
\{A, BC\} &= (-1)^{|A||B|} B\{A, C\} + \{A, B\}C \\
(-1)^{|A||C|}\{\{A, B\}, C\} &+ (-1)^{|B||A|}\{\{B, C\}, A\} + (-1)^{|C||B|}\{\{C, A\}, B\} = 0,
\end{aligned}$$

which makes it a graded Lie algebra[7].

Every Lie group has an associated Lie algebra. In extending the group concept to graded Lie groups there is also the notion of a graded Lie algebra. This is spanned by a set of even and odd generators $X = \{X^A\}$ with a mapping $X \times X \to X$:

$$\{X^A, X^B\} := X^A X^B - (-1)^{|A||B|} X^B X^A = \sum_C \gamma^{ABC} X^C \tag{B.45}$$

with the shorthand notation $|A| = |X_A|$, i.e. $+1(-1)$ if $|X_A|$ is bosonic(fermionic). The graded Lie algebra properties request that the structure functions γ^{ABC} must satisfy the conditions

$$\begin{aligned}
&\gamma^{BAC} + (-1)^{|A||B|}\gamma^{ABC} = 0 \\
&(-1)^{|A||C|}\gamma^{ABD}\gamma^{DCE} + (-1)^{|B||A|}\gamma^{BCD}\gamma^{DAE} + (-1)^{|C||B|}\gamma^{CAD}\gamma DBE = 0.
\end{aligned}$$

[7] The designations are not unique in the literature: You find the names “superalgebra” and “super Lie algebra”, although mathematically there are subtle differences [101].

B.3 *Supergeometry

The supermathematics behind local and global supersymmetry became quite advanced from the 1980s on. If you are interested in this aspect of supersymmetry you might like to refer to [62, 122].

B.3.1 Superspace

Supersymmetry is most compactly and elegantly described in terms of superfields. These are defined in superspace, an "upgrade" of Minkowski space. Fields in Minkowski space are in one-to-one correspondence with its isotropy group. Superfields are in a similar way tight to the symmetry group of superspace.

The idea of superspace was developed by A. Salam and J. Strathdee in 1974 [457]. Superspace has, besides the coordinates x^μ ($\mu = 0, \ldots, 3$), which commute, additional anti-commuting coordinates θ^a, ($a = 1, \ldots, 3$). (Everything can of course be generalized to higher dimensions.)

Remembering that in the case of Minkowski space, the elements of the isometry group are directly linked to the motions in Minkowski space, we expect for the superspace a group element of the form

$$G(x, \theta) = e^{i(x^\mu T_\mu + \bar{\theta}^a Q_a)}. \tag{B.46}$$

Multiplying two group elements we apply the Campbell-Baker-Hausdorff's formula (A.6)

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\ldots}$$

generalized to Grassmann objects. Calculate for instance $G(0, \epsilon)G(x, \theta)$. Here $A = i\bar{\epsilon}^a Q_a$ and $B = i(x^\mu P_\mu + \bar{\theta}^b Q_b)$ for which–due to the Wess-Zumino algebra (specifically (8.58) and (8.57))–one obtains

$$\begin{aligned}[A, B] &= \Big[i(\bar{\epsilon}^a Q_a), i(x^\mu P_\mu + \bar{\theta}^b Q_b)\Big] = -\Big[\bar{\epsilon}^a Q_a, \bar{Q}_b \theta^b\Big] \\ &= -\bar{\epsilon}^a \{Q_a, \bar{Q}_b\}\theta^b = -2(\bar{\epsilon}^a (\gamma^\mu)_{ab}\theta^b) P_\mu.\end{aligned}$$

Since the higher nested brackets of A and B vanish in this case, the result is

$$G(0, \epsilon)G(x^\mu, \theta) = e^{i\left[\bar{\epsilon}^a Q_a + x^\mu T_\mu + i\bar{\theta}^b Q_b + i(\bar{\epsilon}^a(\gamma^\mu)_{ab}\theta^b)T_\mu\right]} = G(x + i\bar{\epsilon}\gamma^\mu\theta,\ \theta + \epsilon).$$

This induces a translation in superspace

$$g(\epsilon) : (x^\mu, \theta^a) \to (x^\mu + i\bar{\epsilon}^a(\gamma^\mu)_{ab}\theta^b,\ \theta^a + \epsilon^a). \tag{B.47}$$

As it turns out, in supersymmetry it is much more appropriate and convenient to work with two-component Weyl spinorial coordinates. Superspace is thus "coordinized" by $z^M = (x^\mu, \theta^\alpha, \bar\theta_{\dot\alpha})$. The equivalent to (B.46) is

$$G(x^\mu, \theta^\alpha, \bar\theta_{\dot\alpha}) = e^{i(x^\mu T_\mu + \theta^\alpha Q_\alpha + \bar\theta_{\dot\alpha}\bar Q^{\dot\alpha})}$$

with

$$g(\epsilon, \bar\epsilon) : (x^\mu, \theta^\alpha, \bar\theta_{\dot\alpha}) \to (x^\mu + \xi^\mu,\ \theta^\alpha + \epsilon^\alpha,\ \bar\theta_{\dot\alpha} + \bar\epsilon_{\dot\alpha}) \tag{B.48}$$

where

$$\xi^\mu = i\epsilon\sigma^\mu\bar\theta - i\theta\sigma^\mu\bar\epsilon. \tag{B.49}$$

The operators that generate these transformations are the supercharges

$$Q_\alpha = \frac{\partial}{\partial\theta^\alpha} - i(\sigma^\mu)_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu \qquad \bar Q_{\dot\alpha} = -\frac{\partial}{\partial\bar\theta^{\dot\alpha}} + i\theta^\beta(\sigma^\mu)_{\beta\dot\alpha}\partial_\mu \tag{B.50}$$

obeying the anti-commutator relations (8.60).

B.3.2 Superfields

Superfields $\mathsf{F}(z^M)$ are mappings of the superspace to a Grassmann algebra, since a Taylor expansion of $\mathsf{F}(x^\mu, \theta^\alpha, \bar\theta_{\dot\alpha})$ in the odd coordinates has at most terms quadratic in θ^α and $\bar\theta_{\dot\alpha}$. Therefore, a general superfield can be written as

$$\begin{aligned}\mathsf{F}(x, \theta, \bar\theta) = f(x) + \theta\,\phi(x) + \bar\theta\,\bar\chi(x) + \theta^2\,m(x) + \bar\theta^2\,\bar n(x) + (\theta\sigma^\mu\bar\theta)\,v_\mu(x)\\ + \theta^2\bar\theta\,\bar\lambda(x) + \bar\theta^2\theta\,\psi(x) + \theta^2\bar\theta^2\,d(x).\end{aligned} \tag{B.51}$$

Here and in the following, I comply with the prevailing convention that when indices are not written explicitly, undotted indices are meant to be descending and dotted indices are meant to be ascending; see (B.15, B.18).

It is easy to verify that the sum of two superfields is again a superfield. Also, a linear combinations of superfields is again a superfield. But notice that also the product of two superfields is again a superfield. The derivative of a superfield is defined through $\delta\mathsf{F} = \delta z^M \frac{\partial\mathsf{F}}{\partial z^M}$ and

$$\frac{\partial}{\partial z^M}(\mathsf{F}_1\mathsf{F}_2) = \frac{\partial\mathsf{F}_1}{\partial z^M}F_2 + (-1)^{|z^M|}\mathsf{F}_1\frac{\partial\mathsf{F}_2}{\partial z^M}.$$

The superfield (B.51) contains as component fields

4 complex (pseudo) scalar fields:	f, m, n, d
4 Weyl spinor fields:	$(\phi, \psi) \in \left(\frac{1}{2}, 0\right), \qquad (\chi, \lambda) \in \left(0, \frac{1}{2}\right)$
1 Lorentz four-vector field:	v_μ.

Thus, there are 16 bosonic and 16 fermionic field components. Let us reflect the mass dimensions of the different quantities. The superspace coordinate transformations

(B.47) imply that ϵ and θ have the same mass dimension, namely $-1/2$. Assuming that the scalar superfield f has mass dimension $[f] = 0$, we find

$$\left[\{f, \phi, \bar{\chi}, m, \bar{n}, v_\mu, \bar{\lambda}, \psi, d\}\right] = \left\{0, \frac{1}{2}, \frac{1}{2}, 1, 1, 1, \frac{3}{2}, \frac{3}{2}, 2\right\}. \tag{B.52}$$

Next, we are interested in the behavior of a superfield under a supersymmetry transformation with parameters $\epsilon, \bar{\epsilon}$:

$$\mathsf{F}(x, \theta, \bar{\theta}) \to \mathsf{F}(x + \xi, \theta + \epsilon, \bar{\theta} + \bar{\epsilon}), \tag{B.53}$$

where ξ is the expression (B.49). Again, for most purposes it is sufficient to consider infinitesimal transformations

$$\delta_{\epsilon\bar{\epsilon}}\mathsf{F}(x, \theta, \bar{\theta}) := \left\{\xi^\mu \partial_\mu + \epsilon \frac{\partial}{\partial\theta} + \bar{\epsilon}\frac{\partial}{\partial\bar{\theta}}\right\}\mathsf{F}. \tag{B.54}$$

A straightforward (although quite lengthy) calculation yields the explicit transformations of the component fields. In fact, the transformation rules of the fields can be deduced–up to numerical constants–by observing the balance of mass dimensions and saturation of Lorentz and Weyl indices. Take for instance the transformation of f. It must be linear in ϵ and $\bar{\epsilon}$: $\delta f = f_1\epsilon + f_2\bar{\epsilon}$. Now $[\delta f] = 0$, and from $[\epsilon] = -1/2$ we deduce $[f_i] = +1/2$. The only fields at our disposal carrying this mass dimension are ϕ and $\bar{\chi}$. No derivative (which has $\left[\partial_\mu\right] = +1$) applied on any of the fields yields terms with mass dimension $+1/2$. Thus $[\delta f]$ must be a linear combination of $\epsilon\phi$ and $\bar{\epsilon}\bar{\chi}$ with numerical coefficients. For another example, take the component field d for which generically $\delta d = d_1\epsilon + d_2\bar{\epsilon}$. Since $[\delta d] = 2$ we read off $[d_i] = 5/2$. Component fields with this mass dimension do not exist in the superfield F. The only way to generate terms with the requested mass dimension is by taking the derivatives of the fields $\bar{\lambda}$ and ψ. Thus

$$\delta d = d_1^\mu \partial_\mu \bar{\lambda}\epsilon + d_2^\mu \partial_\mu \psi\bar{\epsilon} = \tilde{d}_1 \epsilon\sigma^\mu\partial_\mu\bar{\lambda} + \tilde{d}_2\bar{\epsilon}\bar{\sigma}^\mu\partial_\mu\psi,$$

where the matrices σ and $\bar{\sigma}$ make their appearance because they are the only objects around by which the Lorentz indices and the Weyl indices can be saturated. (The result that the highest component in the superfield expansion transforms as a derivative will be seen to be of utmost importance in the construction of supersymmetric actions.) The only things left open are numerical constants $\tilde{d}_i$. In a quite similar way, one may derive the transformation rules of the other fields. I refrain from doing this here, since the most generic superfield is not of prime importance–mainly because it is not irreducible.

A superfield is reducible if it contains a subset of fields which is closed under supersymmetry transformations. One may identify irreducible multiplets by imposing restrictions on the superfield that are compatible with supersymmetry. It turns out that two types of specific superfields are sufficient to construct the most general actions in superspace, namely vector and chiral superfields. As will be seen later, this conforms with fields occurring in the Standard Model, namely chiral fermions and vector bosons.

Chiral Superfields

For the specification of a chiral superfield, we need the notion of covariant spinor derivatives, defined as

$$D_\alpha = \frac{\partial}{\partial\theta^\alpha} + i(\sigma^\mu)_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu \qquad \bar D_{\dot\alpha} = -\frac{\partial}{\partial\bar\theta^{\dot\alpha}} - i\theta^\beta(\sigma^\mu)_{\beta\dot\alpha}\partial_\mu. \tag{B.55}$$

These anti-commute with the generators of the supersymmetry transformations (B.50) and all the Lorentz generators. Furthermore,

$$\{D_\alpha, D_\beta\} = 0 = \{\bar D_{\dot\alpha}, \bar D_{\dot\beta}\} \tag{B.56a}$$

$$\{D_\alpha, \bar D_{\dot\beta}\} = -2i(\sigma^\mu)_{\alpha\dot\beta}\partial_\mu. \tag{B.56b}$$

These operators have the same mass dimension as the superymmetry generators, namely $[D] = +1/2$.

A chiral superfield (sometimes called a left-chiral superfield) is indirectly defined by the condition

$$\bar D_{\dot\alpha}\Phi = 0. \tag{B.57}$$

This constraint on a generic superfield can conveniently be solved in terms of the variables θ and

$$y^\mu := x^\mu + i\theta\sigma^\mu\bar\theta,$$

since

$$\bar D_{\dot\alpha} y^\mu := \bar D_{\dot\alpha}(x^\mu + i\theta\sigma^\mu\bar\theta) = 0 \qquad \bar D_{\dot\alpha}\theta = 0.$$

Thus any superfield depending only on y^μ and θ is a chiral superfield. Expanded in the odd coordinates

$$\begin{aligned}\Phi &= A(y) + \sqrt{2}\,\theta\,\psi(y) + \theta^2\,F(y)\\ &= \left[A(x) + i(\theta\sigma^\mu\bar\theta)\,\partial_\mu A(x) - \frac{1}{2}(\theta\sigma^\mu\bar\theta)(\theta\sigma^\nu\bar\theta)\,\partial_\mu\partial_\nu A(x)\right]\\ &\qquad + \left[\sqrt{2}\theta\,\psi(x) + i\sqrt{2}(\theta\sigma^\mu\bar\theta)\theta\,\partial_\mu\psi(x)\right] + \theta^2\,F(x)\end{aligned} \tag{B.58}$$

where the terms in the squared brackets are Taylor expansion around x–the expansion again yielding only few terms due to the Grassmann character of the θ- and $\bar\theta$-coordinates. By using the Fierz identities in the form (B.23) the latter expression can be written as

$$\Phi(x) = A + \sqrt{2}\theta\,\psi + \theta^2\,F + i(\theta\sigma^\mu\bar\theta)\,\partial_\mu A + \frac{i}{\sqrt{2}}\theta^2(\bar\theta\bar\sigma^\mu\partial_\mu\psi) - \frac{1}{4}\theta^2\bar\theta^2\,\partial_\mu\partial^\mu A.$$

On the other hand, this is the most general solution to condition (B.57), since in terms of $(y, \theta, \bar{\theta})$ the spinor derivatives (B.55) are

$$D_\alpha = \frac{\partial}{\partial\theta^\alpha} + 2i(\sigma^\mu)_{\alpha\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\frac{\partial}{\partial y^\mu} \qquad \bar{D}_{\dot{\alpha}} = -\frac{\partial}{\partial\bar{\theta}^{\dot{\alpha}}}.$$

The chiral superfield (B.58) has in its multiplet two complex scalars A and F, and a Weyl spinor ψ_α. It seems that there is a mismatch between the four bosonic and the two fermionic degrees of freedom. However, as seen later, not all bosonic fields represent physical degrees of freedom since the F-fields do not propagate. In any case, with regard to a supersymmetric extension of the standard model, the chiral superfield contains not only the chiral quarks and leptons (in terms of ψ), but also a scalar (Higgs) boson.

Recalling that θ has mass dimension $[\theta] = -1/2$, and assigning the usual mass dimension $[A] = +1$ we deduce the (usual) mass dimension $+3/2$ for the fermionic field ψ_α, and the (unusual) mass dimension $[F] = +2$.

The superfield $\Phi^\dagger$ obeys $D_\alpha\Phi^\dagger = 0$. It is a function of $y^{*\mu} := x^\mu - i\theta\sigma^\mu\bar{\theta}$ and $\bar{\theta}$ only, also called a right-chiral or an anti-chiral superfield. The explicit form of $\Phi^\dagger$ in terms of components is obtained from (B.58) by conjugation:

$$\Phi^\dagger = A^*(y^*) + \sqrt{2}\bar{\theta}\,\bar{\psi}(y^*) + \bar{\theta}^2\, F^*(y^*).$$

If a supersymmetry transformation (B.54) is applied to a chiral superfield, it turns out to be convenient to phrase this in the variables y and θ as

$$\begin{aligned}\delta_{\epsilon\bar{\epsilon}}\Phi(y,\theta) &= \{\epsilon\frac{\partial}{\partial\theta} + 2i\theta\sigma^\mu\bar{\epsilon}\frac{\partial}{\partial y^\mu}\}\Phi \\ &= \sqrt{2}\,\epsilon\,\psi + 2\epsilon\theta\,F + 2i\theta\sigma^\mu\bar{\epsilon}\,\partial_\mu A + 2\sqrt{2}i\theta\sigma^\mu\bar{\epsilon}\theta\,\partial_\mu\psi \\ &\equiv \delta_{\epsilon\bar{\epsilon}}\,A + \sqrt{2}\,\theta\delta_{\epsilon\bar{\epsilon}}\,\psi + \theta^2\delta_{\epsilon\bar{\epsilon}}\,F.\end{aligned} \tag{B.59}$$

The last line displays the fact that the supersymmetry algebra closes, i.e. that the supersymmetry transformation of a chiral superfield is again a chiral superfield. By comparison of the last two expressions in (B.59) we deduce the transformation behavior of the different fields:

$$\delta_{\epsilon\bar{\epsilon}}A = \sqrt{2}\epsilon\psi \tag{B.60a}$$

$$\delta_{\epsilon\bar{\epsilon}}\psi = \sqrt{2}\epsilon F + i\sqrt{2}\sigma^\mu\bar{\epsilon}\partial_\mu A \tag{B.60b}$$

$$\delta_{\epsilon\bar{\epsilon}}F = i\sqrt{2}\bar{\epsilon}\bar{\sigma}^\mu\partial_\mu\psi. \tag{B.60c}$$

In order to obtain the transformation for the F field one needs to make use of the identity $\theta^\alpha\theta^\beta = -\frac{1}{2}\epsilon^{\alpha\beta}\theta^2$ according to (B.22). Notice that the F field transforms into a total derivative. The transformations (B.60) are related to those of the fields in the Wess-Zumino model (8.44a, 8.48, 8.49) if one defines from S, P and from the F_i the complex fields

$$A := \frac{1}{\sqrt{2}}(S + iP) \qquad F := \frac{1}{\sqrt{2}}(F_1 + iF_2). \tag{B.61}$$

Products of chiral superfields are again chiral superfields. The product of two left-chiral superfields for instance is

$$\begin{aligned}\Phi_i\Phi_j &= (A_i + \sqrt{2}\theta\psi_i + \theta\theta F_i)(A_j + \sqrt{2}\theta\psi_j + \theta\theta F_j)\\ &= A_iA_j \ + \ \sqrt{2}\theta(\psi_i A_j + A_i\psi_j) \ + \ \theta^2(A_iF_j + A_jF_i - \psi_i\psi_j).\end{aligned}$$

It has become conventional in the supersymmetry community to define

$$\Phi_i\Phi_j\big|_F := \int d^2\theta \ \ \Phi_i\Phi_j = A_iF_j + A_iF_j - \psi_i\psi_j.$$

By induction, also the product of any number of left-chiral superfields is a left-chiral superfield. Thus for instance

$$\begin{aligned}\Phi_i\Phi_j\Phi_k\big|_F &:= \int d^2\theta \ \ \Phi_i\Phi_j\Phi_k\\ &= F_iA_jA_k + F_jA_kA_i + F_kA_iA_j - \psi_i\psi_jA_k - \psi_j\psi_kA_i - \psi_k\psi_jA_i.\end{aligned}$$

Everything derived above for (left-)chiral superfields applies *mutatis mutandis* to right-chiral superfields.

Vector Superfields

A superfield is called a vector superfield $V(x, \theta, \bar{\theta})$ if it is self-conjugate $(V = V^\dagger)$. (The naming is a little confusing here. The term "vector" superfield originates from the fact that it contains a real Minkowski vector field v_μ.) This restriction cuts the number of fields in the general superfield (B.51) in half. Expressed as a power series expansion in $(\theta^\alpha, \bar{\theta}_{\dot{\alpha}})$ a vector superfield can be written as

$$\begin{aligned}V(x, \theta, \bar{\theta}) = C(x) &+ i\theta\chi(x) + i\bar{\theta}\bar{\chi}(x)\\ &+ \frac{i}{2}\theta^2\,[M(x) + iN(x)] - \frac{i}{2}\bar{\theta}^2\,[M(x) - iN(x)] - \theta\sigma^\mu\bar{\theta}\ v_\mu(x)\\ &+ i\theta^2\bar{\theta}\left[\bar{\lambda}(x) + \frac{i}{2}\bar{\sigma}^\mu\partial_\mu\chi(x)\right] - i\bar{\theta}^2\theta\left[\lambda(x) + \frac{i}{2}\sigma^\mu\partial_\mu\bar{\chi}(x)\right]\\ &+ \frac{1}{2}\theta^2\bar{\theta}^2\left[D(x) + \frac{1}{2}\Box C(x)\right].\end{aligned} \tag{B.62}$$

where C, D, M, N and v_μ are real fields[8]. The notation of the component fields as in (B.62) will become comprehensible immediately. Applying a supersymmetric transformation (B.54) to a vector superfield in order to find the transformations for its component fields is a little elaborate and gives lengthy expressions (as compared to the chiral superfield). In any case, as will become clear in the next section, the only important result is

$$\delta_{\epsilon\bar{\epsilon}} D = -\epsilon\sigma^\mu\, \partial_\mu \bar{\lambda} + \bar{\epsilon}\sigma^\mu\, \partial_\mu \lambda. \tag{B.63}$$

This shows that the component with $\theta^2\bar{\theta}^2$ transforms into a total derivative.

There are two distinguished vector superfields, both derived from chiral and anti-chiral superfields, and both serving as building blocks of supersymmetric actions. These are for one $(\Phi + \Phi^\dagger)$, and secondly $\Phi^\dagger\Phi$.

- Take

$$\begin{aligned}\Phi + \Phi^\dagger =& (A + A^*) + \sqrt{2}(\theta\psi + \bar{\theta}\bar{\psi}) + (\theta^2 F + \bar{\theta}^2 F^*) + i\theta\sigma^\mu\bar{\theta}\partial_\mu(A - A^*)\\ &+ \frac{i}{\sqrt{2}}\theta^2\bar{\theta}\bar{\sigma}^\mu\partial_\mu\psi + \frac{i}{\sqrt{2}}\bar{\theta}^2\bar{\theta}\sigma^\mu\partial_\mu\bar{\psi} + \frac{1}{4}\theta^2\bar{\theta}^2\Box(A + A^*).\end{aligned}$$

  The appearance of a gradient of the scalar field $(A\text{–}A^*)$ motivates the definition

$$V \longmapsto V + (\Phi + \Phi^\dagger) \tag{B.64}$$

  as the superspace version of a $\mathbf{U(1)}$ gauge transformation. Under this transformation, the component fields in (B.62) transform as

$$\begin{aligned}C &\longmapsto C + (A + A^*)\\ \chi &\longmapsto \chi - i\sqrt{2}\psi\\ (M + iN) &\longmapsto (M + iN) - 2iF\\ v_\mu &\longmapsto v_\mu - i\partial_\mu(A - A^*)\\ \lambda &\longmapsto \lambda\\ D &\longmapsto D.\end{aligned}$$

  From this we learn two things: First, the fields λ and D are invariant under (B.64). And secondly, since the fields C, χ, M, N simply transform by shifts, there is a gauge choice through which these fields can be brought to vanish. In this *Wess-Zumino gauge* the vector superfield simplifies to

$$V(x, \theta, \bar{\theta})\big|_{WZ} = -\theta\sigma^\mu\bar{\theta}v_\mu + i\theta^2\bar{\theta}\bar{\lambda} - i\bar{\theta}^2\theta\lambda(x) + \frac{1}{2}\theta^2\bar{\theta}^2 D. \tag{B.65}$$

  Of course, in choosing this gauge, supersymmetry is broken, but it still allows the "ordinary" gauge transformations $v_\mu \mapsto v_\mu + \partial_\mu f$. Later, in gauging a super-

[8] In the naming of the component fields, I follow the established notation, which of course is historically influenced.

symmetric theory, use will be made of the exponential e^V, which in the WZ-gauge becomes the finite expression

$$e^V\Big|_{WZ} = 1 - \theta\sigma^\mu\bar{\theta}v_\mu + i\theta^2\bar{\theta}\bar{\lambda} - i\bar{\theta}^2\theta\lambda(x) + \frac{1}{2}\theta^2\bar{\theta}^2\left[D - \frac{1}{2}v_\mu v^\mu\right] \qquad \text{(B.66)}$$

because $V^k\big|_{WZ} = 0$ for $k > 2$.

- Since $\Phi^\dagger\Phi$ is obviously self-conjugate, it is a vector superfield *per se*. For reasons explained below, one is interested only in the $\theta^2\bar{\theta}^2$-term:

$$\begin{aligned}\Phi^\dagger\Phi\Big|_D &:= \int d^2\theta d^2\bar{\theta}\ \Phi^\dagger\Phi \\ &= F^*F + \frac{1}{4}A^*\Box A + \frac{1}{4}\Box A^*A - \frac{1}{2}\partial_\mu A^*\partial^\mu A \\ &\quad + \frac{i}{2}(\partial_\mu\bar{\psi}\bar{\sigma}^\mu\psi - \bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi).\end{aligned} \qquad \text{(B.67)}$$

B.3.3 Superactions

In Sect. 5.3 action functionals are constructed for fields which are irreducible under Poincaré transformations. The respective Lagrangians become largely fixed by power counting arguments. With the techniques from superspace and superfields, one can bring forth super-Lagrangians and superactions in a similar manner.

By definition of variational symmetry, we want the action to be invariant under SuSy transformations. This is satisfied if $\mathcal{L}$ transforms into a total derivative. As a matter of fact, $\delta\mathcal{L} \neq 0$, since otherwise this would imply that $[\delta_1, \delta_2]\mathcal{L} \sim \partial\mathcal{L} = 0$, or that $\mathcal{L}$ must be a constant. In the previous treatment of superfields, we already met objects which transform into a total derivative, namely the highest components of chiral superfields (the "F-term") and of vector superfields (the "D-term"). Thus schematically a viable action is

$$S = \int d^4x\left(\int d^2\theta\ \mathcal{L}_F + \int d^2\bar{\theta}\ \mathcal{L}_F^\dagger + \int d^2\theta d^2\bar{\theta}\ \mathcal{L}_D\right) \qquad \text{(B.68)}$$

where $\mathcal{L}_F$ and $\mathcal{L}_D$ are a chiral and a vector superfield.

Lagrangian Components Constructed from Chiral Superfields

As shown, products of chiral superfields (and products of anti-chiral superfields) are chiral (anti-chiral) superfield themselves. These are thus possible components in $\mathcal{L}_F$. However, multiplying more then three chiral superfields results in terms with mass dimension $d > 4$, which would lead to non-renormalizable interactions. Thus there are three possible contributions to $\mathcal{L}_F$ which may be combined to

$$\mathcal{L}_F = \mathcal{W}(\Phi_i) := c_i \Phi_i + \frac{1}{2} m_{ij} \Phi_i \Phi_j + \frac{1}{3} g_{ijk} \Phi_i \Phi_j \Phi_k, \tag{B.69}$$

called the *superpotential*[9].

What about kinetic energy terms, that is terms with derivatives? Derivatives are present in the D-component of $\Phi^\dagger \Phi$, see (B.67). If this is taken as $\mathcal{L}_D$ in (B.68), it allows for a straightforward interpretation in terms of interacting fields: First of all it contains the proper kinetic terms for the scalar A and the fermion ψ. Furthermore, there is no kinetic term for F. This field is not propagating; it is an auxiliary field. As a matter of fact, dimensional arguments exclude higher powers of $\Phi^\dagger \Phi$, since $\left[\Phi^\dagger \Phi\right] = 2$.

Thus we arrive at the action that describes the interacting field theory of a chiral superfield:

$$S = \int d^8 z \;\; \Phi^\dagger \Phi + \int d^6 z \; \mathcal{W}(\Phi_i) + \int d^6 \bar{z} \; \mathcal{W}^\dagger(\Phi_i). \tag{B.70}$$

The simplest supersymmetric theory is the one consisting of the kinetic term only: The Lagrangian for the free massless Wess-Zumino model (8.43) is simply $\Phi^\dagger \Phi|_D$ with the identifications (B.61) except for total derivatives and after using the "field equation" $F = 0$. And we regain the massive and interacting Wess-Zumino model (8.52) by adding a superpotential, i.e.

$$\begin{aligned}
\mathcal{L}_{WZ} &= \Phi^\dagger \Phi\big|_D + [(\frac{1}{2} m \Phi^2 + \frac{1}{3} g \Phi^3) + h.c.]\big|_F \\
&= F^* F - |\partial_\mu A|^2 + i\bar{\psi}\bar{\sigma}^\mu \partial_\mu \psi \\
&\quad + \Big[\big(m\{AF - \psi^2\} + g\,\{FA^2 - \psi^2\, A\}\big) + h.c.\Big],
\end{aligned} \tag{B.71}$$

again up to a total derivative. The field equations with respect to F and F^* are

$$[\mathcal{L}_{WZ}]^F = F^* + mA + gA^2 = 0, \qquad [\mathcal{L}_{WZ}]^{F^*} = F + m^* A^* + g^* A^{*2} = 0$$

which shows that the auxiliary fields can be expressed by the scalar fields. In observing that the terms depending on F, F^* in (B.71) can be rewritten as

$$\begin{aligned}
&F F^* + F(mA + gA^2) + F^*(m^* A^* + g^* A^{*2}) \\
&\qquad\qquad = |F + m^* A^* + gA^{*2}|^2 - |(mA + gA^2)|^2,
\end{aligned}$$

and that the first terms vanishes on-shell, (B.71) becomes

$$\mathcal{L}_{WZ} = -|\partial_\mu A|^2 + i\bar{\psi}\bar{\sigma}^\mu \partial_\mu \psi - m(\psi^2 + \bar{\psi}^2) - g\psi^2\, A - g\bar{\psi}^2\, A^* - V(A, A^*). \tag{B.72}$$

Note that the scalar potential $V = |(mA + gA^2)|^2 \geqslant 0$, and that one is no longer free to add an arbitrary constant as one could do in a non-supersymmetric theory.

[9] This superpotential is by no means related to the superpotential arising in the context of Klein-Noether identities and for energy-momentum pseudo-tensors in GR.

Note further that supersymmetry enforces relations among masses and couplings

$$m_A = m = m_\psi \qquad \lambda_{A^4} = \lambda_{\psi\psi A}.$$

The expression (B.67) can be generalized to a term built from different chiral superfields and thus allows for a kinetic term

$$\mathcal{L}^H = H^{ij}\Phi_i^\dagger \Phi_j\big|_D$$

where H must be a positive definite Hermitean matrix. By taking appropriate linear combinations of the superfields, one may assume that $H^{ij} = \delta^{ij}$:

$$\mathcal{L}^{kin} = \sum_i \Phi_i^\dagger \Phi_i\big|_D. \tag{B.73}$$

The most general supersymmetric Lagrangian (built out of chiral and anti-chiral superfields) is made from the sum of this kinetic term and the superpotential (B.69):

$$\begin{aligned} \mathcal{L} &= \mathcal{L}^{kin} + \left[\mathcal{W}(\Phi_i)\big|_F + h.c.\right] \\ &= F_i^* F_i - |\partial_\mu A_i|^2 + i\bar{\psi}_i \bar{\sigma}^\mu \partial_\mu \psi_i + \left[\frac{\partial \mathcal{W}(A_i)}{\partial A_j} F_j - \frac{1}{2}\frac{\partial^2 \mathcal{W}(A_i)}{\partial A_j \partial A_k}\psi_j\psi_k + \text{h.c.}\right]. \end{aligned} \tag{B.74}$$

Again the F-fields are non-propagating due to their Euler-Lagrange equations

$$F_k^* + \left[\frac{\partial \mathcal{W}(A_i)}{\partial A_k}\right] = 0, \qquad F_k + \left[\frac{\partial \mathcal{W}(A_i)}{\partial A_k}\right]^* = 0. \tag{B.75}$$

Plugging this back into (B.74) one obtains

$$\mathcal{L} = -|\partial_\mu A_i|^2 + i\bar{\psi}_i \bar{\sigma}^\mu \partial_\mu \psi_i - \left|\frac{\partial \mathcal{W}(A_i)}{\partial A_k}\right|^2 - \left[\frac{1}{2}\frac{\partial^2 \mathcal{W}(A_i)}{\partial A_j \partial A_k}\psi_j\psi_k + \text{h.c.}\right]. \tag{B.76}$$

Again we detect in the third term

$$V = \left|\frac{\partial \mathcal{W}}{\partial A_k}\right|^2 = \sum_k |(c_k + m_{kj}A_j + g_{kji}A_j A_i)|^2$$

the scalar masses and their self-interaction, and in the last term

$$\left[\frac{\partial^2 \mathcal{W}}{\partial A_j \partial A_k}\psi_j\psi_k + \text{h.c.}\right] = (m_{jk} + 2g_{jki}A_i)\psi_j\psi_k$$

we see the fermion masses and three-point couplings between two fermions and a scalar. With regard to the Standard Model this "smells" like the Yukawa interaction between a Higgs boson and two fermions.

Vectorfield Lagrangians

In the previous subsection, the most general supersymmetric Lagrangian containing scalar and fermionic fields was constructed. As an intermediate step towards supersymmetric Yang-Mills theories, the next task is the construction of terms containing the typical field strengths for vector fields. As will be seen these can be derived from the expressions

$$W_\alpha := -\frac{1}{4}\bar{D}^2 D_\alpha V \qquad \overline{W}^{\dot\alpha} := +\frac{1}{4}D^2\bar{D}^{\dot\alpha}V, \tag{B.77}$$

where V is a vector superfield (and $D^2 := D^\alpha D_\alpha$, $\bar{D}^2 := \bar{D}_{\dot\alpha}\bar{D}^{\dot\alpha}$). From their algebra (B.56) the derivatives can be shown to obey $D_\alpha D^2 \equiv 0 \equiv \bar{D}^{\dot\alpha}\bar{D}^2$. This in turn entails that the fields W_α and $\bar{W}^{\dot\alpha}$ are chiral and antichiral, respectively. Under the SuSy gauge transformations (B.64), W_α transforms into

$$W'_\alpha = -\frac{1}{4}\bar{D}^2 D_\alpha V - \frac{1}{4}\bar{D}^2 D_\alpha \Phi - \frac{1}{4}\bar{D}^2 D_\alpha \Phi^\dagger.$$

The last term vanishes since $\Phi^\dagger$ is anti-chiral. But the second term also vanishes because of

$$\bar{D}_{\dot\alpha}\bar{D}^{\dot\alpha}D_\alpha\Phi = \bar{D}_{\dot\alpha}\{\bar{D}^{\dot\alpha}, D_\alpha\}\Phi = -2i\epsilon^{\dot\alpha\dot\beta}(\sigma^\mu)_{\alpha\dot\beta}\bar{D}_{\dot\alpha}\partial_\mu\Phi = 0.$$

Thus W_α is SuSy gauge invariant ($W'_\alpha = W_\alpha$). The same can be proven for $\overline{W}_{\dot\alpha}$. The derivation of the explicit expressions for these objects in terms of ordinary fields turns out to be rather elaborate (see e.g. [545]). In the Wess-Zumino gauge, the result is

$$W_\alpha = -i\lambda_\alpha(y) + D(y)\theta_\alpha - i(\sigma^{\mu\nu}\theta)_\alpha F_{\mu\nu}(y) + \theta^2(\sigma^\mu\partial_\mu\bar{\lambda}(y))_\alpha \tag{B.78a}$$

$$\overline{W}^{\dot\alpha} = i\bar{\lambda}^{\dot\alpha}(y^*) + D(y^*)\bar{\theta}^{\dot\alpha} + i(\bar{\sigma}^{\mu\nu}\bar{\theta})^{\dot\alpha}F_{\mu\nu}(y^*) - i\bar{\theta}^2(\bar{\sigma}^\mu\partial_\mu\bar{\lambda}(y^*))^{\dot\alpha}. \tag{B.78b}$$

The vector field v_μ only appears in the form $F_{\mu\nu} := \partial_\mu v_\nu - \partial_\nu v_\mu$. This justifies calling the W fields supersymmetric field strengths.

Quite in analogy with "ordinary" gauge fields, it makes sense to build from the W-fields quadratic expressions $W^\alpha W_\alpha$ and $\overline{W}_{\dot\alpha}\overline{W}^{\dot\alpha}$. These are chiral or anti-chiral superfields, respectively, whose F-terms are suited to define a kinetic Lagrangian

$$\mathcal{L}_k := \frac{1}{4}(W^\alpha W_\alpha|_F + \bar{W}_{\dot\alpha}\bar{W}^{\dot\alpha}|_F) = \frac{1}{2}D^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - i\lambda\sigma^\mu\partial_\mu\bar{\lambda}. \tag{B.79}$$

This Lagrangian component reproduces the kinetic term for an Abelian gauge theory (the photon v_μ) together with a "photino" λ. Since there is no kinetic term for D, this auxiliary superfield can be removed from the theory by its field equation.

Local Gauge Symmetry by Minimal Coupling

Let us next mimic the procedure of introducing a locally gauge-invariant interaction term amongst the photon and the Fermi fields via the introduction of covariant derivatives. A global $\mathbf{U(1)}$-transformation of a chiral superfield is given by

$$\Phi'_k = e^{-iq_k\Lambda}\Phi_k,$$

where q_k is the $\mathbf{U(1)}$-charge of Φ_k. The Fermi fields are components in the Lagrangian (B.74). The kinetic terms $\Phi_k^\dagger\Phi_k|_D$ are globally gauge invariant. However, the part in (B.74) with the superpotential $\mathcal{W}$ is not invariant in general. Its gauge invariance depends on the coefficients in (B.69): For $\mathcal{W}$ to be $\mathbf{U(1)}$-invariant, we must demand

$$c_i = 0 \quad \text{if} \quad q_i \neq 0 \tag{B.80a}$$
$$m_{ij} = 0 \quad \text{if} \quad q_i + q_j \neq 0 \tag{B.80b}$$
$$g_{ijk} = 0 \quad \text{if} \quad q_i + q_j + q_k \neq 0. \tag{B.80c}$$

In turning the global invariance into a local invariance by allowing the gauge parameter Λ to become spacetime dependent, we first observe that the transformed chiral superfields fields Φ'_k are chiral only if Λ is chiral: $\bar{D}^{\dot\alpha}\Lambda = 0 = D_\alpha\Lambda^\dagger$. The kinetic energy term transforms as

$$\Phi'^\dagger_k\Phi'_k = \Phi_k^\dagger\Phi_k e^{iq_k(\Lambda^\dagger-\Lambda)} \qquad \text{(no sum with respect to k)}.$$

The extra factor may be compensated by introducing the vector superfield V with a transformation behavior according to (B.64): $V' = V + i(\Lambda^\dagger - \Lambda)$. Then the Lagrangian

$$\mathcal{L}_G = \frac{1}{4}(W^\alpha W_\alpha|_F + \bar{W}_{\dot\alpha}\bar{W}^{\dot\alpha}|_F) + \Phi_k^\dagger e^{q_k V}\Phi_k|_D + \left[\mathcal{W}(\Phi_i)\big|_F + h.c.\right] \tag{B.81}$$

is invariant under local $\mathbf{U(1)}$ gauge transformations–provided of course that (B.80) is fulfilled. The first term contains the gauge-invariant kinetic energy expression for the $\mathbf{U(1)}$-gauge fields. The second term contains the gauge-invariant kinetic energy expression for the matter fields. A somewhat tedious calculation yields (in the WZ gauge)

$$\begin{aligned}\Phi^\dagger e^{gV}\Phi\big|_D = |F|^2 + (D_\mu A)^* D^\mu A - i\bar\psi\sigma^\mu D_\mu\psi + q|A|^2 D \\ + iq\sqrt{2}(A^*\lambda\psi - \bar\lambda\bar\psi A)\end{aligned} \tag{B.82}$$

with the covariant derivative $D_\mu := \partial_\mu + igv_\mu$. (Just for clarification: The F in this expression is the field strength with respect to the vector field v_μ–and not the auxiliary field. The field A is a scalar field as it arises from the Fourier expansion of a chiral superfield according to (B.58).)

This shows that there is a fermion field λ acting as the supersymmetric partner of the gauge boson v_μ. Therefore it is justified to call λ the "photino"–or more general–"gaugino". Both v_μ and λ transform in the same way under the gauge group. And both the photon and the photino fields are massless in this theory. Furthermore, aside of the (minimal) couplings of the gauge field v_μ to the fermionic matter field ψ–as familiar from gauge invariance–there is an additional interaction between ψ, the scalar field A, and the gaugino λ. The strength of this interaction is also given by the gauge coupling g.

Supersymmetric Yang-Mills Lagrangian

The previous considerations can be generalized to non-Abelian groups; see the original articles by S. Ferrara and B. Zumino [177] and by A. Salam and J. Strathdee [458]. This generalization refers to the vector superfield which now becomes Lie-algebra valued

$$V = V^a T^a,$$

and to W-fields generalized to

$$W_\alpha := -\frac{1}{4\,g}\bar{D}^2 e^{-gV} D_\alpha e^{gV} \qquad \overline{W}^{\dot\alpha} = \frac{1}{4\,g} D^2 e^{gV} \bar{D}^{\dot\alpha} e^{-gV}.$$

Under the transformations $e^V \to e^{i\Lambda^\dagger} e^V e^{-i\Lambda}$, the superfields transform covariantly as

$$W_\alpha \to e^{i\Lambda} W_\alpha e^{-i\Lambda} \qquad\qquad \bar{W}^{\dot\alpha} \to e^{i\Lambda^\dagger} \overline{W}^{\dot\alpha} e^{-i\Lambda^\dagger}.$$

In expanding e^V in the Wess-Zumino gauge according to (B.66), one obtains after some moderate calculation the generalization of (B.78)

$$W_\alpha = -i\lambda_\alpha(y) + D(y)\theta_\alpha - i(\sigma^{\mu\nu}\theta)_\alpha F_{\mu\nu}(y) + \theta^2(\sigma^\mu D_\mu \bar\lambda(y))_\alpha$$

where now the field strength and the covariant derivative take the well-known form

$$F_{\mu\nu} = \partial_\mu v_\nu - \partial_\nu v_\mu - ig\left[v_\mu, v_\nu\right] \qquad D_\mu\bar\lambda = \partial_\mu\bar\lambda - ig\left[v_\mu, \bar\lambda\right].$$

The supersymmetric free Yang-Mills action is

$$S_{\text{S-YM}} = -\frac{1}{4}\int d^4x d^2\theta \;\mathrm{Tr}\; W^\alpha W_\alpha = \int d^4x \;\mathrm{Tr}\left[-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - i\lambda\sigma^\mu D_\mu\bar\lambda + \frac{1}{2}D^2\right].$$

In addition to the familiar kinetic term for the gauge fields, this expression contains a kinetic energy term for the gauginos, as well as the coupling of the gauginos to the gauge fields–its strength dictated by the group structure functions. There is no kinetic term for the D-field, which can be eliminated as an auxiliary field.

A (chiral) matter field Φ can be minimally coupled to the Yang-Mills field by putting it into the adjoint representation of the gauge group, and replacing (B.70) by

$$S_{\text{matter}} = \int d^8z \;\operatorname{Tr}\; \Phi^\dagger e^{-gV}\Phi + \int d^6z \;\mathcal{W}(\Phi_i) + \int d^6\bar{z}\; \mathcal{W}^\dagger(\Phi_i).$$

B.4 *Supergroups

The classification of "ordinary" Lie groups can be extended to graded Lie groups ('supergroups'). Similar to "ordinary" Lie groups one finds families of simple supergroups. There are two main series, which are called orthosymplectic $\mathbf{OSp(N/M)}$ and super-unitary $\mathbf{SU(N/M)}$. Additionally there are exceptional simple groups. The following is only a sketch of the representation theory of supergroups; for details see e.g. Section 3.1 in [519] or [196].

B.4.1 OSp(N/M) and the Super-Poincaré Algebra

The graded Lie group $\mathbf{OSp(N/M)}$ can be understood from its building blocks $\mathbf{O(N)}$ and $\mathbf{Sp(N)}$. As explained in Appendix A, the group $\mathbf{O(N)}$ is represented by the real orthogonal $N \times N$ matrices that leave invariant the expression $\sum^{N} x_i x_i$, and the symplectic group $\mathbf{Sp(M)}$ is represented by those $M \times M$ matrices that leave $\sum^{M} \theta_a C_{ab} \theta_b$ invariant (where C_{ab} is a real antisymmetric matrix and the θ_a are anticommuting). The orthosymplectic group $\mathbf{OSp(N/M)}$ is the set of matrices that leaves invariant the "line element"

$$\sum^{N} x_i x_i \;+\; \sum^{M} \theta_a C_{ab} \theta_b.$$

The elements of the orthosymplectic group can be organized into block form:

$$\begin{pmatrix} O(N) & A \\ B & Sp(M) \end{pmatrix}.$$

Here the entries on the diagonal are bosonic, whereas A and B are fermionic, but of no further interest here. We have

$$\mathbf{O(N)} \otimes \mathbf{Sp(M)} \subset \mathbf{OSp(N/M)}.$$

In supergravity one is interested in the orthosymplectic groups $\mathbf{OSp(N/4)}$. Its building block $\mathbf{Sp(4)}$ is isomorphic to $\mathbf{O(3,2)}$ (see e.g. [519]). By a group contraction, the $\mathbf{OSp(N/4)}$ becomes the Poincaré group. The $\mathbf{O(N)}$ -part describes the extensions from $N = 1$ to $N = 8$.

B.4.2 SU(N/M) and the Super-Conformal Algebra

The matrices of $\mathbf{SU(N/M)}$ leave invariant the expression

$$\sum^{N}(z_i)^* z_i \quad + \quad \sum^{M}(\theta_a)^* \delta_{ab}\theta_b \qquad \text{with} \qquad \delta_{ab} = \pm\delta_{ab}.$$

It holds that

$$\mathbf{SU(N)} \otimes \mathbf{SU(M)} \otimes \mathbf{U(1)} \subset \mathbf{SU(N/M)}.$$

If one takes for the first factor the special unitary group $\mathbf{SU(2,2)}$ the supergroup $\mathbf{SU(2,2/1)}$ has–after group contraction–the algebra of super-conformal symmetry. The expression left invariant is

$$|z_1|^2 + |z_2|^2 - |z_3|^2 - |z_4|^2 + \theta_1^*\theta^1 + \theta_2^*\theta^2 - \theta_3^*\theta^3 - \theta_4^*\theta^4.$$

This super-conformal group–also used as a gauge group of supergravity–has in its algebra aside from the 15 generators $T_\mu, M_{\mu\nu}, C_\mu, S$ of the conformal group (see Sect. 3.5.1) and aside from the fermionic generators Q^a, further spinorial elements and–for closure of the algebra–also another bosonic element; the full algebra and two feasible representations of the generators can be found for instance in Sect. 14.5 of [369].

Appendix C
Symmetries and Constrained Dynamics

Symmetries make Lagrangians look compact and aesthetic, but–as will be seen in this section–they render Hamiltonians ugly. This is because the standard procedure–the Legendre transformation–of going from the configuration-velocity space to the phase space is obstructed. As pointed out in Subsect. 3.3.3, the Noether identities imply in the case of a local symmetry that the canonically conjugate momenta cease to be independent: They are constrained; see (3.81). Thus fundamental physics, abundant in local symmetries, is full of constraints. This point becomes virulent if one wants to step from the classical theory to its quantized version in a canonical way.

This appendix has three sections. In the first one, the basic terminology is exposed, more detailed for systems with finitely many degrees of freedom, and a little sketchy for field theories. The second section treats Yang-Mills type theories including good-old electrodynamics, and the last one deals with re-parametrization invariant–or–generally covariant theories, in this case the relativistic point particle and general relativity. The plain existence of two separate sections devoted to those theories describing particle fields on the one hand side and gravity on the other hand may be seen as an indication that the unification of both worlds is still not reached, despite some (or, if you like, many) formal similarities in their Hamiltonian formulation.

Most of the material in this appendix is treated in more detail in [491]. The core techniques for constrained dynamics are dealt with in [563]. The original ideas are described in [129]. More advanced techniques, especially if it comes to the quantization of constrained systems by BRST-BFV-techniques, can be found in [218], [273].

C.1 Constrained Dynamics

C.1.1 Singular Lagrangians

Assume a classical theory with a finite number of degrees of freedom q^k $(k = 1, \ldots, N)$ defined by its Lagrange function $L(q, \dot{q})$ with the equations of motion

$$\begin{aligned}[L]_k := \frac{\partial L}{\partial q^k} - \frac{d}{dt}\,\frac{\partial L}{\partial \dot{q}^k} &= \left(\frac{\partial L}{\partial q^k} - \frac{\partial^2 L}{\partial \dot{q}^k \partial q^j}\dot{q}^j\right) - \frac{\partial^2 L}{\partial \dot{q}^k \partial \dot{q}^j}\ddot{q}^j \\ &=: V_k - W_{kj}\ddot{q}^j = 0. \end{aligned} \tag{C.1}$$

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

For simplicity, it is assumed that the Lagrange function does not depend on time explicitly; all the following results can readily be extended. As demonstrated in the main text, a crucial role is played by the matrix (sometimes called the "Hessian")

$$W_{kj} \ := \ \frac{\partial^2 L}{\partial \dot{q}^k \partial \dot{q}^j}. \tag{C.2}$$

If $\det W = 0$, not only the Lagrangian but the system itself is termed 'singular', and 'regular' otherwise. This appendix deals exclusively with singular systems.

From the definition of momenta by

$$p_k(q,\dot{q}) = \frac{\partial L}{\partial \dot{q}^k}, \tag{C.3}$$

we immediately observe that only in the regular case can the $p_k(q,\dot{q})$ be solved for all the velocities in the form $\dot{q}^j(q, p)$–at least locally.

In the singular case, $\det W = 0$ implies that the $N \times N$ matrix W has a rank R smaller than N - or that there are $P = N\text{-}R$ null eigenvectors ξ^k_ρ:

$$\xi^k_\rho \ W_{kj} \equiv 0 \qquad \text{for} \qquad \rho = 1, \ldots, P \ (= N - R). \tag{C.4}$$

This rank is independent of which generalized coordinates are chosen for the Lagrange function. The null eigenvectors serve to identify those of the equations in (C.1) which are not of second order. By contracting these with $\ddot{q}^j$ we get the P on-shell equations

$$\chi_\rho = \xi^k_\rho \ V_k(q,\dot{q}) = 0.$$

In analogy with the terminology introduced in the next subsection, these may be called 'primary Lagrangian constraints'. Being functions of $(q,\dot{q})$ they are not genuine equations of motion but–if not fulfilled identically–they restrict the dynamics to a subspace within the configuration-velocity space (or in geometrical terms, the tangent bundle $T\mathbb{Q}$). For reasons of consistency, the time derivative of these constraints must not lead outside this subspace. This condition possibly enforces further Lagrangian constraints and by this a smaller subspace of allowed dynamics, etc. Thus eventually one derives a chain of Lagrangian constraints that define the "true" subspace in which a consistent dynamics occurs. I shall not go into further details here (see for instance [353], [490], [491]), but rather leave the Lagrangian description in favor of the Hamiltonian treatment.

The previous considerations are carried over to a field theory with a generic Lagrangian density $\mathcal{L}(Q^\alpha, \partial_\mu Q^\alpha)$. Rewrite the field equations similar to (C.1)

$$\begin{aligned}[\mathcal{L}]_\alpha &:= \frac{\partial \mathcal{L}}{\partial Q^\alpha} - \partial_\mu \frac{\partial \mathcal{L}}{\partial Q^\alpha_{,\mu}} = \left(\frac{\partial \mathcal{L}}{\partial Q^\alpha} - \frac{\partial^2 \mathcal{L}}{\partial Q^\alpha_{,\mu} \partial Q^\beta} Q^\beta_{,\mu}\right) - \frac{\partial^2 \mathcal{L}}{\partial Q^\alpha_{,\mu} \partial Q^\beta_{,\nu}} \ \ Q^\beta_{,\mu\nu} \\ &:= \mathcal{V}_\alpha - \mathcal{W}^{\mu\nu}_{\alpha\beta} \ Q^\beta_{,\mu\nu}. \end{aligned} \tag{C.5}$$

Depending now on the choice of a time variable T (e.g. $T = x^0,\ T = x^0 - x^{D-1}, \ldots$) the Hessian is defined by

$$\overset{T}{\mathcal{W}}_{\alpha\beta} := \frac{\partial^2 \mathcal{L}}{\partial(\partial_T Q^\alpha)\partial(\partial_T Q^\beta)}. \tag{C.6}$$

If the rank of this matrix is $R < N$, it has $P = N\text{-}R$ null eigenvectors

$$\overset{T}{\xi}{}^{\alpha}{}_{\rho} \overset{T}{\mathcal{W}}_{\alpha\beta} \equiv 0. \tag{C.7}$$

Observe that the Hessian (and by this also the zero eigenvectors) depend on the selection of the time variable. This indeed has the direct consequence that the number of zero eigenvectors and of (primary) constraints depends on this choice.

C.1.2 Constraints as a Consequence of Local Symmetries

In investigating the consequences of Noether's second theorem, we found that the momenta are not independent but obey constraints; see (3.81). Here the argumentation will be slightly modified: The Noether identities for local symmetries of the form (3.68) are according to (3.75)

$$[\mathcal{L}]_\alpha (\mathcal{A}_r^\alpha - Q^\alpha_{,\mu} \mathcal{D}_r^\mu) - \partial_\mu ([\mathcal{L}]_\alpha \mathcal{B}_r^{\alpha\mu}) = 0.$$

Let us identify the terms with the highest possible derivatives of the fields Q^α by using the expression (C.5) which already isolates the second derivatives. A further derivative possibly originates from the last term in the Noether identity. It reads $\mathcal{B}_r^{\alpha\mu} \mathcal{W}^{\lambda\nu}_{\alpha\beta}\, Q^\beta_{,\lambda\nu\mu}$. This term must vanish itself for all third derivatives of the fields, and therefore

$$\mathcal{B}_r^{\alpha(\mu} \mathcal{W}^{\lambda\nu)}_{\alpha\beta} = 0,$$

where the symmetrization goes over μ, λ, ν. Among these identities is the Hessian, and we find

$$\overset{T}{\mathcal{B}}{}^{\alpha}{}_{r}\ \overset{T}{\mathcal{W}}_{\alpha\beta} = 0. \tag{C.8}$$

Thus the non-vanishing $\overset{T}{\mathcal{B}}_r$ are null-eigenvectors of the Hessian. And comparing this with (C.7) there must be linear relationships

$$\overset{T}{\mathcal{B}}{}^{\alpha}{}_{r} = \lambda_r^\rho\ \overset{T}{\xi}{}^{\alpha}{}_{\rho}$$

with coefficients λ_r^ρ. In case all or some of the $\overset{T}{\mathcal{B}}{}^{\alpha}{}_{r}$ are zero, we can repeat the previous argumentation by singling out the terms with second derivatives, and find again that the Hessian has a vanishing determinant.

We conclude that every action which is invariant under local symmetry transformations necessarily describes a singular system. This, however, should not lead to the impression that any singular system exhibits local symmetries. As a matter of fact, a system can become singular just by the choice of the time variable.

C.1.3 Rosenfeld-Dirac-Bergmann Algorithm

Since the fields and the canonical momenta are not independent, they cannot be taken as coordinates in a phase space as one is accustomed in the unconstrained case. This difficulty was known already by the end of the 1920's, and after unsatisfactory attempts by eminent physicists such as W. Pauli, W. Heisenberg, and E. Fermi this problem was attacked by L. Rosenfeld. He undertook the very ambitious effort of obtaining the Hamiltonian version for the Einstein-Maxwell theory as a preliminary step towards quantization [449]. But only in the late forties and early fifties did the Hamiltonian version of constrained dynamics acquire a substantially mature form due to P.A.M. Dirac on the one hand and to P. Bergmann and collaborators on the other hand; for historical aspects, see [460], [461].

In order to get an easier access on the notions of constrained Hamiltonian dynamics, let us restrict ourselves again to a system with finitely many degrees of freedom.

Primary Constraints

The rank of the Hessian (C.2) being $R = N - P$ implies that–at least locally–the Eqs. (C.3) can be solved for R of the velocities in terms of the positions, some of the momenta and the remaining velocities. Furthermore, there are P relations

$$\phi_\rho(q, p) = 0 \qquad \rho = 1, \ldots, P \tag{C.9}$$

which restrict the dynamics to a subspace $\Gamma_P \subset \Gamma$ of the full phase space Γ. These relations were dubbed *primary constraints* by Anderson and Bergmann [9], a term suggesting that there are possibly secondary and further generations of constraints.

For many of the calculations below, one needs to set regularity conditions, namely (1) the Hessian of the Lagrangian has constant rank, (2) there are no ineffective constraints, that is constraints whose gradients vanish on Γ_P, (3) the rank of the Poisson bracket matrix of constraints remains constant in the stabilization (RDB) algorithm described below.

Weak and Strong Equations

It will turn out that even in the singular case, one can write the dynamical equations in terms of Poisson brackets. But one must be careful in interpreting them in the presence of constraints. In order to support this precaution, Dirac [127] contrived the concepts of "weak" and "strong" equality.

If a function $F(p, q)$ which is defined in the neighborhood of Γ_P becomes identically zero when restricted to Γ_P it is called "weakly zero", denoted by $F \approx 0$:

$$F(q, p)\big|_{\Gamma_P} = 0 \longleftrightarrow F \approx 0.$$

(Since in the course of the algorithm the constraint surface is possibly narrowed down, a better notation would be $F \approx_{|\Gamma_P} 0$.) If the gradient of F is also identically zero on Γ_P, F is called "strongly zero", denoted by $F \eqsim 0$:

$$F(q, p)\big|_{\Gamma_P} = 0 \qquad \Big(\frac{\partial F}{\partial q^i}, \frac{\partial F}{\partial p_k}\Big)\big|_{\Gamma_P} = 0 \qquad \longleftrightarrow \qquad F \eqsim 0.$$

It can be shown (see e.g. [491]) that if F vanishes weakly, it strongly is a linear combination of the functions defining the constraint surface–or -

$$F \approx 0 \longleftrightarrow F - f^\rho \phi_\rho \eqsim 0.$$

Indeed, the subspace Γ_P can itself be defined by the weak equations

$$\phi_\rho \approx 0, \tag{C.10}$$

since due to the regularity assumptions, the constraints do not vanish strongly.

Canonical and Total Hamiltonian

Next we introduce the "canonical" Hamiltonian

$$H_C = \; p_i \dot{q}^i - L(q, \dot{q}).$$

Its variation yields

$$\delta H_C = \; (\delta p_i)\dot{q}^i + \; p_i \delta \dot{q}^i - \frac{\partial L}{\partial q^i}\delta q^i - \frac{\partial L}{\partial \dot{q}^i}\delta \dot{q}^i = \dot{q}^i \; \delta p_i - \frac{\partial L}{\partial q^i} \; \delta q^i$$

(after using the definition of momenta), revealing the remarkable fact that the canonical Hamiltonian can be written in terms of q's and p's. No explicit dependence on any velocity variable is left, despite the fact that the Legendre transformation is non-invertible[1]. Observe, however, that the expression for δH_C given in terms of

[1] In geometrical terms, this is explained by the Legendre projectability of the Lagrangian energy function; see Section C.5.1.

the variations δq^i and δp_i does not allow the derivation of the Hamilton equations of motion, since the variations are not independent due to the primary constraints (C.10). In order that these be respected, the variation of H_C needs to be performed together with Lagrange multipliers. Thus we introduce the "total" Hamiltonian

$$H_T = H_C + u^\rho \phi_\rho \tag{C.11}$$

with arbitrary multiplier functions u^ρ in front of the primary constraint functions. Varying the total Hamiltonian with respect to (u, q, p) we obtain the constraints (C.10) and

$$\frac{\partial H_C}{\partial p_i} + u^\rho \frac{\partial \phi_\rho}{\partial p_i} = \dot{q}^i \tag{C.12a}$$

$$\frac{\partial H_C}{\partial q^i} + u^\rho \frac{\partial \phi_\rho}{\partial q^i} = -\frac{\partial L}{\partial q^i} = -p_i \tag{C.12b}$$

where the last relation follows from the definition of momenta and the Euler-Lagrange equations. This recipe of treating the primary constraints with Lagrange multipliers sounds reasonable and may work, but it leaves one unsure about a mathematical justification. This justification must guarantee that the Hamiltonian description of the dynamics is ultimately equivalent to the Lagrangian formalism; a proof is given in [33]. The Eqs. (C.12) are reminiscent of the Hamilton equations for regular systems. However, there are extra terms depending on the primary constraints and the multipliers. Nevertheless, (C.12) can be written in terms of Poisson brackets, provided one adopts the following convention: Consider

$$\{F, H_T\} = \{F, H_C + u^\rho \phi_\rho\} = \{F, H_C\} + u^\rho \{F, \phi_\rho\} + \{F, u^\rho\}\phi_\rho.$$

Since the multipliers u_ρ are not phase-space functions, the Poisson brackets $\{F, u^\rho\}$ are not defined. However, these appear multiplied with constraints and thus the last term vanishes weakly. Therefore the dynamical equations for any phase-space function $F(q, p)$ can be written as

$$\dot{F}(p, q) \approx \{F, H_T\}. \tag{C.13}$$

Before continuing to scrutinize how many of the functions u^ρ can be determined, and before we discover that the undetermined ones are connected to local symmetries of the theory in question, let us relate the primary constraints to the eigenvectors of the Hessian. The primary constraints can be thought of as functions of the velocities: $\phi_\rho = \phi_\rho(q, \partial L/\partial \dot{q})$. Then

$$0 \approx \frac{\partial \phi_\rho}{\partial \dot{q}^k} = \frac{\partial \phi_\rho}{\partial p_i}\frac{\partial p_i}{\partial \dot{q}^k} = \frac{\partial \phi_\rho}{\partial p_i} W_{ik}.$$

Thus $\partial \phi_\rho / \partial p_i$ is a null eigenvector of the Hessian and can be expressed as a linear combination of the null eigenvectors from (C.4). They even may serve to canonize the null eigenvectors:

$$\xi^i_\rho \equiv \frac{\partial \phi_\rho}{\partial p_i}. \tag{C.14}$$

A more precise definition in terms of Legendre projectability is given in Sect. C.5.

Stability of Constraints

For consistency of a theory, one must require that the primary constraints are conserved during the dynamical evolution of the system:

$$0 \overset{!}{\approx} \dot{\phi}_\rho \approx \{\phi_\rho, H_C\} + u^\sigma\{\phi_\rho, \phi_\sigma\} := h_\rho + C_{\rho\sigma}u^\sigma. \tag{C.15}$$

There are essentially two distinct situations, depending on whether the determinant of $C_{\rho\sigma}$ vanishes (weakly) or not

- $\det C \neq 0$: In this case (C.15) constitutes an inhomogeneous system of linear equations with solutions
$$u^\rho \approx -\bar{C}^{\rho\sigma} h_\sigma$$
where $\overline{C}$ is the inverse of the matrix C. Therefore, the Hamilton equations of motion (C.13) become
$$\dot{F} \approx \{F, H_C\} - \{F, \phi_\rho\}\bar{C}^{\rho\sigma}\{\phi_\sigma, H_C\},$$
which are free of any arbitrary multipliers.
- $\det C \approx 0$: In this case, the multipliers are not uniquely determined and (C.15) is only solvable if the h_ρ fulfill certain relations, derived as follows: Let the rank of C be M. This implies that there are $(P\text{-}M)$ linearly-independent null eigenvectors, i.e. $w^\rho_\alpha C_{\rho\sigma} \approx 0$ from which by (C.15) we find the conditions
$$0 \overset{!}{\approx} w^\rho_\alpha h_\rho.$$
These either are fulfilled or lead to a certain number S' of new constraints
$$\phi_{\bar{\rho}} \approx 0 \qquad \bar{\rho} = P+1, \ldots, P+S'$$
called "secondary" constraints according to Anderson and Bergmann [9]. The primary and secondary constraints define a hypersurface $\Gamma_2 \subseteq \Gamma_P$. In a further step–for consistency again–one has to check that the original and the newly generated constraints are conserved on Γ_2. This might imply another generation of constraints, called tertiary constraints, defining a hypersurface $\Gamma_3 \subseteq \Gamma_2$, etc., etc. It seems harder to formulate the algorithm in all its details (these may be found in [491]) than to apply it to the field theories of interest. In most physically relevant cases, the algorithm terminates with the secondary constraints, and there are only few cases where also at most tertiary constraints appear.[2]

The algorithm terminates when the following situation is attained: There is a hypersurface Γ_C defined by the constraints

$$\phi_\rho \approx_{|\Gamma_C} 0 \quad \rho = 1, \ldots, P \quad \text{and} \quad \phi_{\bar{\rho}} \approx_{|\Gamma_C} 0 \quad \bar{\rho} = P+1, \ldots, P+S. \tag{C.16}$$

[2] I thank J.B. Pitts for pointing out to me that pure spin-2 massive gravity has a quaternary constraint.

The first set $\{\phi_\rho\}$ contains all P primary constraints, the other set $\{\phi_{\bar{\rho}}\}$ comprises all secondary, tertiary, etc. constraints, assuming there are S of them. It turns out to be convenient to use a common notation for all constraints as $\phi_{\hat{\rho}}$ with $\hat{\rho} = 1, \ldots, P+S$. Furthermore, for every left null-eigenvector $w_\alpha^{\hat{\rho}}$ of the matrix

$$\hat{C}_{\hat{\rho}\rho} = \{\phi_{\hat{\rho}}, \phi_\rho\} \tag{C.17}$$

the conditions $w_\alpha^{\hat{\rho}}\{\phi_{\hat{\rho}}, H_C\} \approx_{|\Gamma_C} 0$ are fulfilled. For the multiplier functions u^ρ, the equations

$$\{\phi_{\bar{\rho}}, H_C\} + \{\phi_{\bar{\rho}}, \phi_\rho\}u^\rho \approx_{|\Gamma_C} 0, \tag{C.18}$$

hold. In the following, weak equality $\approx$ is always understood with respect to the "final" constraint hypersurface Γ_C. This hypersurface is also termed 'reduced phase-space'. Its description in local coordinates and its symplectic structure are far from obvious.

First- and Second-Class Constraints

We are still curious about the fate of the multiplier functions. Our curiosity leads to the notion of first- and second-class objects.

Some of the Eqs. (C.18) may be fulfilled identically, others represent conditions on the u^ρ. The details depend on the rank of the matrix $\hat{C}$ defined by (C.17). If the rank of $\hat{C}$ is P, all multipliers are fixed. If the rank of $\hat{C}$ is $K < P$ there are P-K solutions of

$$\hat{C}_{\hat{\rho}\rho}V_\alpha^\rho = \{\phi_{\hat{\rho}}, \phi_\rho\}V_\alpha^\rho \approx 0. \tag{C.19}$$

The most general solution of the linear inhomogeneous Eqs. (C.18) is the sum of a particular solution U^ρ and a linear combination of the solutions of the homogeneous part:

$$u^\rho = U^\rho + v^\alpha V_\alpha^\rho \tag{C.20}$$

with arbitrary coefficients v^α. Together with ϕ_ρ, also the linear combinations

$$\phi_\alpha := V_\alpha^\rho \phi_\rho \tag{C.21}$$

constitute constraint functions. According to (C.19), these have the property that their Poisson brackets with all constraints vanish on the constraint surface.

A phase-space function $\mathcal{F}(p, q)$ is said to be *first class* (FC) if it has a weakly vanishing Poisson bracket with all constraints in the theory:

$$\{\mathcal{F}(p, q), \phi_{\hat{\rho}}\} \approx 0.$$

If a phase-space object is not first class, it is called *second class* (SC). Due to the definitions of weak and strong equality a first-class quantity obeys the strong equation

$$\{\mathcal{F}, \phi_{\hat{\rho}}\} \simeq f^{\hat{\sigma}}_{\hat{\rho}}\ \phi_{\hat{\sigma}},$$

from which by virtue of the Jacobi identity we infer that the Poisson bracket of two FC objects is itself an FC object:

$$\begin{aligned}\{\{\mathcal{F}, \mathcal{G}\}, \phi_{\hat{\rho}}\} &= \{\{\mathcal{F}, \phi_{\hat{\rho}}\}, \mathcal{G}\} - \{\{\mathcal{G}, \phi_{\hat{\rho}}\}, \mathcal{F}\} \\ &= f^{\hat{\sigma}}_{\hat{\rho}}\{\phi_{\hat{\sigma}}, \mathcal{G}\} + \{f^{\hat{\sigma}}_{\hat{\rho}}, \mathcal{G}\}\phi_{\hat{\sigma}} - g^{\hat{\sigma}}_{\hat{\rho}}\{\phi_{\hat{\sigma}}, \mathcal{F}\} - \{g^{\hat{\sigma}}_{\hat{\rho}}, \mathcal{F}\}\phi_{\hat{\sigma}} \approx 0.\end{aligned}$$

It turns out to be advantageous to reformulate the theory completely in terms of its maximal number of independent FC constraints and the remaining SC constraints. Assume that this maximal number is found after building suitable linear combinations of constraints. Call this set of FC constraints Φ_I ($I = 1, \ldots, L$) and denote the remaining second class constraints by χ_A. Evidence that one has found the maximal number of FC constraints is the non-vanishing determinant of the matrix built by the Poisson brackets of all second class constraints

$$(\Delta_{AB}) = \{\chi_A, \chi_B\}. \tag{C.22}$$

If this matrix were singular, then there would exist a null vector $0 \approx e^A \Delta_{AB} \approx \{e^A \chi_A, \chi_B\}$, meaning that the constraint $e^A \chi_A$ commutes with all SC constraints. Since by the definition of first class the Poisson brackets with all FC constraints also vanish, $e^A \chi_A$ would be a first-class constraint itself, contradicting the assumption that all independent first-class constraints were found already. As a corollary we find that the number of SC constraints must be even[3], because otherwise $\det \Delta = 0$.

Rewriting the total Hamiltonian (C.11) with the aid of (C.20) as

$$H_T = H' + v^\alpha \phi_\alpha \qquad \text{with} \qquad H' = H_C + U^\rho \phi_\rho, \tag{C.23}$$

we observe that H' is itself first class, and that the total Hamiltonian is a sum of a first-class Hamiltonian and a linear combination of primary first-class constraints (PFC).

Consider again the system of Eqs. (C.18). They are identically fulfilled for the FC constraints. For a SC constraint, we can write these equations as

$$\{\chi_A, H_C\} + \Delta_{AB}\ u^B \approx 0$$

with the understanding that $u^B = 0$ if χ^B is a secondary constraint (SC). For the other multipliers we obtain

$$u^B = \overline{\Delta}^{BA}\ \{\chi_A, H_c\} \qquad \text{for} \qquad \chi_B \quad \text{primary}. \tag{C.24}$$

---

[3] This holds true only for finite-dimensional (and bosonic) systems.

where $\overline{\Delta}$ is the inverse of Δ. As a result, we find that all multipliers belonging to the primary second-class constraints in H' of (C.23) are determined, and that only the v^α are left open: There are as many arbitrary functions in the Hamiltonian as there are (independent) primary first-class constraints (PFC).

Inserting the solutions (C.24) into the Hamilton Eqs. (C.13), they become

$$\dot{F}(p, q) \approx \{F, H_T\} \approx \{F, H_C\} + \{F, \phi_\alpha\}v^\alpha - \{F, \chi_A\}\overline{\Delta}^{AB}\{\chi_B, H_C\}. \quad \text{(C.25)}$$

C.1.4 First-Class Constraints and Symmetries

"First-Class Constraints are Gauge Generators": Maybe…,… Some

It was argued that a theory with local variational symmetries necessarily is described by a singular Lagrangian and that it acquires constraints in its Hamiltonian description. In the previous section, we at last saw the essential difference between regular and singular systems in that for the latter, there might remain arbitrary functions as multipliers of primary first-class constraints. An educated guess leads us to suspect that these constraints are related to the local symmetries on the Lagrange level. This guess points in the right direction, but things aren't that simple[4].

Dirac, in his famous lectures [129], introduced an influential invariance argument (in more detail described below) by which he conjectured that also secondary first-class constraints lead to invariances. His argumentation gave rise to the widely-held view that "first-class constraints are gauge generators". Aside from the fact that Dirac did not use the term "gauge" anywhere in his lectures, later work on relating the constraints to variational symmetries revealed that a detailed investigation on the full constraint structure of the theory in question is needed. The following analysis of Dirac's arguments is heavily influenced by [418].

Dirac's Conjecture and Variational Symmetries

Dirac's deliberations in [129] are as follows (in his own words): "For a general dynamical variable g with initial value g_0, its value at time δt is

$$g(\delta t) = g_0 + \delta t\ \dot{g} = g_0 + \delta t\ \{g, H_T\} = g_0 + \delta t\big[\{g, H'\} + v^\alpha\{g, \phi_\alpha\}\big].$$

The coefficients v are completely arbitrary and at our disposal. Suppose we take different values, v', for these coefficients. That would give a different $g(\delta t)$, the difference being $\Delta g(\delta t) = \delta t(v^{'\alpha} - v^\alpha)\{g, \phi_\alpha\}$. We may write this as

$$\Delta g(\delta t) = \epsilon^\alpha\{g, \phi_\alpha\}, \quad \text{(C.26)}$$

where

$$\epsilon^\alpha = \delta t(v^{'\alpha} - v^\alpha)$$

[4] Although one can simply check on the example of electrodynamics that there is no one-to-one relation between a constraint and a symmetry transformation: J.B. Pitts, "A First Class Constraint Generates Not a Gauge Transformation, But a Bad Physical Change: The Case of Electromagnetism", arXiv: 1310.2756

is a small arbitrary number, small because of the coefficient δt and arbitrary because the v's and v''s are arbitrary. We can change all our Hamiltonian variables in accordance with the rule (C.26) and the new Hamiltonian variables will describe the same state. This change in the Hamiltonian variables consists in applying an infinitesimal contact transformation with a generating function $\epsilon^\alpha \phi_\alpha$. We come to the conclusion that the ϕ_α's, which appeared in the theory in the first place as the primary first-class constraints, have this meaning: *as generating functions of infinitesimal contact transformations, they lead to changes in the q's and p's that do not affect the physical state*." (pp. 20–21 in [129]; italics by Dirac himself)

Dirac next calculates the commutator of two contact transformation with parameter ϵ, ϵ',

$$(\Delta_\epsilon \Delta_{\epsilon'} - \Delta_{\epsilon'} \Delta_\epsilon) g = \epsilon^\alpha \epsilon'^\beta \{g, \{\phi_\alpha, \phi_\beta\}\}$$

to show that $\{\phi_\alpha, \phi_\beta\}$ is also a generating function of an infinitesimal contact transformation in the sense outlined above. Because the ϕ_α are FC constraint functions their Poisson bracket is–as shown before–strongly equal to a linear combination of FC constraints. Since in general within this linear combination also secondary FC constraints can appear, one may conjecture that *all* FC constraints give rise to "changes in the q's and p's that do not affect the physical state"–and this is known as "Dirac's conjecture". It also led Dirac to introduce an "extended" Hamiltonian which contains all FC constraints

$$H_E = H_T + v^{\alpha'} \phi_{\alpha'} = H' + v^I \Phi_I; \tag{C.27}$$

and he remarks: "The presence of these further terms in the Hamiltonian will give further changes in g, but these further changes in g do not correspond to any change of state and so they should certainly be included, even though we did not arrive at these further changes of g by direct work from the Lagrangian" ([129], p. 25). Although, as shown below, this extended Hamiltonian has some interesting features, it is not logically substantiated by the consistent derivation of phase-space dynamics from configuration-velocity space dynamics.

Dirac connected first-class constraints with "changes in the q's and p's that do not affect the physical state" in an infinitesimal neighborhood of the initial conditions. Somehow, it became common wisdom to call these "gauge transformations"; but this has to be taken with caution. The term "gauge transformation" is otherwise mainly used in the context of Yang-Mill theories where it has the meaning of "local phase transformations". However by some, it is also used for reparametrization-invariant theories; there it has the meaning of "general coordinate transformations".

Lagrangian and Hamiltonian Symmetries

Before trying to specify the relation between symmetries in configuration-velocity space and those in phase space one needs to state which symmetries one is talking about in both spaces. In Subsect. 2.2 the Noether symmetries are defined with respect to the Lagrangian and restricted to point transformations. The Lie symmetries refer to point transformations of the second-order equations of motion. We saw that even for seemingly simple examples (like the free nonrelativistic particle) the discovery of these symmetries is far from trivial. How does one find all invariances of an action for more complicated cases, if not already obvious by construction of the action? There are examples from supersymmetry where the action was known first, but the symmetries were found only later. If it comes to the Hamiltonian version of a theory it is not at all obvious how its symmetries (which are defined below) are related to the Lagrangian symmetries, especially in the case of singular systems. It is not evident *a priori* that all Lagrangian symmetries are Hamiltonian symmetries and *vice versa*. There even are counterexamples where Lagrangian symmetries in the tangent space cannot be projected to Hamiltonian symmetries in the cotangent space, and where Hamiltonian symmetries do not have a Lagrangian counterpart.

In order to simplify the following consideration consider classical mechanics. In asking for Lagrangian symmetries one seeks all transformations $\delta_\epsilon q^i = f^i(q, \dot{q}, \ddot{q}, \ldots \epsilon^r, \dot{\epsilon}^r, \ddot{\epsilon}^r, \ldots)$ that leave the action $S_L = \int dt \;\; L(q, \dot{q})$ invariant. (Here, for simplicity, I refer to transformations of the dependent variables only.) In turns out that for a complete and consistent investigation of tangent-space symmetries and their relation to cotangent-space symmetries one must allow also for "trivial" transformations

$$\delta_\rho q^i = \int dt' \rho^{ij}(t, t') \frac{\delta S_L}{\delta q^j(t')} = [L]_k \, \rho^{ik}$$

where the coefficient functions ρ are antisymmetric: $\rho^{ij}(t, t') = \rho^{ji}(t', t)$. These transformations are trivial in the sense that they vanish on-shell.

For regular systems we derived in Sect. 2.2.3 that for the Hamiltonian form of the action $S_H = \int dt \; \big(p_i q^i - H(q, p)\big)$ Noether symmetry transformations can on-shell be expressed as canonical transformations. Also in the singular case, we are interested in finding a symmetry generator such that $g(\underline{\epsilon}) = g_I(q, p)\underline{\epsilon}^I$ with $\delta_{\underline{\epsilon}}(-) = \{-, g(\underline{\epsilon})\}$. Again there are trivial transformations

$$\delta_{\xi,\alpha} q^i = EM(q^j)\xi_j{}^i + EM(p_j)\alpha^{ji} \qquad \delta_{\xi,\beta} p_i = EM(p_j)\xi^j{}_i + EM(q^j)\beta_{ji}$$

with antisymmetric matrices α and β and where $EM(F) := \dot{F} - \{F, H\}$ are the Euler-Hamilton expressions, that is $EM(F) = 0$ is the Hamilton equation for the phase-space variable F. As it turns out, sometimes one needs to invoke trivial transformations in order to fully restore an equivalence between Lagangian and Hamiltonian constraints [374].

Under certain regularity restrictions and under conditions of projectability, the symmetries for singular systems can be algorithmically constructed from the chain of Lagrangian constraints required by the quest for time conservation of the primary (Lagrangian) constraints [84].

The authors of [274] base their derivation of symmetry transformations on an observation by C. Teitelboim[5]: Under the assumption that only first-class constraints

5 In his unpublished 1973 PhD thesis "The Hamiltonian structure of spacetime".

are present, the action with the extended Hamiltonian

$$S_E = \int d^4x(p_i\dot{q}^i - H_E) = \int d^4x(p_i\dot{q}^i - H_C - v^I\Phi_I) \tag{C.28}$$

changes only by a boundary term under the transformations

$$\delta_\epsilon q^i = \{q^i, \Phi_I\}\epsilon^I, \qquad \delta_\epsilon p_i = \{p_i, \Phi_I\}\epsilon^I \tag{C.29}$$

if the multipliers transform as

$$\delta_\epsilon v^I = \dot{\epsilon}^I + v^K\epsilon^J C_{KJ}^{\ \ I} - \epsilon^K H_K^I,$$

and where

$$\{\Phi_I, \Phi_J\} =: C_{IJ}^{\ \ K}\Phi_K \qquad \{H_C, \Phi_I\} =: H_I^K\Phi_K. \tag{C.30}$$

Under further specified regularity conditions, relations between Lagrangian symmetries and phase-space symmetries are derived, and it turns out that the generators for phase-space symmetries are linear combinations of all first-class constraints, with the peculiarity that the coefficients depend on the multipliers v^I.

Indeed, in the the literature you will find many articles in which the authors argue with the extended Hamiltonian. Since up to now there is no consensus about the relevance of this heuristically introduced object, it might be worthwhile how the previous consideration changes in case one starts with the total Hamiltonian

$$S_T = \int d^4x(p_i\dot{q}^i - H_T) = \int d^4x(p_i\dot{q}^i - H_C - u^\rho\phi_\rho). \tag{C.31}$$

Assume that it is invariant with respect to the transformations (C.29). First, we have

$$\delta_\epsilon(p_i\dot{q}^i) = \dot{\epsilon}^I\Phi_I + \frac{d}{dt}\left[p_i\delta_\epsilon q^i - \epsilon^I\Phi_I\right]$$

and

$$\delta_\epsilon H_T = \epsilon^I\{H_C, \Phi_I\} + \phi_\rho\delta_\epsilon u^\rho + u^\rho\epsilon^I\{\phi_\rho, \Phi_I\}.$$

Since the Φ_I are FC constraints, we find

$$\{\phi_\rho, \Phi_I\} =: C_{\rho I}^{\ \ \hat{\rho}}\phi_{\hat{\rho}} \qquad \{H_C, \Phi_I\} =: H_I^{\hat{\rho}}\phi_{\hat{\rho}}$$

where the summation (with respect to $\hat{\rho}$) extends over all constraints defining the final constraint surface in phase space. Thus as a condition for the invariance of the action (C.31), we derive

$$0 \stackrel{!}{=} \dot{\epsilon}^I\Phi_I - \epsilon^I G_I^{\hat{\rho}}\phi_{\hat{\rho}} - \phi_\rho\delta_\epsilon u^\rho \qquad \text{with} \qquad G_I^{\hat{\rho}} = H_I^{\hat{\rho}} + u^\rho C_{\rho I}^{\ \ \hat{\rho}}. \tag{C.32}$$

This is, as it stands, a relation involving constraints of different characteristics, the first term involving all first-class constraints, the second containing all constraints, and the third term primary constraints only. In this form one cannot say whether it has solutions for the symmetry parameters $\underline{\epsilon}^I$, or how many solutions there are. But there is a specific case in which the previous expression can be analyzed further: If all constraints are first-class primary, the transformations of the multipliers can be expressed as $\delta_{\underline{\epsilon}} u^I = \dot{\underline{\epsilon}}^I - \underline{\epsilon}^J G^I_J$. Notice that in this case the total and the extended Hamiltonians are identical and we are back at the previous consideration.

In [421], the condition for $G_\zeta(t)$ to be a canonical generator of infinitesimal phase-space transformations in the sense that

$$\delta_\zeta g = \{g, G_\zeta(t)\} \tag{C.33}$$

is derived as

$$\frac{\partial G_\zeta}{\partial t} + \{G_\zeta, H_T\} \simeq \text{PFC}$$

where PFC stands for a linear combination of primary first-class constraints. Since $H_T = H' + v^\alpha \phi_\alpha$, and remembering that the functions v^α which multiply the primary first-class constraints are arbitrary, we find the three conditions

$$G_\zeta(t) \qquad \text{is a first-class function} \tag{C.34a}$$

$$\frac{\partial G_\zeta}{\partial t} + \{G_\zeta, H'\} \simeq \text{PC} \tag{C.34b}$$

$$\{G_\zeta, PFC\} \simeq \text{PC}. \tag{C.34c}$$

where PC stands for a linear combinations of primary constraints, and PFC for any of the primary first class constraints.

Relating Lagrangian and Hamiltonian Symmetries

We saw that for all theories and models of fundamental physics the infinitesimal transformations on fields Q^α are of the generic form (3.68), that is

$$\delta_\epsilon Q^\alpha = \mathcal{A}^\alpha_r(Q) \cdot \epsilon^r(x) + \mathcal{B}^{\alpha\mu}_r(Q) \cdot \epsilon^r_{,\mu}(x) + \ldots, \tag{C.35}$$

and these are assumed to leave the action $S = \int d^D x \mathcal{L}(Q, \dot{Q})$ of the theory in question invariant. So why at all should the transformations (C.26)–and tentatively extended to all first-class constraints–investigated by Dirac in an infinitesimal neighborhood of the initial conditions, that is

$$\delta_\epsilon Q^\alpha = \epsilon^I \{Q^\alpha, \Phi_I\} \tag{C.36}$$

be related to the transformations (C.35) of a variational symmetry? Is there a mapping between the parameter functions ϵ^r and $\underline{\epsilon}^I$? Can one specify an algorithm to calculate the generators of Noether symmetries in terms of constraints? Or, in starting from the symmetry generating function G with the properties (C.34), is there a relation to the Dirac transformations (C.36) in terms of first-class constraints?

It seems that the very first people to address these questions were J. L. Anderson and P. G. Bergmann [9]–even before the Hamiltonian procedure for constrained systems was fully developed. N. Mukunda [376] started off from the chain (3.77) of Noether identities and built symmetry generators as linear combinations of first class primary and secondary constraints from them, assuming that no tertiary constraints are present. L. Castellani [81] devised an algorithm for calculating symmetry generators for local symmetries, implicitly neglecting possible second-class constraints. His algorithm was proven to be correct for those Noether transformations that are projectable from tangent to cotangent space.

The set of cotangent-space symmetry transformations characterized by (C.34) includes transformations that do not have a counterpart in tangent space. As shown in [34], the generator for those Noether symmetries that are projectable to phase space are characterized by

$$G_\zeta(t) \qquad \text{is a first-class function} \tag{C.37a}$$

$$\frac{\partial G_\zeta}{\partial t} + \{G_\zeta, H_C\} = \text{PC} \tag{C.37b}$$

$$\{G_\zeta, PC\} = \text{PC}. \tag{C.37c}$$

These conditions are more restrictive then (C.34) in that (1) "ordinary" equalities arise instead of strong ones, (2) instead of H' the canonical Hamiltonian enters the condition, (3) the last Poisson bracket contains all primary constraints (PC) and not only the PFC's.

In observing that for both Yang-Mills theories and for general relativity, the infinitesimal symmetry transformations can be expanded in terms of infinitesimal parameters and their derivatives, Anderson and Bergmann [9] already assumed an expansion

$$G_\zeta = \sum_{k=0}^{N} \zeta^{(k)}(t) G_k(q,p) \qquad \text{with} \qquad \zeta^{(k)}(t) := \frac{d^k\zeta}{dt^k}. \tag{C.38}$$

(In its full glory the ζ carries a further index denoting different symmetries.) If this *ansatz* is plugged into (C.37) then due to the arbitrariness of the ζ, one obtains

$$\{G_k, PC\} \simeq \text{PC} \tag{C.39}$$

and the conditions for Noether projectable transformations is expressed as the chain

$$G_N = \text{PFC} \tag{C.40a}$$

$$G_{k-1} + \{G_k, H_C\} = \text{PC}, \qquad \text{k} = \text{K}, \ldots, 1 \tag{C.40b}$$

$$\{G_0, H_C\} = \text{PC}. \tag{C.40c}$$

Thus one may start with taking G_N to be identified with an arbitrarily chosen primary first-class constraint. Next G_{N-1} is found to be a secondary first-class constraint (up to PC pieces), etc. One expects that any independent PFC can be taken as the head of the chain (C.40a). However, both conceptually and algorithmically, it makes a difference how the chain algorithm is started. It is even not clear from the very outset that solutions of (C.39) and (C.40) exist. It was shown in [233] under which conditions one can–at least locally–construct a complete set of independent symmetry transformations according to (C.38). As a corollary of this proof it is found that the number of independent Lie symmetry transformations is identical to the number of primary first-class constraints.

The symmetry parameter $\underline{\epsilon}^I$ in the generating function $g = \underline{\epsilon}^I \Phi_I$ can be expressed by the symmetry parameter ζ in the generator G_ζ. This was made specific for pure first-class systems in [103] and later in this appendix it is exemplified for the relativistic particle.

Finally, it remains to be settled how the Dirac-type transformations (C.26) are related to the transformations (C.35). For both transformations, it holds that

$$\delta g(t) = \{g, G_\zeta(t)\} = \sum_{k=0}^{N} \zeta^{(k)}(t)\{g, G_k\}.$$

Dirac's transformations are characterized by infinitesimal time intervals δt near $t = 0$. In order to guarantee $\delta g(0) = 0$, one must impose $\zeta^{(k)}(0) = 0\ \forall k$. Then to first order in δt,

$$\zeta^{(k)}(\delta t) = \zeta^{(k)}(0) + \delta t \cdot \zeta^{(k+1)}(0)$$

implying

$$\begin{aligned} \zeta^{(k)}(\delta t) &= 0 \qquad \text{for} \qquad k = 0, \ldots, N-1 \\ \zeta^{(N)}(\delta t) &= \delta t \cdot \zeta^{(N+1)}(0) \end{aligned}$$

where the value of $\zeta^{(N+1)}(0)$ is arbitrary. Choosing $\zeta^{(N+1)}(0) = (v'^{\alpha} - v^{\alpha})$ yields a transformation for each α of the form

$$\delta g(t) = \zeta\{g, G_N\}$$

which is a Dirac-type transformation (C.26), since, due to (C.40), G_N is a primary first class function.

Observables

A classical *observable* $\mathcal{O}(t)$ is defined as a phase space function that is invariant under the transformations generated by the $G(t)$, that is for which

$$\{\mathcal{O}(t), G(t)\} \approx 0. \tag{C.41}$$

This is equivalent to the vanishing of the Poisson bracket of an observable with all first-class constraints:

$$\{\mathcal{O}(t), \Phi_I\} \approx 0.$$

This property incidentally makes the discussion of whether to employ the total or the extended Hamiltonian superfluous: Although, in general, the equations of motions are different in both cases for generic fields, they are identical for observables.

C.1.5 Second-Class Constraints and Gauge Conditions

The previous subsection dealt at length with first-class constraints because they are related to variational symmetries of the theory in question. Second-class constraints χ_A enter the Hamiltonian equations of motion (C.25) without arbitrary multipliers. If there are no first-class constraints the dynamics is completely determined by

$$\dot{F}(p, q) \approx \{F, H_T\} \approx \{F, H_C\} - \{F, \chi_A\}\overline{\Delta}^{AB}\{\chi_B, H_C\}$$

without any ambiguity.

Dirac Bracket

Dirac introduced in [127] a “new P.b.”:

$$\{F, G\}^* := \{F, G\} - \{F, \chi_A\}\ \overline{\Delta}^{AB}\ \{\chi_B, G\} \tag{C.42}$$

nowadays called the *Dirac bracket* (DB). Sometimes for purposes of clarity it is judicious to indicate in the notation $\{F, G\}^*_\Delta$ that the DB is built with respect to the matrix Δ. The Dirac bracket satisfies the same properties as the Poisson bracket, i.e. it is antisymmetric, bilinear, and it obeys the product rule and the Jacobi identity. Furthermore, the DBs involving SC and FC constraints obey

$$\{F, \chi_A\}^* \equiv 0 \qquad \{F, \Phi_I\}^* \approx \{F, \Phi_I\}.$$

Thus when working with Dirac brackets, second-class constraints can be treated as strong equations. The equations of motion (C.25) written in terms of DBs are

$$\dot{F}(p, q) \approx \{F, H_T\}^*. \tag{C.43}$$

For later purposes let me mention an iterative property of the DB: First define the Dirac bracket for a subset of the SC constraints (denoted by $\chi_{A'}$) with $\det\{\chi_{A'}, \chi_{B'}\} \neq 0$:

$$\{F, G\}^*_{\Delta'} := \{F, G\} - \{F, \chi_{A'}\}\overline{\Delta'}^{A'B'}\{\chi_{B'}, G\}.$$

Next select further second-class constraints $\chi_{A''}$, such that $\{\chi_{A''}\} \cap \{\chi_{A'}\} = \{\}$, $\Delta''_{A''B''} = \{\chi_{A''}, \chi_{B''}\}$ with $\det \Delta''_{A''B''} \neq 0$ and

$$\{F, G\}^*_{\Delta''} := \{F, G\}^*_{\Delta'} - \{F, \chi_{A''}\}^*_{\Delta'} \overline{\Delta''} A'' B'' \{\chi_{B''}, G\}^*_{\Delta'}.$$

This iterative process can be pursued until all second-class constraints are exhausted. The result is the same as (C.42), that is as one would have computed the DB with respect to the full set of second-class constraints.

Just as Poisson brackets have a geometric and group-theoretical meaning, also the Dirac brackets are more than just a convenient compact notation; see e.g. [491].

Gauge Fixing

The existence of unphysical symmetry transformations notified by the presence of first-class constraints may make it necessary to impose conditions on the dynamical variables. This is specifically the case if the observables cannot be constructed explicitly–and this is notably true in general for Yang-Mills and for gravitational theories. These extra conditions are further constraints

$$\Omega_a(q, p) \approx 0, \tag{C.44}$$

where now weak equality refers to the hypersurface Γ_R defined by the weak vanishing of all previously found first- and second-class constraints, that is the hypersurface Γ_C together with the constraints (C.44). The idea is that the quest for stability of these constraints, namely

$$0 \overset{!}{\approx} \dot{\Omega}_a \approx \{\Omega_a, H_C\} + \{\Omega_a, \phi_\alpha\} v^\alpha - \{\Omega_a, \chi_A\} \overline{\Delta}^{AB} \{\chi_B, H_C\}$$

is meant to uniquely determine the multiplier v^α. At least for finite-dimensional systems, the previous condition can be read as a linear system of equations which has unique solutions if the number of independent gauge constraints is the same as the number of primary FC constraints and if the gauge constraints are chosen so that the determinant of the matrix

$$\Lambda_{\beta\alpha} := \{\Omega_\beta, \phi_\alpha\} \tag{C.45}$$

does not vanish[6]. In this case the multipliers are fixed to:

$$v^\alpha = \overline{\Lambda}^{\alpha\gamma} \big[- \{\Omega_\gamma, H_C\} + \{\Omega_\gamma, \chi_A\} \overline{\Delta}^{AB} \{\chi_B, H_C\} \big].$$

If there are no second-class constraints the Hamilton equations can be shown (see e.g. [491]) to be equivalent to

$$\dot{F}(p, q) \approx \{F, H_C\}^*_{\hat{\Lambda}} \qquad \text{with} \qquad \hat{\Lambda} = \begin{pmatrix} \{\phi_\alpha, \phi_\beta\} & \{\phi_\alpha, \Omega_\beta\} \\ \{\Omega_\alpha, \phi_\beta\} & \{\Omega_\alpha, \Omega_\beta\} \end{pmatrix}. \tag{C.46}$$

[6] This determinant is related to the Faddeev-Popov determinant which plays a central rôle in the path integral formulation of gauge systems; see Appendix D.1.3.

It is quite obvious how, by making use of the iterative property of the Dirac bracket, one can extend this to situations with SC constraints: Build the Dirac brackets with respect to an enlarged matrix containing both the Δ-matrix from the Poisson brackets of the SC constraints and the matrix Λ made from the gauge fixings and the primary first-class constraints.
Some remarks concerning the choice of gauge constraints Ω^α :

- The condition of a non-vanishing determinant $(\det(\Lambda_{\alpha\beta}) \neq 0)$ is only a sufficient condition for determining the arbitrary multipliers connected with the primary FC constraints.
- The gauge constraints must not only be such that the "gauge" freedom is removed (this is guaranteed by the non-vanishing of $\det \Lambda$), but also the gauge constraints must be accessible, that is for any point in phase space with coordinates (q, p), there must exist a transformation $(q, p) \to (q', p')$ such that $\Omega_\alpha(q', p') \approx 0$. This may be achievable only locally.
- We will see that in case of reparametrization invariance (at least one of) the gauge constraints must depend on the parameters explicitly–and not only on the phase-space variables.
- Especially in field theories it may be the case that no globally admissible (unique and accessible) gauge constraints exist. An example is given by the Gribov ambiguities, as they were first found in Yang-Mills theories.

C.1.6 Constraints in Field Theories: Some Remarks

The canonical formulation for singular Lagrangians was derived in the previous section for systems with a finite number of degrees of freedom. However, the motivation for the work of Rosenberg, the Syracuse group around Bergmann, and of Dirac were constraints in field theory (or even more specifically: in general relativity). Although many of the previous techniques and results can be directly transferred to field theories, one should be aware about some peculiarities.

First, constraints in a field theory are in general no longer functions of the phase-space fields alone, but they may also depend on spatial derivatives:

$$\phi_\rho[Q, \Pi, \partial_i Q, \partial_i \Pi] = 0.$$

Here "spatial" (and the index i) refer to $(D$-$1)$ dimensions other than the chosen time variables. In the instant form (with $x^0 \triangleq T$) it has the meaning of the Euclidean "spatial". As a consequence, constraints in field theories are no longer only pure algebraic relations among phase-space variables, but can very well be differential equations. Furthermore, with $\phi_\rho \approx 0$, also spatial derivatives and integrals of ϕ_ρ are weakly zero. Constraints in a field theory must be understood as having not only a discrete index (ρ) but additionally a continuous one, namely one for each point in spacetime.

In the context of identifying the constraints and the multiplier functions we need to solve linear equations which contain matrices defined in terms of Poisson brackets of constraints:

$$P_{rs}(\bar{x}, \bar{y}) = \{\phi_r(x), \phi_s(y)\}_{x^T = y^T}.$$

If P is non-singular, its inverse $\bar{P}^{rs}$ is not uniquely defined by

$$\int dy \; P_{rs}(x, y)\bar{P}^{st}(y, z) = \delta_r{}^t\delta(x - z)$$

and, in case $\det P = 0$, the null eigenvectors of P are also not defined uniquely.

We saw that a main step towards the final Hamiltonian for constrained systems is to identify the maximal number of first-class constraints and to distinguish these from the rest of all constraints. This task is purely algebraic for non-field theories. In field theories, first-class constraints may also be built by taking into account for instance spatial derivatives of other constraints.

C.1.7 Quantization of Constrained Systems

Already from "ordinary" quantum mechanics–and without phase-space constraints–we know that there is no royal road[7] connecting a classical theory with its quantum counterpart.

The "Dirac quantization" for systems with constraints largely builds on the Hilbert space method sketched in Sect. 4.1.4. (switching from classical observables to Hermitean operators in a Hilbert space, replacing Poisson brackets of phase-space objects by commutators of observables):

$$\{F, G\} \longrightarrow \frac{1}{i\hbar}[\hat{F}, \hat{G}]. \tag{C.47}$$

However, one has to keep in mind that due to the existence of constraints, the Hilbert space in which the quantum operators act is itself restricted–and this can be described explicitly only for specific cases. In any case, one may try one of the following:

1. Imagine having fixed the multipliers "in front" of the constraints by one way or another. Call the full set of constraints (including the gauge constraints) $\mathcal{C}_\alpha$. Use the substitution rule (C.47) as if there were no constraints. Require that matrix elements built with the constraints vanish. That is, define the Hilbert space $\mathcal{H}_P$ implicitly by those states which obey

$$\langle\psi'|\hat{\mathcal{C}}_\alpha|\psi\rangle \; = 0. \tag{C.48}$$

[7] This phrase is said to stem from Euclid, when he was asked by the ruler Ptolemy I Soter if there was a shorter road to learning geometry than through Euclid's Elements: "There is no royal road to geometry".

Aside from difficulties arising from finding the quantum counterparts $\hat{\mathcal{C}}_\alpha$ of the phase-space constraints, this condition is not very handy because you have to know the full spectrum of the Hamiltonian before you can decide which states do belong to $\mathcal{H}_P$.

2. Instead of (C.48) one may consider the stronger conditions

$$\hat{\mathcal{C}}_\alpha|\psi\rangle = 0.$$

But indeed this is much to strong, since for consistency one has to demand that also $[\hat{\mathcal{C}}_\alpha, \hat{\mathcal{C}}_\beta]|\psi\rangle = 0$, and this does not hold in general. It at most holds weakly for first-class constraints ϕ_α because $\{\phi_\alpha, \phi_\beta\} = g^\gamma_{\alpha\beta}\phi_\gamma$. But even in this case, one must allow for the possibility that the quantum analogue of the Poisson brackets does not neatly read $[\hat{\phi}_\alpha, \hat{\phi}_\beta] = \hat{g}^\gamma_{\alpha\beta}\hat{\phi}_\gamma$, since the structure coefficients g may become operators.

3. In the case of second-class constraints χ_A, there is no way to impose consistently the conditions $\hat{\chi}_A|\psi\rangle = 0$. This would make sense only if the operators $\hat{\chi}_A$ themselves vanish. Nevertheless, this opens another perspective of quantizing a system with phase-space constraints. We know that after complete gauge fixing, all constraints become second class and that, when using Dirac brackets instead of Poisson brackets, the constraints can be set to zero identically. Thus it is suggestive to replace the "rule" (C.47)

$$\{F, G\}^* \longrightarrow \frac{1}{i\hbar}[\hat{F}, \hat{G}]. \tag{C.49}$$

But of course the problems with operator ordering remain. Furthermore, the quantization may work in one gauge, but not in another.

Path Integral

Assume we are dealing with a theory with N degrees of freedom that has first-class constraints ϕ_α $(\alpha = 1, \dots M)$ only. According to [169], we may formally introduce the path-integral

$$\int \mathcal{D}p\mathcal{D}q\mathcal{D}\lambda \; \exp\left\{i\int_0^T [p_k\dot{q}^k - H(p,q) - \lambda^\alpha\phi_\alpha(p,q)]\right\}$$
$$= \int \mathcal{D}p\mathcal{D}q \; \delta\big(\phi_\alpha(p,q)\big) \exp\left\{i\int_0^T [p_k\dot{q}^k - H(p,q)]\right\}.$$

Introduce M complementary gauge constraints $\Omega_\alpha(p,q) = 0$, such that the original symmetry is completely broken. As mentioned, a necessary condition is $\det\{\Omega_\alpha, \phi_\beta\} \neq 0$. (For simplicity we may assume that there are no Gribov ambiguities.) Then, as shown by Faddeev,

$$\int \mathcal{D}p\mathcal{D}q\ \delta\big(\phi_\alpha(p,q)\big)\delta\big(\Omega_\alpha(p,q)\big)\det\{\Omega_\alpha,\phi_\beta\}\exp\left\{i\int_0^T [p_k\dot{q}^k - H(p,q)]\right\}$$
$$= \int \mathcal{D}p^*\mathcal{D}q^*\ \exp\left\{i\int_0^T [p^*_j\dot{q}^{*j} - H(p^*,q^*)]\right\}$$

where the (q^*, p^*) are the coordinates in the $(N\text{-}M)$-dimensional reduced phase space. Faddeev's idea was extended to theories with quite general gauge constraints and additional second-class constraints by Batalin, Vilkovisky, Fradkin, Fradkina,… in the seventies (more details may be found [491], Chap. IV).

Field-Antifield Formalism

This was the forerunner of BRST ghost/anti-ghost techniques, and their specific embodiment for constrained dynamics; see [268], [273]. These techniques enable the quantization of any kind of system with variational symmetries in the broadest sense (reducible gauge algebra, soft algebra, open algebra, etc.)

C.2 Yang-Mills Type Theories

In order to get a feeling about the generic Hamiltonian structure of Yang-Mills fields coupled to spinor fields, we first consider vacuum electrodynamics, next the Maxwell-Dirac theory ("classical QED"), and then the pure Yang-Mills theory.

C.2.1 Electrodynamics

Start from the Lagrangian $\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$, and opt for the instant form $(x^0 \mathrel{\hat{=}} T)$. The Hessian

$$\mathcal{W}^{\mu\nu} := \frac{\partial^2\mathcal{L}}{\partial(\partial_0 A_\mu)\partial(\partial_0 A_\nu)} = \eta^{0\mu}\eta^{0\nu} - \eta^{\mu\nu}\eta^{00},$$

has rank $(D\text{-}1)$ in D dimensions. Thus, one primary constraint can be expected. From

$$\Pi^\mu := \frac{\partial\mathcal{L}}{\partial(\partial_0 A_\mu)} = -F^{0\mu} = \partial^\mu A^0 - \partial^0\, A^\mu, \tag{C.50}$$

the momentum canonically conjugate to A_0, namely

$$\phi_1 := \Pi^0(x) \approx 0 \tag{C.51}$$

is detected as the primary constraint. The canonical Hamiltonian is

$$\mathcal{H}_C = \Pi^\mu\dot{A}_\mu - \mathcal{L} = \frac{1}{2}\Pi_i\Pi^i - A_0\,\partial_i\Pi^i + \frac{1}{4}F_{ik}F^{ik} + \partial_i(\Pi^i A_0). \tag{C.52}$$

The total Hamiltonian $\mathcal{H}_T = \mathcal{H}_C + u_1\Pi^0$ contains the arbitrary function $u_1(x)$ in correlation with the primary constraint. Because of the canonical Poisson brackets

$\{A_\mu(x), \Pi^\nu(y)\}_{x^0=y^0} = \delta^\nu_\mu\, \delta^3(\vec{x}-\vec{y})$–all others vanishing–one finds

$$0 \overset{!}{\approx} \dot{\phi}_1 = \{\Pi^0, H_T\} = -\{\Pi^0, \int d^3x\ A_0\partial_i\Pi^i\} = \partial_i\Pi^i.$$

and thus discovers the secondary constraint

$$\phi_2 := \partial_i\Pi^i \approx 0. \tag{C.53}$$

Requiring that the secondary constraint vanish weakly does not enforce further constraints. Since $\{\phi_1, \phi_2\} = \{\Pi^0, \partial_i\Pi^i\} = 0$ the two constraints are first class. Observe that because of $\Pi^i = -F^{0i} = -E^i$ the constraint $\partial_i\Pi^i \approx 0$ is simply Gauß' Law, that is the first of the Maxwell Eqs. (3.1a).

Section C.1.5 dealt with the relationship of first-class constraints and symmetry generators. One may try to confine oneself to the only PFC and set $\Pi^0 = G_0$ in (C.40a). However, for this choice, $\{G_0, H_C\} \neq$ PC. The next attempt is $\Pi^0 = G_1$ and indeed this gives the solution

$$G_\zeta = \int d^3x(\dot{\zeta}\Pi^0 - \zeta\partial_i\Pi^i) = \int d^3x\ (\partial_\mu\zeta)\Pi^\mu + \text{b.t.} \tag{C.54}$$

The transformations mediated by this generator on the phase-space variables are

$$\begin{aligned}\delta_\zeta A_\mu = \{A_\mu, G_\zeta\} &= \{A_\mu(x), \int d^3y\ (\partial_\nu\zeta)\Pi^\nu\}\\ &= \int d^3y\ \partial_\nu\zeta(y)\ \{A_\mu(x), \Pi^\nu(y)\} = \partial_\mu\zeta\\ \delta_\zeta\Pi^\mu = \{\Pi^\mu, G_\zeta\} &= \{\Pi^\mu(x), \int d^3y\ (\partial_\nu\zeta)\Pi^\nu\} = 0.\end{aligned}$$

The transformation of the A_μ are indeed the Noether symmetries (5.47) identifying the infinitesimals as $\zeta = (1/e)\theta$–e being the charge. The cotangent space transformation of Π^μ is completely compatible with the transformation in tangent space, since by the definition (C.50), one obtains $\delta\Pi^\mu = \partial^\mu\delta A^0 - \partial^0\delta A^\mu = 0$.

So far, the treatment of Hamiltonian electrodynamics fully respects the gauge symmetry of the theory. The gauge symmetry can be broken by choosing two gauge conditions Ω_I (two, because both first-class constraints make their appearance in G_ϵ according to (C.54)), such that the matrix $\{\Omega_I, \Phi_J\}$ is nonsingular. An admissible gauge choice is e.g. the radiation gauge

$$\Omega_1 := A_0 \approx 0, \quad \Omega_2 := \partial_i A_i \approx 0.$$

The matrix $\hat{\Lambda}$ defined in (C.46) is

$$\hat{\Lambda}(\vec{x}, \vec{y})) = \begin{pmatrix} 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -\nabla^2\\ -1 & 0 & 0 & 0\\ 0 & \nabla^2 & 0 & 0\end{pmatrix}\ \delta(\vec{x}-\vec{y}).$$

Indeed this matrix is nonsingular; it can also be shown that by this choice the original gauge freedom $\delta A_\mu = \partial_\mu \epsilon$ is completely broken [491]: the function $\epsilon(x)$ is fully determined by the radiation gauge. The Dirac brackets is shaped by the inverse of $\hat{\Lambda}$, which is

$$\hat{\Lambda}^{-1} = \begin{pmatrix} 0 & 0 & -\delta(\vec{x}-\vec{y}) & 0 \\ 0 & 0 & 0 & -\frac{1}{4\pi|\vec{x}-\vec{y}|} \\ \delta(\vec{x}-\vec{y}) & 0 & 0 & 0 \\ 0 & \frac{1}{4\pi|\vec{x}-\vec{y}|} & 0 & 0 \end{pmatrix}.$$

From this, the fundamental Dirac brackets are calculated as

$$\{A_\mu(x), \Pi^\nu(y)\}^* = (\delta_\mu^\nu + \delta_\mu^0 \eta^{\nu 0})\delta^3(\vec{x}-\vec{y}) - \partial_\mu \partial^\nu \frac{1}{4\pi|\vec{x}-\vec{y}|}$$

–all others vanishing. The DB's different from the PB's are

$$\begin{aligned} \{A_0(x), \Pi^\nu(y)\}^* &= 0 \qquad \{A_\mu(x), \Pi^0(y)\}^* = 0 \\ \{A_i(x), \Pi_j(y)\}^* &= \delta_{ij}\, \delta^3(\vec{x}-\vec{y}) - \partial_i \partial_j \frac{1}{4\pi|\vec{x}-\vec{y}|} =: \delta_{ij}^{\mathrm{tr}}(\vec{x}-\vec{y}). \end{aligned}$$

The first two of these bracket relations are in accordance with $\Pi^0 = 0 = A_0$, and the latter one is in line with $\partial_i \Pi_i = 0 = \partial_i A_i$. Here $\delta_{ij}^{\mathrm{tr}}(\vec{x}-\vec{y})$ denotes the transversal delta function which was introduced *ad hoc* in previous treatments of electrodynamics.

While in the case of vacuum electrodynamics it is possible to identify the true degrees of freedom and thus to specify explicitly the reduced phase space, this is no longer true in non-Abelian gauge theories.

If *in lieu* of the instant form, the front form (that is $T \hat{=} x^+ = (x^0 - x^3)$) had been chosen, then instead of one primary constraint there would have been three primary constraints and one secondary constraint. Two linear combinations can be made first class and two are second class. The latter serve to define the Dirac brackets, and the Hamiltonian formulation in front form eventually turns out to be completely equivalent to the one in instant form, for details see [491].

C.2.2 Maxwell-Dirac Theory

Our starting point is the QED-Lagrangian (5.51)

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + e\bar{\psi} A_\mu \gamma^\mu \psi. \tag{C.55}$$

for which the canonical momenta (in the instant form) are computed as

$$\begin{aligned} \Pi^\mu &:= \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = -F^{0\mu} \\ \pi^\alpha &:= \frac{\partial \mathcal{L}}{\partial(\partial_0 \psi_\alpha)} = -i(\bar{\psi}\gamma^0)^\alpha \qquad \bar{\pi}^\alpha := \frac{\partial \mathcal{L}}{\partial(\partial_0 \bar{\psi}_\alpha)} = 0. \end{aligned}$$

We spot the primary constraints directly:

$$\phi := \ \Pi^0 \ \approx 0, \qquad \phi^\alpha := \pi^\alpha + \ i(\bar{\psi}\gamma^0)^\alpha \approx 0, \qquad \bar{\phi}^\alpha := \bar{\pi}^\alpha \ \approx 0. \tag{C.56}$$

The canonical Hamiltonian density is found to be

$$\mathcal{H}_C \ = \ \mathcal{H}_C^{ED} - i\bar{\psi}\gamma^k\partial_k\psi + m\bar{\psi}\psi + e\bar{\psi}A_\mu\gamma^\mu\psi,$$

where $\mathcal{H}_C^{ED}$ is the canonical Hamiltonian density for the free electromagnetic field (C.52). The total Hamiltonian is thus

$$H_T = H_C + \int d^3x \ (u\phi + \ \phi_\alpha v_\alpha + w_\alpha\bar{\phi}_\alpha) \tag{C.57}$$

with multiplier functions $\{u, \ v_\alpha, \ w_\alpha\}$. Stabilization of the primary constraints requires

$$0 \overset{!}{\approx} \{\phi, H_T\} = \ \partial_k\Pi^k \ + \ e\bar{\psi}\gamma^0\psi \tag{C.58a}$$

$$0 \overset{!}{\approx} \{\phi_\alpha, H_T\} = \ i\partial_k(\bar{\psi}\gamma^k)_\alpha + m\bar{\psi}_\alpha - e(\bar{\psi}\gamma_\mu A^\mu)_\alpha - i(w\gamma_0)_\alpha \tag{C.58b}$$

$$0 \overset{!}{\approx} \{\phi_\alpha, H_T\} = \ i\partial_k(\gamma^k\psi)_\alpha - m\psi_\alpha + e(\gamma_\mu A^\mu\psi)_\alpha - i(\gamma_0 v)_\alpha. \tag{C.58c}$$

To calculate these expressions, one needs the Bose-Fermi Poisson brackets (see Appendix B.2.4), that is specifically the non-vanishing equal-time brackets

$$\{A_\mu(x), \Pi^\nu(y)\} = \ \delta^\nu_\mu\,\delta(\vec{x} - \vec{y}) \tag{C.59a}$$

$$\{\psi_\alpha, \pi^\beta\} = -\delta^\beta_\alpha\,\delta(\vec{x} - \vec{y}) \qquad\qquad \{\bar{\psi}_\alpha, \bar{\pi}^\beta\} = -\delta^\beta_\alpha\,\delta(\vec{x} - \vec{y}). \tag{C.59b}$$

Conditions (C.58b) and (C.58c) can be fulfilled by an appropriate choice of the multiplier functions w_α and v_α. On the other hand, (C.58a) constitutes a secondary constraint

$$\phi_G := \ \partial_k\Pi^k \ + \ e\bar{\psi}\gamma^0\psi \ \approx 0. \tag{C.60}$$

Observe that this is again Gauß's law, now with the charge density $j_0 = e\bar{\psi}\gamma^0\psi$. Requiring next that the time derivative of ϕ_G vanish on the constraint surface does not lead to further constraints–the straightforward but tedious calculation is suppressed here. Thus at the end of the algorithm we are left with a set of 10 constraints[8] (in four dimensions): $\{\phi, \phi^\alpha, \bar{\phi}^\alpha, \phi_G\}$.

What about the maximal number of first-class constraints that can be built from these constraints? One immediately finds that the constraint function ϕ has a vanishing Poisson bracket with all other nine constraints. Thus, $\Phi_1 = \phi \ \ \approx 0$ is a first-class constraint. A further one is found to be the combination

$$\Phi_2 := \phi_G + ie(\phi^\alpha\psi_\alpha + \bar{\psi}_\alpha\bar{\phi}^\alpha) = \partial_k\Pi^k - ie(\pi^\alpha\psi_\alpha + \bar{\pi}^\alpha\bar{\psi}_\alpha).$$

[8] or $10 \times \infty$–to be pedantic

Dirac brackets are to be built with the second-class constraint functions ϕ^α and $\bar{\phi}^\alpha$. The matrix corresponding to (C.22) is

$$\Delta(\vec{x},\vec{y}) = -i \begin{pmatrix} 0 & (\gamma^0)^T \\ \gamma^0 & 0 \end{pmatrix} \delta(\vec{x}-\vec{y}) \quad \Delta^{-1}(\vec{x},\vec{y}) = i \begin{pmatrix} 0 & \gamma^0 \\ (\gamma^0)^T & 0 \end{pmatrix} \delta(\vec{x}-\vec{y})$$

and the Dirac bracket becomes

$$\begin{aligned}\{F(x),G(y)\}^*_\Delta =& \{F(x),G(y)\} - i(\gamma^0)_{\alpha\beta}\int dz\,\{F(x),\phi^\alpha\}\{\bar{\phi}^\beta,G(y)\} \\ & - i(\gamma^0)_{\alpha\beta}\int dz\,\{F(x),\bar{\phi}^\alpha\}\{\phi^\beta,G(y)\}.\end{aligned}$$

The Dirac brackets differing from the fundamental brackets (C.59b) are

$$\{\bar{\psi}_\alpha,\bar{\pi}^\beta\}^*_\Delta = 0 \tag{C.61a}$$

$$\{\psi_\alpha,\bar{\psi}_\beta\}^*_\Delta = -i(\gamma^0)_{\alpha\beta}\,\delta(\vec{x}-\vec{y}) \tag{C.61b}$$

which are compatible with the constraints $\phi^\alpha \approx 0 \approx \bar{\phi}^\alpha$ and confirm that one is allowed to interpret the second-class constraints as strong equations. Finally the gauge generator is

$$G_\zeta = \int d^3x\left[\dot{\zeta}\Pi^0 - \zeta\big(\partial_k\Pi^k - ie(\pi^\alpha\psi_\alpha + \bar{\pi}^\alpha\bar{\psi}_\alpha)\big)\right].$$

C.2.3 Non-Abelian Gauge Theories

The Yang-Mills Lagrangian

$$L = -\frac{1}{4}\int d^3x\; F^a_{\mu\nu}F^{\mu\nu}_a \qquad \text{with} \qquad F^a_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - gA^b_\mu A^c_\nu f^{bca}$$

contains the momenta (canonically conjugate to the A^a_μ)

$$\Pi^\mu_a = -F^{0\mu}_a. \tag{C.62}$$

Here, we assume again the instant form ($x^0 = T$) and thus the fundamental Poisson brackets are

$$\{A^a_\mu(x),\Pi^\nu_b\}_{x^0=y^0} = \delta^\nu_\mu\delta^a_b\delta(\vec{x}-\vec{y}).$$

The antisymmetry of the field-strength components gives the primary constraints

$$\phi_a = \Pi^0_a \approx 0. \tag{C.63}$$

The canonical Hamiltonian

$$H_C = \int d^3x\ (\Pi^\mu_a \dot{A}^a_\mu - L) = \int d^3x\ \big(\Pi^k_a \dot{A}^a_k - \frac{1}{2}\Pi^\mu_a \Pi^a_\mu + \frac{1}{4}F^a_{kl}F^{kl}_a\big)$$

becomes independent of the “velocities”, due to $\dot{A}^a_k = F^a_{0k} + \partial_k A^a_0 + gA^b_0 A^c_k f^{bca}$:

$$H_C = \int d^3x\ \big(\frac{1}{2}\Pi^\mu_a \Pi^a_\mu + \frac{1}{4}F^a_{kl}F^{kl}_a - A^a_0(\partial_k \Pi^k_a - g f_{ac}{}^b A^c_k \Pi^k_b) + b.t.\big) \quad \text{(C.64)}$$

(after another partial integration). The algorithm for finding possible further constraints starts with requiring that

$$0 \overset{!}{\approx} \{\phi_a, H_T\} = \{\phi_a, H_C\} + \{\phi_a, \int d^3x\ u^b\ \phi_b\} =: \varphi_a$$

which leads to the secondary constraints

$$\varphi_a = \partial_k \Pi^k_a - g f_{ac}{}^b A^c_k \Pi^k_b \approx 0. \quad \text{(C.65)}$$

With the gauge group covariant derivative (5.63), this constraint can be written as $\varphi_a = D_k \Pi^k_a$, which directly shows the generalization from the secondary constraint in the Abelian (Gauß’s Law) to the non-Abelian case. Checking for stability of the constraints φ_a as $0 \overset{!}{\approx} \{\varphi_a, H_T\} \approx \{\varphi_a, H_C\} = g f_{ab}{}^c A^b_0\ \varphi_c$, one observes that this vanishes on the hypersurface defined by $\phi_a \approx 0 \approx \varphi_a$. Thus, there are no further constraints. Both the primary and secondary constraints are first class, since

$$\{\Pi^0_a(x), D_i \Pi^i_a(y)\} = 0,$$
$$\{D_i \Pi^i_a(x), D_j \Pi^j_b(y)\} = -g f_{ab}{}^c\ (D_k \Pi^k_c(x))\ \delta(\vec{x} - \vec{y}) \approx 0.$$

In order to find the gauge generators one may start the chain (C.40) with $G_{1(a)} = \Pi^0_a$ and

$$G_{0(a)} = -\{\Pi^0_a, H_C\} + \lambda_{ab}\Pi^0_b = -\varphi_a + \lambda_a{}^b \Pi^0_b.$$

The coefficients $\lambda_a{}^b$ in the linear combination of the primary first-class constraints are determined from

$$PC \overset{!}{=} \{G_{0(a)}, H_C\} = \{\varphi_a, H_C\} + \lambda_a{}^b \{\Pi^0_b, H_C\} = g f_{ac}{}^b A^c_0 \varphi_b + \lambda_a{}^b \varphi_b$$

with the result

$$G_\zeta = \int d^3x (\dot{\zeta}^a \Pi^0_a - \zeta^a \big(D_i \Pi^i_a - g f_{ac}{}^b A^c_0 \Pi^0_b)\big) = \int d^3x (D_\mu \zeta^a) \Pi^\mu_a + \text{b.t.} \quad \text{(C.66)}$$

The transformations generated by G_ζ

$$\delta_\zeta A^a_\mu = \{A^a_\mu(x) \int d^3y\ (D_\nu \zeta^b)\Pi^\nu_b\} = D_\mu \zeta^a$$
$$\delta_\zeta \Pi^\mu_a = \{\Pi^\mu_a(x)\ \int d^3y\ (D_\nu \zeta^b)\Pi^\nu_b\}$$
$$= \int d^3y\ \Pi^\nu_b(y)\{\Pi^\mu_a(x)\ g f_{cd}{}^b A^c_\nu \zeta^d(y)\} = -g f_{ad}{}^b \Pi^\mu_b \zeta^d$$

correspond to the Noether symmetry transformations (5.62) if $\zeta^a = (1/g)\theta^a$, those for the momenta being compatible with the identifications (C.62).

C.3 Reparametrization-Invariant Theories

The terms in this section heading are essentially synonymous to "generally covariant" or "diffeomorphism invariant", the prime example from fundamental physics being general relativity. But even classical physics can be reformulated in a reparametrization-invariant way, and this has various benefits, see Chap. 3 of [451]. Many of the results in Subsect. C.3.1 hold independently of the specific example. Some of these are directly visible in the most simple case ("GR in one dimension"), the relativistic point particle. In stepping up to higher dimensions, one would arrive at relativistic string models, at membranes, and many other extended entities which are more or less fashionable in modern theoretical physics.

The reason why Yang-Mills type theories are treated in a section separate from reparametrization-invariant theories is substantiated by the fact that there is an essential difference in their Hamiltonian description: As will be demonstrated, the total Hamiltonian for reparametrization-invariant theories is a sum of constraints. This odd feature led to a plethora of prejudices, misunderstandings and still unsettled questions in the Hamiltonian treatment of generally covariant systems, and as such for the canonical formulation of GR.

In this section, I start with disclosing some basic features of reparametrization-invariant theories. Then the Hamiltonian analysis of the relativistic particle serves as a preliminary stage to canonical GR. In this book, I only sketch the treatment of vacuum GR (ADM approach for metric gravity, tetrad gravity, Ashtekar variables). But I close the section by presenting some essential peculiarities in the treatment of GR in the presence of Yang-Mills and/or Dirac fields.

C.3.1 Immediate Consequences of Reparametrization Invariance

The immediate consequences of reparametrization invariance are explicitly demonstrated to hold for systems with a finite number of degrees of freedom, but can straightforwardly shown to retain validity also for field theories.

1. Consider a system with an action $S = \int_{t_1}^{t_2} dt\; L(q^k, \dot{q}^k)$. If the Lagrange function fulfills
$$L(q, \lambda\dot{q}) = \lambda L(q, \dot{q}), \tag{C.67}$$
then the action is invariant with respect to the reparametrization $t \to \hat{t}(t)$. In order to interpret $\hat{t}(t)$ again as "time", the transformations need to be restricted to strictly monotonous functions, or $d\hat{t}/dt > 0$. Furthermore, $\hat{t}(t_i) = t_i$, that is, the boundary is mapped to itself.
2. The relation (C.67) expresses the fact that the Lagrange function is homogeneous (of first degree) in the velocities. Differentiate both sides with respect to λ:
$$\frac{\partial L(q, \lambda\dot{q})}{\partial \lambda\dot{q}^j}\dot{q}^j = L(q, \dot{q})$$
and use (C.67) again to derive (after renaming/rescaling the velocities)
$$\frac{\partial L}{\partial \dot{q}^j}\dot{q}^j - L = 0. \tag{C.68}$$
On the other hand if (C.68) holds true, we regain (C.67) by integration.
3. The condition (C.68) is equivalent to a vanishing canonical Hamilton function (by definition of the canonical momenta as $p_j = \partial L/\partial \dot{q}^j$):
$$H_C = p_j\dot{q}^j - L = 0. \tag{C.69}$$
4. Differentiate (C.68) with respect to the velocities:
$$0 = \frac{\partial}{\partial \dot{q}^k}\left[\frac{\partial L}{\partial \dot{q}^j}\dot{q}^j - L\right] = W_{kj}\dot{q}^j + \frac{\partial L}{\partial \dot{q}^k} - \frac{\partial L}{\partial \dot{q}^k}$$
from which it is seen that the Hessian has the zero eigenvector $(\dot{q}^j)$:
$$W_{kj}\dot{q}^j = 0. \tag{C.70}$$
5. The invariance with respect to the reparametrization $t \to \hat{t}$ brings about a Noether identity. Instead of deriving it according to the procedure of Sect. 3.3., proceed in the following way: Contract the Euler-Lagrange derivatives $[L]_i$ with $\dot{q}^i$, i.e. consider
$$[L]_i\dot{q}^i = \left[\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right]\dot{q}^i = \frac{\partial L}{\partial q^i}\dot{q}^i - \frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\right] + \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i.$$
Now make use of (C.68) in the second term to replace it by $\frac{d}{dt}L$. Next compute the time derivative of the Lagrangian in order to find that it cancels with the two other terms. Thus
$$[L]_i\dot{q}^i = 0 \tag{C.71}$$
is the Noether identity.

6. With a similar consideration, we derive the variation $\bar{\delta}L$ assuming that the q^i transform as scalars under the reparametrization; infinitesimally:

$$\delta t = \epsilon \qquad \text{and} \qquad \delta q^i = 0, \qquad \bar{\delta} q^i = -\dot{q}_i \epsilon.$$

From this

$$\bar{\delta}L = \frac{\partial L}{\partial q^i}\bar{\delta}q_i + \frac{\partial L}{\partial \dot{q}^i}\bar{\delta}\dot{q}^i = -\frac{\partial L}{\partial q^i}\dot{q}^i\epsilon - \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i\epsilon - \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\dot{\epsilon}.$$

Now using (C.68), the last term becomes $(-L\dot{\epsilon})$ and the two first terms add up to $(-\dot{L}\epsilon)$. Thus, independently of the specific or explicit form of the Lagrangian, it transforms as a scalar density:

$$\bar{\delta}L = -\frac{d}{dt}(L\epsilon). \tag{C.72}$$

All previous results can be extended to reparametrization-invariant field theories in D dimensions. The homogeneity conditions (C.67) and (C.68) become

$$\mathcal{L}(Q, \lambda^\nu_\mu \partial_\nu Q) = \det(\lambda)\mathcal{L} \quad \longleftrightarrow \quad \frac{\partial \mathcal{L}}{\partial(\partial_\mu Q^\alpha)}\partial(\partial_\nu Q^\alpha) = \delta^\nu_\mu \mathcal{L}.$$

From these, one derives the vanishing of the canonical Hamiltonian according to (C.69) and the singularity of the Hessian according to (C.70) (depending on one's choice of the temporal parameter). The D Noether identities (C.71) correspond to

$$[L]_\alpha \partial_\mu Q^\alpha = 0.$$

Finally for $\delta x^\mu = \epsilon^\mu$, (C.72) becomes generalized to

$$\bar{\delta}\mathcal{L} = -\partial_\mu(\mathcal{L}\epsilon^\mu).$$

C.3.2 Free Relativistic Particle

World-Line Description

The trajectory of a massive relativistic particle can be written in parametric form $x^\mu(\tau)$ where τ labels the trajectory. The action from which to derive its equation of motion may be taken to be proportional to the length of the world line as

$$S = -m \int d\tau (\eta_{\mu\nu}\dot{x}^\mu \dot{x}^\nu)^{1/2}. \tag{C.73}$$

As shown in Subsect. 3.2.3 this reproduces in the nonrelativistic limit the dynamics for the free particle. The Lagrangian to (C.73) is homogeneous in the velocities: $L(\lambda\dot{x}^\mu) = \lambda L(\dot{x}^\mu)$. The $(D \times D)$ Hessian associated with this Lagrangian is

$$W_{\mu\nu} := \frac{\partial^2 L}{\partial\dot{x}^\mu\partial\dot{x}^\nu} = \frac{m}{(\dot{x}^2)^{3/2}}(\dot{x}_\mu\dot{x}_\nu - \eta_{\mu\nu}\,\dot{x}^2),$$

and it has rank $(D\text{-}1)$. The zero eigenvector is easily recognized as being proportional to $\dot{x}^\mu$ in accordance with (C.70). The Euler-Lagrange derivatives are

$$[L]_\mu = -\frac{d}{d\tau}\frac{\partial L}{\partial\dot{x}^\mu} = -\frac{\partial^2 L}{\partial\dot{x}^\mu\partial\dot{x}^\nu}\,\ddot{x}^\nu = -W_{\mu\nu}\ddot{x}^\nu$$

from which the identity $[L]_\mu\dot{x}^\mu \equiv 0$ immediately results, revealing that only $(D\text{-}1)$ equations of motion are independent. This is off course the Noether identity, corresponding to (C.71), and stemming from the invariance of the action under reparametrization, infinitesimally

$$\delta\tau = \epsilon(\tau) \qquad \text{and} \qquad \delta_\epsilon x^\mu = 0 \qquad \bar{\delta}_\epsilon x^\mu = -\dot{x}^\mu\epsilon, \tag{C.74}$$

where $\bar{\delta}L = -d/d\tau(\epsilon L)$, again in compliance with the generic results of the previous subsection.

Let us proceed to the Hamiltonian formulation for this system: The canonical momenta to the x^μ are

$$p_\mu = \frac{\partial L}{\partial\dot{x}^\mu} = -m\frac{\dot{x}_\mu}{\sqrt{\dot{x}^2}}. \tag{C.75}$$

We expect one primary constraint since the rank of the Hessian is 1. This constraint can either be found by solving the $p_\mu(\dot{x}^\nu)$ for as many "velocities" as possible, or by observing that

$$p_\mu p^\mu = m^2\frac{\dot{x}_\mu\dot{x}^\mu}{\dot{x}^2} = m^2.$$

This is just the mass-shell condition, in this context to be understood as a constraint among the momenta which we write as

$$H_\perp := \frac{1}{2}(p^2 - m^2) \approx 0. \tag{C.76}$$

In going through the Rosenfeld-Dirac-Bergmann procedure, we first find that the canonical Hamiltonian $H_C = p_\mu\dot{x}^\mu - L$ vanishes identically–in agreement with the findings in the previous paragraph. Thus the total Hamiltonian is proportional to the constraint: $H_T = vH_\perp$ with the multiplier v. No further constraints exist.

In what sense is the constraint $H_\perp$ a symmetry generator? In order to answer this question we need to relate the infinitesimal transformation in phase space

$$\delta_\xi x^\mu := \{x^\mu, \xi H_\perp\} = \xi p^\mu \qquad \delta_\xi p_\mu := \{p_\mu, \xi H_\perp\} = 0$$

to the reparametrization (C.74). Thus due to (C.75) we seek the correspondence

$$-\xi m(\dot{x}^2)^{-1/2}\dot{x}^\mu \longleftrightarrow -\dot{x}^\mu\epsilon.$$

This demonstrates that the infinitesimal parameter ξ mediating the symmetry transformation in phase space must depend on $\dot{x}^\mu$, that is the "time" derivative of the "fields" x^μ. This is a typical feature of reparametrization-invariant theories (and as such also for general relativity) which seems to be little known. As pointed out by P. Bergmann and A. Komar [44], the action (C.73) is not only invariant with respect to transformations $\delta\tau = \epsilon(\tau)$, but also to those where $\delta\tau = \epsilon[\tau, x^\mu]$, that is where ϵ is a functional of τ and x^μ. In phase space, however, these "field"-dependent transformations are not arbitrary. In this case of a relativistic particle we have $\delta\tau = (\dot{x}^2)^{-1/2}\epsilon(\tau)$. The deeper reason of this dependency is the demand for projectability of the Lagrangian transformation $\delta_\epsilon x^\mu$ under the Legendre transformation (see C.5.1.): With the zero eigenvector $(\dot{x}^\mu)$, the operator for checking projectability is $\Gamma = \dot{x}^\nu\ \partial/\partial\dot{x}^\nu$ according to (C.118), so that

$$0 \stackrel{!}{=} \Gamma(\delta_\epsilon x^\mu) = \dot{x}^\nu\frac{\partial}{\partial\dot{x}^\nu}(-\dot{x}^\mu\epsilon) = -\dot{x}^\mu\epsilon - \dot{x}^\nu\dot{x}^\mu\frac{\partial\epsilon}{\partial\dot{x}^\nu}.$$

This has indeed the solution $\epsilon \propto (\dot{x}^2)^{-1/2}\xi(\tau)$. The variation of the momenta (expressed in configuration-velocity space variables) becomes

$$\bar{\delta}_\epsilon p_\mu = \frac{\partial p_\mu}{\partial\dot{x}^\nu}\bar{\delta}_\epsilon\dot{x}^\nu = \frac{\partial p_\mu}{\partial\dot{x}^\nu}(-\dot{x}^\nu\epsilon)^{\cdot} = -\frac{\partial p_\mu}{\partial\dot{x}^\nu}\ddot{x}^\nu\epsilon - \frac{\partial p_\mu}{\partial\dot{x}^\nu}\dot{x}^\nu\dot{\epsilon} = [L]_\mu\epsilon - (W_{\mu\nu}\dot{x}^\nu)\dot{\epsilon}.$$

The second term vanishes identically. But the first term does not: The transformation mediated by the first-class constraint is compatible with the Noether symmetry transformation only on-shell. This is typically the case for reparametrization-invariant theories, in contrast to Yang-Mills type theories, the deeper reason being that the Hamiltonian is quadratic in the momenta.

A point of confusion is the seemingly double role of the constraint as a Hamiltonian and as a generator of symmetry transformations. The vanishing of the total Hamiltonian is discussed with terms like "frozen time" or "nothing happens". This attitude comes about from the formal similarity of the two entities by choosing $\xi = (v/2)$. Although both have the same mathematical form, the gauge generator and the Hamiltonian act in different spaces: The gauge generator takes a complete solution and maps a point in the space of solutions to another point in this space, i.e. another solution, while the Hamiltonian takes initial data in this point and maps these to later data in the same point. This issue of time is not specific to the relativistic particle but arises in all generally covariant theories, and has been clarified in [422].

As for gauge fixing, one first verifies that any gauge constraint function Ω necessarily must depend on the parameter τ since it must be conserved (weakly), that is

$$0 \stackrel{!}{\approx} \frac{d\Omega}{d\tau} = \frac{\partial\Omega}{\partial\tau} + \{\Omega, H_T\} = \frac{\partial\Omega}{\partial\tau} + vp^\mu\frac{\partial\Omega}{\partial x^\mu}.$$

A non-vanishing solution for the free parameter v exists only if both $\frac{\partial\Omega}{\partial\tau} \neq 0$ and if $\frac{\partial\Omega}{\partial x^\mu} \neq 0$, at least for one index μ. Of course one can imagine many gauge conditions fulfilling these requests. A choice which eventually leads to simple equations of motions is $\Omega = x^0$-τ, from which $v = 1/p^0$. The ensuing Hamilton function

$$H = \frac{1}{2\,p^0}(p^2 - m^2)$$

in turn entails the simple Hamilton equations of motion $\dot{x}^\mu = p^\mu/p^0$ and $\dot{p}_\mu = 0$.

In order to complete the story, we might construct the Dirac brackets with respect to the two constraints Ω and $H_\perp$. I leave it as a simple exercise to show that the fundamental Dirac brackets become

$$\{x^\mu, x^\nu\}^*_\Delta = 0 \qquad \{p_\mu, p_\nu\}^*_\Delta = 0 \qquad \{x^\mu, p_\nu\}^*_\Delta = \delta^\mu_\nu - \frac{p^\mu}{p^0}\,\delta^0_\nu$$

which indeed allows one to take the constraint as a strong equation: $\{x^\mu, p^2 - m^2\}^*_\Delta = 0$.

Einbein Description

The dynamics of the relativistic particle can also be derived from the Lagrange function

$$L(\dot{x}^\mu, e) = \frac{1}{2e}\eta_{\mu\nu}\dot{x}^\mu\dot{x}^\nu + \frac{m^2}{2}e. \tag{C.77}$$

This Lagrangian contains a further function e, which one may call the einbein[9], because of a resemblance with the notion of a vierbein field, for instance in 4D general relativity. The Euler-Lagrange derivatives are

$$[L]_\mu = -\frac{d}{d\tau}\frac{\partial L}{\partial \dot{x}^\mu} = -\frac{d}{d\tau}\left(\frac{\dot{x}_\mu}{e}\right) \qquad\qquad [L]_e = \frac{\partial L}{\partial e} = -\frac{1}{2e^2}\dot{x}^2 + \frac{1}{2}m^2.$$

Observe that the Euler-Lagrange equation for the einbein is the Lagrangian primary constraint. On-shell with $[L]_e = 0$ holds

$$e \doteq \pm\frac{\sqrt{\dot{x}^2}}{m},$$

and replacing this in the original Lagrangian (C.77), one indeed returns to the description of the relativistic particle with the action (C.73)[10]. Notice that this holds only for $m \neq 0$. And also, be aware that both Lagrangians for the massive relativistic

---

[9] If the relativistic particle is treated in the ADM manner (see C.3.3) as one-dimensional gravity, the e is more appropriately interpreted as the lapse function N.

[10] This substitution is allowed in the sense that it does not change the dynamic content of the model, since the field equation for e itself is used to express e by the other fields. The field e is indeed an auxiliary field; see Sect. 3.3.4.

particles are equivalent only with respect to the classical solutions, but not necessarily equivalent in their quantum version. The Lagrangian (C.77) has the advantage of being free of the nasty square root and of having a zero mass limit. Being quadratic in the velocities, it can be brought into to a path-integral formulation one is accustomed to from scalar and vector field theories.

The Euler derivatives are not independent but obey the identity

$$[L]_\mu \dot{x}^\mu - e\partial_\tau [L]_e \equiv 0.$$

This "smells" like a Noether identity, and indeed it can be rewritten in a form that allows one to read off the infinitesimal transformations that stand behind the symmetry:

$$\big([L]_\mu \dot{x}^\mu + [L]_e \dot{e}\big) - \partial_\tau\big([L]_e e\big) \equiv 0.$$

Compare this with the generic expression (3.85) of the Noether identity to find that for $\delta\tau = \epsilon$ the "fields" transform as $\delta x^\mu = 0$ and $\delta e = -e\dot{\epsilon}$, or

$$\bar{\delta} x^\mu = -\dot{x}^\mu \epsilon, \qquad \bar{\delta} e = -e\dot{\epsilon} - \dot{e}\epsilon = -\partial_\tau(e\epsilon). \tag{C.78}$$

The Lagrangian (C.77) is quasi-invariant with respect to these transformations

$$\begin{aligned}\delta L = \frac{\partial L}{\partial \dot{x}^\mu}\bar{\delta}\dot{x}^\mu + \frac{\partial L}{\partial e}\bar{\delta} e &= -\Big(\frac{\partial L}{\partial \dot{x}^\mu}\ddot{x}^\mu + \frac{\partial L}{\partial e}\dot{e}\Big)\epsilon - \Big(\frac{\partial L}{\partial \dot{x}^\mu}\dot{x}^\mu + \frac{\partial L}{\partial e}e\Big)\dot{\epsilon} \\ &= -\dot{L}\epsilon - \Big(\frac{\dot{x}^2}{e} - \frac{\dot{x}^2}{2e^2}e + \frac{m^2}{2}e\Big)\dot{\epsilon} = -\partial_\tau(L\epsilon).\end{aligned}$$

Thus this model exhibits features of reparametrization invariance considered in the previous subsection. But observe that the Lagrangian (C.77) is not homogeneous of first degree, and thus we cannot infer immediately that the canonical Hamiltonian vanishes. But we will see shortly that the total Hamiltonian vanishes weakly, since it becomes a sum of first-class constraints.

The momenta canonically conjugate to x^μ and e are

$$p_\mu := \frac{\partial L}{\partial \dot{x}^\mu} = \eta_{\mu\nu}\frac{\dot{x}^\nu}{e}, \qquad P := \frac{\partial L}{\partial \dot{e}} = 0. \tag{C.79}$$

The latter is the primary constraint we are inclined to expect from the invariance related to (C.78). The canonical Hamiltonian is

$$H_C = \dot{x}^\mu p_\mu + \dot{e}P - L = ep^2 - \frac{1}{2}\frac{e^2\, p^2}{e} - \frac{1}{2}m^2 e = \frac{1}{2}e(p^2 - m^2),$$

and thus

$$H_T = \frac{1}{2}e(p^2 - m^2) + uP. \tag{C.80}$$

The requirement of time conservation of the primary constraint leads to

$$0 \overset{!}{\approx} \{P, H_T\} = -\frac{1}{2}(p^2 - m^2) := -H_\perp .$$

This is a secondary constraint–and in accordance with the general findings–the phase-space version of the primary Lagrangian constraint $[L]_e = 0$. Actually it is identical to the primary Hamiltonian constraint derived from the action (C.73). Both constraints are first class, so that in the end the total Hamiltonian (C.80) is a linear combination of first-class constraints. The equations of motion induced by this Hamiltonian are

$$\dot{x}^\mu = ep^\mu, \qquad \dot{p}_\mu = 0, \qquad \dot{e} = u, \qquad \dot{P} = -\frac{1}{2}(p^2 - m^2) \approx 0.$$

The equation of motion for x^μ reproduces the momentum in (C.79), the equation of motion for the einbein e determines the multiplier u.
Both constraints enter the symmetry generator

$$G_\zeta = H_\perp \zeta + P\dot{\zeta}.$$

As in the situation of the worldline description of the relativistic particle, the phase-space symmetry generator can be used to generate Noether symmetries only for a specific choice of descriptors ζ. Namely from

$$\delta_\zeta x^\mu = \{x^\mu, G_\zeta\} = \eta^{\mu\nu} p_\nu \zeta, \qquad \delta_\zeta e = \{e, G_\zeta\} = \dot{\zeta}$$

we get only the correspondence with (C.78) if $\zeta = -e\epsilon$. Again, the descriptor in the symmetry generator depends on fields, here on the einbein e. And again this directly follows from the demand for projectability, which in this case amounts to

$$0 \overset{!}{=} \frac{\partial}{\partial \dot{e}} \bar{\delta} e = -e \frac{\partial}{\partial \dot{e}} \dot{\epsilon} - \epsilon - \dot{e} \frac{\partial}{\partial \dot{e}} \epsilon,$$

and which indeed has the solution $\epsilon = \xi(\tau)/e$. The variation of the momenta vanish on-shell, i.e. either on solutions or on the constraint surface.

In order to complete the story, let us investigate the invariance of the action

$$S_T = \int d\tau \Big(p_\mu \dot{x}^\mu + P\dot{e} - \frac{e}{2}(p^2 - m^2) - u\Phi_1 \Big)$$

under the transformations canonically resulting from the generator $\delta_{\underline{\epsilon}}(-) = \{-, \Phi_I\}\underline{\epsilon}^I$ with the two first class constraints $\Phi_1 = P$ and $\Phi_2 = 1/2(p^2 - m^2)$. Following the argumentation after (C.31), the condition (C.32) reads in this case

$$0 \overset{!}{=} \dot{\epsilon}^1 \Phi_1 + \dot{\epsilon}^2 \Phi_2 - \epsilon^1 \Phi_2 - \delta_\epsilon u \; \Phi_1 \ ,$$

and thus $\delta_{\underline{\epsilon}} u = \dot{\underline{\epsilon}}^1$ and $\dot{\underline{\epsilon}}^2 = \underline{\epsilon}^1$, this being consistent with previous results.

Supergravity in One Dimension

Off course the Hamiltonian approach can (and has) also been applied to supersymmetric theories. Although this is not the place to display for example the canonical treatment of generic (or even specific) supergravity theories, we will be able to identify some typical features of Hamiltonian supersymmetry with one-dimensional models. "Supergravity in one dimension" has also been titled "The spinning classical particle", although the spin is commonly understood as an outgrowth of relativistic quantum physics. This section will be a bit sketchy, since otherwise it could become a book of its own.

"Just for fun", consider the following Lagrange function

$$L = \frac{1}{2}\Big\{\frac{\dot{x}^2}{e} - i\psi_\mu\dot{\psi}^\mu - \frac{2i}{e}\chi(\psi_\mu\dot{x}^\mu)\Big\} \tag{C.81}$$

in terms of commuting "fields" x^μ, e and anti-commuting fields ψ^μ, χ. Instead of dealing immediately with the (quasi-)invariances of the Lagrangian (which you most probably would not be able to guess) we turn directly to the phase space and discover the invariances from the constraints. The momenta are[11]

$$p_\mu := \frac{\partial L}{\partial \dot{x}^\mu} = \frac{1}{e}(\dot{x}^\mu - i\chi\psi_\mu) \qquad\qquad p_e := \frac{\partial L}{\partial \dot{e}} = 0$$
$$\pi_\mu := \frac{\partial L}{\partial \dot{\psi}^\mu} = \frac{i}{2}\psi_\mu \qquad\qquad \pi_\chi := \frac{\partial L}{\partial \dot{\chi}} = 0.$$

From this, one reads off the primary constraints

$$\phi_e := p_e \approx 0, \qquad \phi_\mu := \pi_\mu - \frac{i}{2}\psi_\mu \approx 0, \qquad \phi_\chi := \pi_\chi \approx 0.$$

The canonical Hamiltonian and the total Hamiltonian are found to be

$$H_C = \frac{e}{2}p^2 + i\chi(p_\mu\psi^\mu) \qquad H_T = H_C + u\phi_e + v^\mu\phi_\mu + v\phi_\chi$$

with an even multiplier u and odd multipliers (v^μ, v). According to the Rosenfeld-Dirac-Bergmann procedure the time (τ) derivatives of the primary constraints are required to vanish weakly. In order to implement these requirements, calculate the Poisson brackets among the constraints, using the canonical Bose-Fermi Poisson brackets

$$\{x^\mu, p_\nu\} = \delta^\mu_\nu \qquad \{e, p_e\} = 1 \qquad \{\psi^\mu, \pi_\nu\} = -\delta^\mu_\nu \qquad \{\chi, \pi\} = -1$$

–all others vanishing. From these, we find the only non-vanishing brackets among the primary constraints: $\{\phi_\mu, \phi_\nu\} = i\eta_{\mu\nu}$. Next, require

[11] Observe that all over this book, derivatives are taken "from the right" in the sense of Appendix B.2.3.

$$
\begin{aligned}
0 &\overset{!}{\approx} \{\phi_e, H_T\} = \{p_e, H_C\} = -\frac{1}{2}p^2 =: \hat{\phi} \\
0 &\overset{!}{\approx} \{\phi_\mu, H_T\} = \{\phi_\mu, i\chi(p\psi)\} - \{\phi_\mu, \phi_\nu\}v^\nu = ip^\mu\chi - iv^\mu \\
0 &\overset{!}{\approx} \{\phi_\chi, H_T\} = \{\pi_\chi, H_C\} = i(p\psi) =: \bar{\phi}.
\end{aligned}
$$

The first and the last requirement result in two secondary constraint functions $\hat{\phi}$ and $\bar{\phi}$. The further requirements lead to a determination of the multipliers v^μ as $v^\mu = p^\mu\chi$. Tertiary constraints do not arise since $\{\hat{\phi}, H_T\} = 0$ and $\{\bar{\phi}, H_T\} = i\{p\psi, v^\mu\pi_\nu\} = ip_\mu v^\mu = ip^2\chi \approx 0$.

The constraints $\phi_e, \phi_\chi, \hat{\phi}$ are immediately recognized as first-class constraints. The constraint $\bar{\phi}$ is not first class, since $\{\bar{\phi}, \phi_\mu\} = ip_\mu$. However, the linear combination of second-class constraints

$$
\Omega := \bar{\phi} + p^\mu\phi_\mu
$$

is first class: Its Poisson brackets with $\phi_e, \phi_\chi, \hat{\phi}$ vanish and

$$
\begin{aligned}
\{\Omega, \phi_\mu\} &= ip^\nu\{\psi_\nu, \phi_\mu\} + p^\nu\{\phi_\nu, \phi_\mu\} = ip^\nu\{\psi_\nu, \pi_\mu\} + ip_\mu = 0 \\
\{\Omega, \Omega\} &= \{ip\psi, p\phi\} - \{p\phi, ip\psi\} + \{p\phi, p\phi\} \propto p^2 \approx 0.
\end{aligned}
$$

With this, the total Hamiltonian can be written as a sum of first-class constraints:

$$
H_T = -e\hat{\phi} + \chi\Omega + u\phi_e + v\phi_\chi.
$$

Dirac brackets are built with respect to the second-class constraints. The matrix inverse to $(\Delta) = \{\phi_\mu, \phi_\nu\}$ is $(\Delta^{-1}) = -i\eta_{\mu\nu}$ and therefore the Dirac brackets are

$$
\{A, B\}^*_\Delta = \{A, B\} + i\{A, \phi_\lambda\}\{\phi^\lambda, B\},
$$

specifically

$$
\{A, H_T\}^*_\Delta = \{A, H_T\} + i\{A, \phi_\lambda\}\{\phi^\lambda, H_T\} = \{A, H_T\}.
$$

Since the only fields which have non-vanishing Poisson brackets with the second-class constraints ϕ_μ are ψ^μ and π_ν, the sole changes in the canonical brackets are

$$
\{\psi_\mu, \psi_\nu\}^*_\Delta = i\eta_{\mu\nu} \qquad \{\psi_\mu, \pi_\nu\}^*_\Delta = -\frac{1}{2}\eta_{\mu\nu} \qquad \{\pi_\mu, \pi_\nu\}^*_\Delta = -\frac{i}{4}\eta_{\mu\nu}.
$$

The Hamilton equations of motion derived from $\dot{F} = \{F, H_T\}^*$ read

$$\dot{x}^\mu = ep^\mu + i\chi\psi^\mu \qquad \dot{p}_\mu = 0 \tag{C.82a}$$
$$\dot{e} = u \qquad \dot{p}_e \approx 0 \tag{C.82b}$$
$$\dot{\psi}^\mu = v^\mu = \chi p^\mu \qquad \dot{\pi}_\mu = (i/2)v^\mu = (i/2)\chi p^\mu \tag{C.82c}$$
$$\dot{\chi} = v \qquad \dot{\pi}_\chi \approx 0. \tag{C.82d}$$

The appearance of two primary first class constraints indicates the presence of symmetries. Since the PFC ϕ_e is even and ϕ_χ is odd, we expect that each of them takes part in generating even and odd symmetry transformations, respectively. Thus we will go through the gauge generating algorithm (C.40) for the two primary first-class constraints separately.

In starting with the PFC $\phi_e = p_e$, we observe from $\{\phi_e, H_T\} = \hat{\phi}$ that this does not yield a PC. Therefore ϕ_e is unsuited for being the generator component G_0. Thus we start the algorithm with $G_1 = \phi_e$ and in the next step we determine $G_0 = -\{\phi_e, H_T\} + PC = -\hat{\phi} + PC$. Here, PC could be any linear combination of primary constraints. However, since $\{\hat{\phi}, H_T\} = 0$ the simplest solution [12] is $G_0 = -\hat{\phi}$. We thus have

$$G_\zeta = \dot{\zeta}\; G_1 + \zeta\; G_0 = \dot{\zeta}\; p_e + \zeta\; \frac{1}{2}p^2.$$

These generate transformations $\delta_\zeta F = \{F, G_\zeta\}$:

$$\begin{aligned}&\delta_\zeta x^\mu = \zeta p^\mu, \qquad \delta_\zeta e = \dot{\zeta}, \qquad \delta_\zeta \psi^\mu = 0 = \delta_\zeta \chi\\ &\delta_\zeta p_\mu = \delta_\zeta p_e = \delta_\zeta \pi_\mu = \delta_\zeta \pi_\chi = 0.\end{aligned} \tag{C.83}$$

As for the fermionic symmetry, we start the algorithm by choosing the PFC $G_1 = \pi_\chi$. Then $G_0 = -\{G_1, H_C\} + PC = \bar{\phi} + PC$. Since $\bar{\phi}$ itself does not obey (C.40c) we are forced to add a PC part, tentatively of the form $G_0 = \bar{\phi} + w^\mu \phi_\mu + \alpha p_e + \beta p_\chi$ with bosonic parameter/fields w^μ, β and fermionic parameter α. In order that this G_0 satifies (C.40c) the parameter become determined and $G_0 = \Omega + 2i\chi p_e$. Therefore

$$G_\theta = \dot{\theta}\; G_1 + \theta\; G_0 = \dot{\theta}\; \pi_\chi + \theta\; (\Omega + 2i\chi p_e),$$

generating

$$\begin{aligned}&\delta_\theta x^\mu = i\theta\psi^\mu, \qquad \delta_\theta e = 2i\theta\chi, \qquad \delta_\theta \psi^\mu = \theta p^\mu, \qquad \delta_\theta \chi = \dot{\theta}\\ &\delta_\theta p_\mu = \delta_\theta p_e = \delta_\theta \pi_\chi = 0, \qquad \delta_\theta \pi_\mu = \frac{i}{2}\theta p_\mu.\end{aligned} \tag{C.84}$$

[12] Indeed, one could add the primary constraint $(\psi\phi + \chi\pi_\chi)$ which at the very end modifies the ensuing infinitesimal transformations by on-shell terms.

The Lagrange function (C.81) is indeed quasi-invariant under the transformations (C.84):

$$\delta_\zeta L = \frac{d}{d\tau}\Big(\frac{\zeta}{e}(L + \frac{i}{2}\psi_\mu\dot{\psi}^\mu)\Big) \tag{C.85}$$

and under the transformations (C.84) as

$$\delta_\theta L = \frac{d}{d\tau}\Big(\frac{\theta}{e}(\frac{i}{2}\psi_\mu\dot{x}^\mu)\Big).$$

Noticing that in the absence of fermionic fields, the action (C.81) reduces to the action of a massless relativistic particle, and that this action is reparametrization-invariant under the transformations $\bar{\delta}_\epsilon x^\mu = -\epsilon\dot{x}^\mu$ and $\bar{\delta}_\epsilon e = -(\dot{\epsilon e})$, we are tempted to recover these again. We observe that

$$\delta_\zeta x^\mu = \zeta p^\mu = \frac{\zeta}{e}\dot{x}^\mu - i\zeta\chi\psi^\mu.$$

In identifying $\epsilon = -\zeta/e$, the first term is recognized as a reparametrization. The second term is a supersymmetry transformation with $\theta = \zeta\chi$. This suggests that we interpret the δ_ζ-transformations as a mixture of a genuine reparametrization $\bar{\delta}_\epsilon$ with a specific supersymmetry transformation $\delta_{\theta(\epsilon)}$:

$$\hat{\delta}_\epsilon := \delta_{(\zeta=-e\epsilon)} + \delta_{(\theta=-e\chi\epsilon)}. \tag{C.86}$$

This results in

$$\delta_\epsilon x^\mu = -\epsilon\dot{x}^\mu \quad \delta_\epsilon e = -\epsilon\dot{e} - e\dot{\epsilon} \quad \delta_\epsilon\psi^\mu = -\chi\epsilon p^\mu \simeq -\epsilon\dot{\psi}^\mu \quad \delta_\epsilon\chi = -\epsilon\dot{\chi} - \chi\dot{\epsilon}.$$

Here the $\simeq$ symbol has the meaning "being the same on the solutions of the Hamilton equations of motion" (C.82).

In conclusion, I want to stress, that we found the symmetry transformations that leave the Lagrangian (C.81) quasi-invariant by analyzing the Hamiltonian constraints. A proper mixing of the symmetry transformations provided by the canonical generators results in symmetry transformations of the Lagrangian. The mixing of different transformations is typical of any generally covariant theory with additional local symmetries. The mathematical reason behind this is the substitution of Lie derivatives by group-covariant Lie derivatives as explained in Appendix F.4. This mixing of symmetries will be seen again in the case of coupling a Yang-Mills theory to the gravitational field.

The reparametrizations and local supersymmetry transformations form a group: Denoting $\tilde{\delta}_j\mathrm{F} = \hat{\delta}_{\epsilon_j}\mathrm{F} + \delta_{\theta_j}\mathrm{F}$, one finds for all fields $\mathrm{F} = \{x^\mu, e, \psi^\mu, \chi\}$ that for two transformations $\tilde{\delta}_1$ and $\tilde{\delta}_2$ the commutators $(\tilde{\delta}_1\tilde{\delta}_2 - \tilde{\delta}_2\tilde{\delta}_1)\mathrm{F} = \tilde{\delta}_3\mathrm{F}$ obey

$$\epsilon_3 = (\dot{\epsilon}_1\epsilon_2 - \dot{\epsilon}_2\epsilon_1) + \frac{2i}{e}\theta_1\theta_2 \qquad \chi_3 = (\dot{\theta}_2\epsilon_1 - \dot{\theta}_1\epsilon_2) + \frac{2i\chi}{e}\theta_1\theta_2. \tag{C.87}$$

The Lagrangian (C.81) is more than just a toy example and exhibits many features of supersymmetric theories. For instance, the commutator (C.87) reveals that the product of two supersymmetry transformations makes up for a reparametrization (i.e. a coordinate transformation). It also exhibits the fact that for reparametrization-invariant theories, we cannot avoid having field-dependent structure functions (here depending on e and χ). The example also illustrates that in order to describe "classical spin", it does not suffice to include just one fermionic field ψ^μ into the Lagrangian (C.81) for the relativistic particle. An additional field (χ) is needed in order to obtain a group of symmetry transformations. In the late 1970's, other descriptions for "supergravity in one dimension " were discussed, also for the massive case. The underlying Lagrangians although being related to (C.81), suffered from being quasi-invariant only "on-shell" (i.e. on the solutions of the equations of motion), or from requiring *ad hoc* assumptions in order to assure congruity of the Lagrangian and Hamiltonian description; see e.g. [492].

If aside from the coordinate τ, odd coordinates θ are introduced, one can write down a compact action for "superfields" $X^\mu(z)$ and $E(z)$ which is invariant under reparametrizations in the superspace in terms of coordinates $z = (\tau, \theta)$; see for instance Sect. 3.7 in [232].

The relativistic (super)particle was here only treated classically. A thorough description of its quantization is given in [270]. The most advanced quantization procedure by BRST techniques can be found in [517].

C.3.3 Metric Gravity

In this subsection, only vacuum metric gravity is treated. Some remarks on both tetrad gravity and on Yang-Mills and Dirac fields coupled to the gravitational field are made in the following subsections.

P.A.M. Dirac was the first to find the Hamiltonian for vacuum metric gravity [128]. His tedious derivation obtained a more streamlined form with the (3+1)-decomposition, today known by the acronym ADM for its three authors R. Arnowitt, S. Deser, and C.W. Misner [15]. Also its coordinate-free variant was investigated [290]. For years, this was the established version, until A. Ashtekar discovered a set of fields by which the Hamiltonian structure of general relativity acquires a form similar to that of Yang-Mills gauge theories [16]. These "new" variables are the ones preferred today, also for loop quantum gravity. A rather novel formulation of general relativity, not treated here, is *shape dynamics* [231]. It relates to the ADM formulation, but replaces the refoliation invariance in the (3+1)-split with local spatial conformal invariance. This approach seems to be more amenable to quantization because of a better suitable phase-space structure.

In the following I will, contrary to the conventions at other places of this book, use the "mostly plus" metric diag $g = (-1, +1, +1, +1)$ because it is the one made use of in the influential ADM paper.

The 3 + 1 Decomposition

Arnowitt, Deser and Misner started with a 3+1 split of the metric:

$$g_{\mu\nu}dx^\mu dx^\nu = -N^2 dt^2 + g_{ij}(dx^i + N^i dt)(dx^j + N^j dt)$$

i.e.

$$(g_{\mu\nu}) = \begin{pmatrix} -N^2 + N^i N^j g_{ij} & g_{ij}N^j \\ g_{ij}N^j & g_{ij} \end{pmatrix} \tag{C.88}$$

where the indices $i, j, \dots$ run from 1 to 3. The objects N and N^i are called the *lapse function* and the *shift functions*, respectively. The geometric interpretation is as follows: Fix a hypersurface $x^0 = \text{const}$ and call it Σ. In the coordinates x^i, the three-dimensional metric h_{ij} induced on Σ coincides with the three-dimensional part of the four-dimensional metric, namely $h_{ij} = g_{ij}$. Let h^{ij} be the inverse to the metric on Σ:

$$h_{ij}h^{jk} = \delta_i^j$$

from which

$$h^{ij} = g^{ij} - \frac{g^{0i}g^{0j}}{g^{00}}, \qquad \text{or explicitly} \qquad (g^{\mu\nu}) = \begin{pmatrix} -1/N^2 & N^j/N^2 \\ N^i/N^2 & h^{ij} - N^i N^j/N^2 \end{pmatrix}.$$

The lapse and the shift functions are thus

$$N := \frac{1}{\sqrt{-g^{00}}} \qquad\qquad N^i := -\frac{g^{0i}}{g^{00}}. \tag{C.89}$$

We notice further $\sqrt{-g} = N\sqrt{h}$, where $g := \det(g_{\mu\nu})$, $h := \det(h_{ij})$. The idea is to use the three metric, and the lapse and the shift as "configuration" variables. As we will see, the canonically conjugate momenta to the three-metric are related to the extrinsic curvature of the 3-dimensional hypersurface Σ.

In choosing coordinates ξ^i for the hyper-surface Σ, the metric h_{ij} induced from the spacetime metric $g^{\mu\nu}$ is

$$h_{ij} = g_{\mu\nu}\frac{\partial x^\mu}{\partial \xi^i}\frac{\partial x^\nu}{\partial \xi^j} := g_{\mu\nu}\, x_i^\mu\, x_j^\nu.$$

To regain the full geometry we need to add a normal vector n^μ which fulfills

$$g_{\mu\nu}n^\mu\, n^\nu = -1, \qquad g_{\mu\nu}n^\mu x^\nu = 0,$$

to the three tangent vectors x_i^μ. Now it can be shown that

$$(n^\mu) = -\frac{(g^{0\mu})}{\sqrt{(-g^{00})}} = \frac{1}{N}(1, -N^i). \tag{C.90}$$

The embedding of Σ in the original 4-dimensional space is characterized by the extrinsic curvature

$$K_{ij} = \frac{1}{2N}\Big(N_{i|j} + N_{j|i} - \partial_0 h_{ij}\Big), \tag{C.91}$$

where the slash stands for the covariant derivative in the 3-metric. According to the Gauss-Codazzi equations[13] the Riemann curvature tensors of the full and the embedded space are

$$\begin{aligned} {}^{(4)}R_{ijkl} &= {}^{(3)}R_{ijkl} + K_{ik}K_{jl} - K_{il}K_{jk} \\ {}^{(4)}R_{0jkl} &= K_{jl|k} - K_{jk|l}. \end{aligned}$$

This already suffices to express the Hilbert Lagrangian within the 3+1-split. The result is:

$$\mathcal{L} = \frac{1}{2\kappa}\sqrt{-g}\;{}^{(4)}R = \frac{1}{2\kappa}\,Nh^{1/2}\big({}^{(3)}R + K_{ij}K^{ij} - K^2\big) + \text{b.t..} \tag{C.92}$$

This is the starting point for the Hamiltonian approach by the Rosenfeld-Dirac-Bergmann algorithm[14].

The Hamiltonian for the Hilbert-Einstein Action

The Lagrange density (C.92) is a functional of the six metric components h_{ij} and the lapse and the shift functions (N, N^i) which are contained in the extrinsic curvature (C.91). Let us use as a short-hand notation $(N^\alpha) = (N, N^i)$ (while remaining aware that this is not a 4-vector). Their canonical momenta are

$$P_\alpha := \frac{\partial\mathcal{L}}{\partial(\partial_0 N^\alpha)} = 0. \tag{C.93}$$

The canonical momenta with respect to h_{ij} are

$$p^{ij} := \frac{\partial\mathcal{L}}{\partial(\partial_0 h_{ij})} = -\frac{1}{2\kappa}h^{1/2}\big(K^{ij} - Kh^{ij}\big). \tag{C.94}$$

---

[13] For details see e.g. [236] or Appendix E of [491].

[14] The 3+1 decomposition is of genuine importance also in other cases and specifically for numerical relativity [236].

This expression can be solved for the K^{ij}, giving

$$K^{ij} = -2\kappa\, h^{-1/2}\left(p^{ij} - \frac{1}{2} p h^{ij}\right),$$

where $p := p^{ij} h_{ij}$. And, recalling (C.91), the "velocities" $\partial_0 h_{ij}$ can therefore be expressed by the momenta

$$h_{ij,0} = -4\kappa\, N h^{-1/2}\left(p^{ij} - \frac{1}{2} p h^{ij}\right) - (N_{i|j} + N_{j|i}).$$

Thus there are the four primary constraints

$$\phi_\alpha := P_\alpha \approx 0. \tag{C.95}$$

Calculating the canonical Hamiltonian density (and after a further partial integration and disregarding a surface term) one eventually obtains

$$\begin{aligned} \mathcal{H}_C &= N\mathcal{H}_\perp + N^i \mathcal{H}_i \\ &= N\left[(2\kappa)\, h^{-1/2}(p^{ij} p_{ij} - \frac{1}{2} p^2) - \frac{h^{1/2}}{2\kappa}\ {}^{(3)}R\right] + N^i\left[-2 p_i{}^j{}_{|j}\right]. \end{aligned} \tag{C.96}$$

The total Hamiltonian is

$$H_T = H_C + \int d^3x\ v^\alpha P_\alpha$$

with multipliers v^α. The consistency (or stabilization) condition $\{P_\alpha,\ H_T\} \overset{!}{\approx} 0$ based on the canonical Poisson brackets

$$\begin{aligned} \{N^\alpha(x),\ P_\beta(y)\}_{x^o=y^o} &= \delta^\alpha_\beta \delta(\vec{x}, \vec{y}) \\ \{h_{ij}(x), p^{kl}(y)\ \}_{x^o=y^o} &= \frac{1}{2}(\delta^k_i \delta^l_j + \delta^l_i \delta^k_j)\delta(\vec{x}, \vec{y}) \end{aligned}$$

yields the secondary constraints

$$\mathcal{H}_\perp \approx 0 \qquad\qquad \mathcal{H}_i \approx 0, \tag{C.97}$$

known as the *Hamiltonian constraint* $\mathcal{H}_\perp$ and the *momentum constraints* $\mathcal{H}_i$. There are no further constraints since the Poisson brackets of these constraints vanish on the constraint surface due to the Poisson brackets

$$
\begin{aligned}
\{\mathcal{H}_\perp(x),\ \mathcal{H}_\perp(y)\} &= -\sigma[h^{ij}\mathcal{H}_j(x) + h^{ij}\mathcal{H}_j(y)]\partial_i\delta(x,y) \\
\{\mathcal{H}_i(x),\ \mathcal{H}_\perp(y)\} &= \mathcal{H}_\perp\partial_i\delta(x,y) \\
\{\mathcal{H}_i(x),\ \mathcal{H}_j(y)\} &= [\mathcal{H}_j(x)\partial_i + \mathcal{H}_i(y)\partial_j]\delta(x,y).
\end{aligned}
\tag{C.98}
$$

The factor σ in the first Poisson bracket is introduced to show the dependence of the constraint algebra on the metric signature $\mathrm{diag}(\sigma, +1, +1, +1)$. Certainly, in GR we are dealing with the Lorentzian case $\sigma = -1$, but we will see later in the Ashtekar version of canonical GR how one can "play" with this signature together with another parameter[15].

The Hamiltonian is–up to boundary terms–a linear combination of constraints, and this can be shown to be typical for all reparametrization-invariant theories. We saw it already for the relativistic particle, other examples are the relativistic string and–amazingly–any theory derivable from an action if the coordinates are treated as configuration space variables. In all these cases the Hamiltonian can be written in the form

$$
H = \int d^{D-1}x\ (N^\alpha\mathcal{H}_\alpha + v^\alpha P_\alpha) + b.t.,
$$

where $\mathcal{H}_\alpha = (\mathcal{H}_\perp, \mathcal{H}_i)\ \ i = 1, \ldots, D$-1. It is remarkable, that although the explicit form of the $\mathcal{H}_\alpha$ is different in the diverse examples, they always obey the algebra (C.98). This universal structure of the Hamiltonian and the constraint algebra reflect the criteria to embed a spatial surface within the spacetime manifold; that is, they are valid if the theory is consistent with the foliation of spacetime into arbitrary spacelike hypersurfaces as shown by S. A. Hojman, K. Kuchař, and C. Teitelboim [278], see also [497]. Be aware that the structure of the Hamiltonian and the constraint algebra holds true for any diffeomorphism-invariant theory of gravity formulated in a Riemann space, that is specifically for many modifications and extensions of GR described in Sect. 7.6. The structural changes that arise in Riemann-Cartan spaces are handled below in Subsect. C.3.4 on tetrad gravity.

It is definitely crucial that (C.98) is not the algebra of the diffeomorphism group (see Appendix A.2.6). Only the spatial constraints form the algebra (A.20). The Poisson brackets among the constraints $\mathcal{H}_\perp$ in (C.98) reveal that the structure coefficient depends on the (D-1)-metric. This is one of the reasons why the canonical quantization program of gravity ran aground. Only for specific cases in which gravity is coupled to (macroscopic) matter it is known that by forming new constraint functions from the $\mathcal{H}_\alpha$ it is possible to get rid of the metric-dependence in the algebra. For instance in the original dust model of Brown and Kuchař [61] the weight-two scalar combination $\mathcal{G} = \mathcal{H}_\perp^2 - h^{ij}\mathcal{H}_i\mathcal{H}_j$ has strongly vanishing Poisson brackets with itself. The full algebra corresponds to the semidirect product of the Abelian algebra generated by $\mathcal{G}$ and the algebra of spatial diffeomorphisms generated by $\mathcal{H}_i$. The

[15] The algebra (C.98) is defined for other values of σ as well. It considerably simplifies for $\sigma = 0$; this corresponds to the limit in which the speed of light is zero, characteristic for the Carroll kinematical group; see Subsect. 3.4.5.

reason behind introducing matter/dust is that it serves as a material reference system which breaks the diffeomorphism invariance and de-parametrizes the theory.

For later purposes the constraint algebra (C.98) is compactly denoted as

$$\{\mathcal{H}_\alpha(x), \mathcal{H}_\beta(y)\}_{x^0=y^0} = \int d^{D-1}z\; C^{\gamma}_{\alpha\beta}(x,y,z)\mathcal{H}_\gamma$$

or, for short

$$\{\mathcal{H}_\alpha, \mathcal{H}_{\beta'}\} = C^{\gamma''}_{\alpha\beta'}\mathcal{H}_{\gamma''}. \tag{C.99}$$

Gauge Generators and Observables

Already at the example of the relativistic particle, it was mentioned that the symmetry transformations in the cotangent space are a restricted subset of the symmetry transformations in the tangent space. As pointed out by P. Bergmann and A. Komar [44] the Hilbert-Einstein Lagrangian is quasi-invariant not only with respect to diffeomorphism (locally: $\hat{x}^\mu = \hat{x}^\mu(x)$) but to the even larger group **Q** with transformations of the form $\hat{x}^\mu = \hat{x}^\mu[x, g_{\rho\sigma}]$. In addition, they showed that in phase space, these generically field-dependent functionals must be restricted. In its infinitesimal form $\hat{x}^\mu = x^\mu + \epsilon[x, g_{\rho\sigma}]$, this restriction can be expressed as a condition of the functional dependence of ϵ on the fields as

$$\epsilon^\mu[x, g_{\rho\sigma}] = n^\mu\xi^0 + \delta^\mu_i\xi^i \qquad \text{where} \qquad \xi^\alpha = \xi^\alpha[g_{ij}, K^{ij}]. \tag{C.100}$$

And it has been shown in [462] that only these restricted symmetry transformations can be realized as canonical transformations in phase space. Namely, for this realizability the commutator $[\delta_1, \delta_2]x^\mu$ of two symmetry transformations $\delta_i x^\mu = -\epsilon^\mu_i(x, g(x))$ must not depend on time derivatives of the ϵ_i. As shown in [423] the deeper reason is again the projectability issue[16]. The null vectors of the Hessian for the Hilbert-Einstein action (expressed in the 3+1 decomposition) are spanned by

$$\Gamma_\lambda = \frac{\partial}{\partial \dot{N}^\lambda}.$$

The condition $\Gamma_\lambda(\delta_\epsilon g_{\mu\nu}) = 0$ leads to a set of differential equations for $\epsilon[x, g_{\rho\sigma}]$ having (C.100) as solution. Thus the transformations must depend explicitly on the lapse and shift function via the normal vector (C.90). Further the descriptors ξ^α are allowed to depend only on the three-geometry or the $(D\text{-}1)$ geometry, respectively. We saw this kind of transformation already with the relativistic particle: There the "normal" vector has only one component, namely $n^0 = 1/N$ and (C.100) simplifies to $\epsilon(x, N) = (1/N)\xi(x)$, where N corresponds to the "einbein" e. The transformations (C.100) form the Bergmann-Komar group **BK**. This is a subgroup of **Q**

[16] In another guise, this was disclosed also in [338].

as is **Diff**(**M**). However, **Diff**(**M**) is not a subgroup of **BK**! I point out here that the "gauge choice" ($N = 1, N^i = 0$), very often seen in the literature, blurs the distinction between these symmetry groups.

The symmetry generator for the Bergmann-Komar transformations can be derived by the algorithm (C.40) and is, according to [81] and [423], given by

$$G_\xi = -\big(\mathcal{H}_\alpha + N^{\gamma''} C^{\beta'}_{\alpha\gamma''} P_{\beta'}\big)\xi^\alpha - P_\alpha \dot{\xi}^\alpha, \tag{C.101}$$

where

$$\dot{\xi}^\alpha = \frac{d}{dt}\xi^\alpha = \frac{\partial}{\partial t}\xi^\alpha + N^{\beta'}\{\xi^\alpha, \mathcal{H}_{\beta'}\}.$$

Observe that the lapse and the shift functions are integral building blocks in this generator–appearing in a specific combination of the first-class constraints. Observables $\mathcal{O}$ are those objects which have weakly vanishing Poisson brackets with the symmetry generators: $\{\mathcal{O}, G_\xi\} \approx 0$. The problem of determining observables in general relativity is as old as GR itself. Only for some spacetimes with special asymptotic behavior or additional Killing symmetries have observables been constructed explicitly. Nevertheless, one can establish a formal expression to derive and prove their properties, like their equations of motion, the relation of their Poisson brackets to Dirac brackets, etc. This was done in [426] with the aid of the symmetry generator (C.101) together with "gauge" conditions[17]. Solutions of the field equations which are related by a symmetry operation can only be discriminated by coordinate conditions in the tangent space or by "gauge" conditions χ^α in the cotangent space. These are (locally) unique and complete if they explicitly depend on coordinates, e.g.

$$\chi^\alpha = x^\alpha - X^\alpha, \tag{C.102}$$

where each of the X^α is a spacetime scalar and one must require $\det(\{X^\alpha, \mathcal{H}_\beta\}) \neq 0$. A formal expression for the observable adjoined to a phase space quantity Φ is obtained through a transformation to coordinates intrinsically defined by (C.102):

$$\mathcal{O}_\Phi = \exp(\{-, G_{\bar{\xi}}\})\Phi|_{\bar{\xi}=\chi}. \tag{C.103}$$

Next, introduce the nonsingular matrix $\mathcal{A}^\alpha{}_\beta := \{X^\alpha, \mathcal{H}_\beta\}$ with its inverse $\mathcal{B}^\alpha{}_\beta := (\mathcal{A}^{-1})^\alpha{}_\beta$. (It might technically be hard or even impossible to find the inverse, but remember that we are seeking a formal definition.) Notice that the requirement $\dot{\chi} \approx 0 = \delta^\alpha_0 - \mathcal{A}^\alpha{}_\beta N^\beta$ determines the lapse and shift: $N^\alpha = \mathcal{B}^\alpha{}_0$. Now define

$$\bar{\mathcal{H}}_\alpha := \mathcal{B}^\beta{}_\alpha\Big(\mathcal{H}_\beta - \mathcal{B}^\gamma{}_\delta N^\epsilon\{\mathcal{A}^\delta{}_\epsilon, \mathcal{H}_\beta\}P_\gamma\Big),$$
$$\bar{P}_\alpha := \mathcal{B}^\beta{}_\alpha P_\beta, \qquad G_{\bar{\xi}} = \bar{\mathcal{H}}_\alpha\bar{\xi}^\alpha + \bar{P}_\alpha\dot{\bar{\xi}}^\alpha.$$

[17] I use quotation marks here because the Bergmann-Komar group is not the gauge group of GR in the sense of a Yang-Mills gauge theory.

Then for any phase-space function Φ, the observable in the gauge χ is

$$\mathcal{O}_\Phi[\chi] \approx \exp(\{-, G_{\bar{\xi}}\})\Phi_{|\bar{\xi}\to\chi} = \Phi + \{\Phi, G_{\bar{\xi}}\}_{|\bar{\xi}\to\chi} + \frac{1}{2!}\{\{\Phi, G_{\bar{\xi}}\}, G_{\bar{\xi}}\}_{|\bar{\xi}\to\chi} + \ldots.$$

For phase-space functions not depending on the lapse and the shift, this expression can be written as

$$\mathcal{O}_\Phi[\chi] = \Phi + \chi^\alpha\{\Phi, \overline{\mathcal{H}}_\alpha\} + \chi^\alpha\chi^\beta\frac{1}{2!}\{\{\Phi, \overline{\mathcal{H}}_\alpha\}, \overline{\mathcal{H}}_\beta\} + \ldots.$$

or in a compact notation

$$\mathcal{O}_\Phi[\chi] =: \sum_{k=0}^{\infty}\frac{1}{k!}\chi^{(k)}\{\Phi, \bar{\mathcal{H}}\}_{(k)}.$$

It was shown in [426] that the time rate of change of an observable is

$$\frac{d}{dt}\mathcal{O}_\Phi \approx \mathcal{O}_{\{\Phi, \bar{\mathcal{H}}_0\}}.$$

Thus, contrary to claims often made in the community, observables can be time dependent. Furthermore, there is a Taylor expansion

$$\mathcal{O}_\Phi \approx \sum_{k=0}^{\infty}\frac{t^k}{k!}C_k \qquad \text{with} \qquad C_k[\Phi, \chi] = \{\Phi, \bar{\mathcal{H}}_0\}_{(k)}. \tag{C.104}$$

The constants C_k depend, as indicated, on the gauge choice, which in turn enters in the definition of $\bar{\mathcal{H}}$. These constants may be either real constant numbers or constants of motion. Quite interestingly, the observables contain generators of rigid Noether symmetries: if C_k is a nontrivial constant of motion, the Noether charge is

$$Q_k := C_k + U^\alpha_{k\beta}N^\beta P_\alpha,$$

where the U-coefficients are indirectly defined from $\{C_k, \mathcal{H}_\beta\} =: U^\alpha_{k\beta}\mathcal{H}_\alpha$. For every Φ the infinitesimal symmetry transformations are thus given by $\delta_k\Phi = \epsilon\{\Phi, Q_k\}$. In general, the explicit form of the Noether charges (and their related symmetries) depends on the choice of the gauge function χ.
The mapping $\Phi \to \mathcal{O}_\Phi$ is not a canonical transformation ($\{\mathcal{O}_\Phi, \mathcal{O}_{\Phi'}\} \neq \mathcal{O}_{\{\Phi,\Phi'\}}$). However,

$$\{\mathcal{O}_\Phi, \mathcal{O}_{\Phi'}\} \approx \mathcal{O}_{\{\Phi,\Phi'\}^*_\Delta}$$

where the Dirac bracket is defined with respect to the matrix $\Delta_{AB} = \{\varphi_A, \varphi_B\}$, built from the Poisson brackets of the set of constraints $\varphi_A = \{\mathcal{H}_\alpha, \mathcal{P}_\alpha, \chi^\alpha, \dot{\chi}^\alpha\}$.

Example: Observables for Finite-Dimensional Systems

Although the previous considerations referred to gravity as a field theory, they can be made explicit for finite-dimensional systems. As specific cases we will consider the free relativistic particle in Minkowski space and an example from cosmology. These can be derived from an action

$$S[q^\alpha, N] = \int dt\ N\left[\frac{1}{2N^2}G_{\alpha\beta}(q)\dot{q}^\alpha\dot{q}^\beta - U(q)\right]. \tag{C.105}$$

with a non-singular matrix $G_{\alpha\beta}$. The canonical momenta are $p_\alpha = N^{-1}G_{\alpha\beta}\dot{q}^\beta$, $P \approx 0$. By this, the total Hamiltonian is $H_T = N\mathcal{H}(q,p) + \lambda P$, with

$$\mathcal{H} = \left[\frac{1}{2}\bar{G}^{\alpha\beta}p_\alpha p_\beta + U(q)\right],$$

where $\bar{G}^{\alpha\beta}$ is the inverse to $G_{\alpha\beta}$.
Several of the previous expressions simplify in one dimension. The gauge condition (C.102) will be written as $\chi = t - T(q, N, p, P)$ and the $\mathcal{A}$ matrix simplifies to $\mathcal{A}(q,p) := \{T, \mathcal{H}\}$. Furthermore,

$$\begin{aligned}
G_\xi &= P\dot{\xi} + \mathcal{H}\xi \qquad \overline{\mathcal{H}} = \mathcal{A}^{-1}\big(\mathcal{H} - \mathcal{A}^{-1}N\{\mathcal{A},\mathcal{H}\}P\big)\\
\mathcal{O}_\Phi[\chi] &\approx \sum_{k=0}^{\infty}\frac{1}{k!}t^k C_k[\Phi,\chi] \qquad \text{with} \qquad C_k[\Phi,\chi] = \{\Phi, \bar{\mathcal{H}}\}_{(k)}\\
Q_k &= C_k + U_k N P \qquad U_k\mathcal{H} = \{C_k, \mathcal{H}\}.
\end{aligned}$$

(A) Relativistic particle in Minkowski space
Comparison with (C.77) yields $G_{\alpha\beta} \Leftrightarrow \eta_{\mu\nu}$, $U(q) \Leftrightarrow \frac{1}{2}m^2$, leading to $\mathcal{H} = \frac{1}{2}(\eta^{\mu\nu}p_\mu p_\nu + m^2)$. Since the Poisson brackets of p_μ and P with the G_ξ vanish, these momenta are recognized as observables. The rescaling function $\mathcal{A}$ is

$$\mathcal{A} = \{T, \mathcal{H}\} = \frac{\partial T}{\partial q^\mu}\{q^\mu, \mathcal{H}\} = \frac{\partial T}{\partial q^\mu}p_\mu.$$

As for a gauge choice, let $T(q)$ be a linear combination of the fields q^μ: $T(q) = \alpha_\mu q^\mu = (\alpha q)$, where the α_μ do not depend on the q^μ. For this choice $\mathcal{A} = (\alpha p)$. Since $\{\mathcal{A}, \mathcal{H}\} = 0$, furthermore $\overline{\mathcal{H}} = (\alpha p)^{-1}\mathcal{H}$ and $G_{\bar{\xi}} = (\alpha p)^{-1}\dot{\bar{\xi}}P + \bar{\xi}\overline{\mathcal{H}}$. The nested Poisson brackets $\{q^\mu, \overline{\mathcal{H}}\}_{(k)}$ are calculated to be

$$\begin{aligned}
&\{q^\mu, \overline{\mathcal{H}}\}_{(1)} = \{q^\mu, \overline{\mathcal{H}}\} \approx (\alpha p)^{-1}\{q^\mu, \mathcal{H}\} = (\alpha p)^{-1}p^\mu\\
&\{q^\mu, \overline{\mathcal{H}}\}_{(j)} = 0 \qquad \text{for} \qquad j \geq 2.
\end{aligned}$$

Therefore the observables associated to q^μ are

$$\mathcal{O}_{q^\mu}[\chi] \approx q^\mu + \chi \frac{p^\mu}{(\alpha p)} = q_\mu - \frac{p^\mu}{(\alpha p)}(\alpha q) + \frac{p^\mu}{(\alpha p)}\, t.$$

Notice especially that $\mathcal{O}_{\alpha q} = t$, which is consistent with the gauge choice $t = (\alpha q)$. From the expression $\mathcal{O}_{q^\mu}$ we read off the conserved quantities:

$$C_0[q^\mu] = q^\mu - (\alpha p)^{-1} p^\mu (\alpha q) \qquad\qquad C_1[q^\mu] = (\alpha p)^{-1} p^\mu.$$

Since we already know that p^μ is conserved, the object C_1 offers no new information. However, multiplying C_0 with the constant (αp) we get the constants of motion

$$C^\mu := (\alpha p) C_0[q^\mu] = (\alpha_\nu p^\nu) q^\mu - (\alpha_\nu q^\nu) p^\mu = \alpha_\nu (p^\nu q^\mu - q^\nu p^\mu) = \alpha_\nu M^{\nu\mu},$$

thus the C^μ are just linear combinations of the angular momenta $M^{\mu\nu}$. Indeed the $C_i[q^\mu]$ can directly be identified with Noether charges since $\{C_i[q^\mu], \mathcal{H}\} = 0$. If now for instance we choose $\alpha = (1, 0, 0, 0)$, we have aside from the space-time translations $T_\mu := p_\mu$ also the boosts $K^j := q^0\, p^j - q^j p^0$, and their algebra delivers the generator of space rotations J^k.
Finally, for completeness, let us determine the observable associated to N. This is derived with the aid of $\{N, G_{\bar{\xi}}\} = (\alpha p)^{-1} \dot{\bar{\xi}}$ and $\{\{N, G_{\bar{\xi}}\}, G_{\bar{\xi}}\} = 0$ as

$$\mathcal{O}_N = N + \{N, G_{\bar{\xi}}\}_{\bar{\xi}=\chi} = N + (\alpha p)^{-1} \dot{\chi} = N + (\alpha p)^{-1} (1 - (\alpha \dot{q})) = (\alpha p)^{-1}$$

which again is consistent with the gauge choice: $N = \mathcal{A}^{-1}$.

Instead of the Minkowski metric any other background metric may be used. The example of a relativistic particle in an anti-de Sitter background is treated in [426]. Let me remark that in any case, those constants of motion that are linear in the momenta are directly related to the Killing symmetries of the background: Let $K^\mu(q) p_\mu$ be a constant of motion. Then

$$\{K^\mu p_\mu, \mathcal{H}\} = \{K^\mu p_\mu, \frac{1}{2}(g^{\rho\sigma} p_\rho p_\sigma + m^2)\} = -(\pounds_K g^{\rho\sigma}) p_\rho p_\sigma$$

where $\pounds_K$ is the Lie derivative with respect to the vector field $K = K^\mu \partial_\mu$. Therefore $\{K^\mu p_\mu, \mathcal{H}\} = 0$ is equivalent to $\pounds_K g^{\rho\sigma} = 0$.

(B) Minisuperspace cosmology
One speaks of minisuperspace[18] gravity if the symmetries imposed upon spacetime makes the system finite dimensional. In cosmology for instance, assuming spatial homogeneity ("looking the same at each location"), the most general line element

---

[18] The name was coined by Ch. Misner; despite the "super" it has nothing at all to do with supersymmetry.

can be written[19] as $ds^2 = -N^2(t)dt^2 + h_{ij}(t)dx^i dx^j$. An important class are the Bianchi-type cosmologies. These are classified according to their isotropy groups. If the structure constants of the associated Lie algebras fulfill $f_{ab}^{\ \ b} = 0$, one is allowed to perform a "symmetry reduction", that is to insert the metric into the action. And it is known that for a large subset of Bianchi-type cosmologies the action can always be brought into the generic form (C.105).

Astrophysical observations are in good agreement with a cosmological model which is not only spatially homogeneous but also spatially isotropic ("looking the same in each direction"). This is based on the Robertson-Walker metric for which

$$ds^2 = -N^2 dt^2 + a^2(d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta d\varphi^2))$$

with the scale factor $a(t)$ describing the expansion of our universe.The function $f(\chi)$ depends on the topology of the three-space, namely $f(\chi) = (\sin\chi, \chi, \sin h\,\chi)$ for $k = (+1, 0, -1)$. Inserting this metric for example into the action describing gravity coupled to a scalar field φ, one obtains

$$S = S_g + S_m = \frac{3}{\kappa}\int dt\ N\big(-\frac{a\dot{a}^2}{N^2} + ka - \frac{\Lambda a^3}{3}\big) + \frac{1}{2}\int dt\ Na^3\big(\frac{\dot{\varphi}^2}{N^2} - m^2\varphi^2\big).$$

Here Λ is the cosmological constant, and m the mass of the scalar field; for a derivation see [316]. By denoting $q^0 = a,\ q^1 = \varphi$ you may realize that this has the form of (C.105). In particular, for a flat universe without a cosmological constant and a massless scalar, the Lagrangian for this model becomes

$$L = \frac{1}{2N}\Big(-\frac{12a}{\kappa}\,\dot{a}^2 + a^3\dot{\varphi}^2\Big) = \frac{1}{2N}e^{-3\Omega}\Big(-\frac{12a}{\kappa}\,\dot{\Omega}^2 + \dot{\varphi}^2\Big) \quad \text{with} \quad a = e^{-\Omega}. \tag{C.106}$$

(We could even get rid of the numerical factor before $\dot{\Omega}^2$ by rescaling Ω.) To see the generality of this approach, let us consider the so-called anisotropic Bianchi type-I cosmology, defined by the metric

$$ds^2 = -N^2 dt^2 + e^{-2\Omega}(e^{2\beta_x}dx^2 + e^{2\beta_y}dy^2 + e^{2\beta_z}dz^2)$$

with the restriction $\beta_x + \beta_y + \beta_z = 0$. By steps similar to those taken in the previous case, one derives the Lagrangian

$$L = \frac{1}{2N}\omega^{-1}(\Omega)\left(-\dot{\Omega}^2 + \dot{\beta}_+^2 + \dot{\beta}_-^2 + \dot{\varphi}^2\right),$$

with $\omega = e^{3\Omega}$ and $\beta_+ = \frac{1}{2}(\beta_x + \beta_y),\ \beta_- = \frac{1}{2\sqrt{3}}(\beta_x - \beta_y)$. The Hamiltonian is determined to be $H = N\mathcal{H} + \lambda P$ with

$$\mathcal{H} = \frac{1}{2}\omega(\Omega)(-p_\Omega^2 + p_+^2 + p_-^2 + p_\varphi^2).$$

[19] In a mathematically correct way, spatial homogeneity–as well as isotropy–is to be defined via Killing symmetries; see [526], [455].

Similar to the relativistic particle, all momenta are observables, independent of a gauge choice. Next, let us find the observables associated to $(\Omega, \beta_{\pm}, \varphi)$. Choosing for instance the gauge $\chi = t - \Omega = 0$, gives $\mathcal{A} = -\omega p_\Omega$ and $\overline{\mathcal{H}} = \mathcal{A}^{-1}\mathcal{H} - 3NP$. From this, a straightforward calculation yields observables

$$\mathcal{O}_\Omega = t, \qquad \mathcal{O}_\Phi = \left(\Phi + \Omega\,\frac{p_\Phi}{p_\Omega}\right) - \frac{p_\Phi}{p_\Omega}\,t, \qquad \mathcal{O}_N = -\frac{1}{p_\Omega}e^{-3t}$$

where Φ stands for the set $\Phi = \{\beta_+, \beta_-, \varphi\}$. Notice that the expression for $\mathcal{O}_\Omega$ is compatible with the gauge condition, and the one for $\mathcal{O}_N$ is consistent with $N = \mathcal{A}^{-1}$ (in the chosen gauge). The two coefficients in $\mathcal{O}_\Phi$ are used to calculate the Noether charges. Introducing a notation by which the three functions in Φ are called Φ_K and their momenta p_K the Noether charges are

$$P_\Omega := p_\Omega - 3NP, \qquad P_K := p_K, \qquad Q_K := (p_\Omega \Phi_K + p_K \Omega) - 3\Phi_K NP.$$

The algebra closes with further charges $R_{KL} = \{Q_K, Q_L\} = p_K \Phi_L - p_L \Phi_K$ and the full algebra becomes

$$\begin{aligned}
\{P_i, P_j\} &= 0 \quad P_i \in \{P_\Omega, P_K\} \\
\{P_\Omega, Q_K\} &= -P_K \quad \{P_\Omega, R_{KL}\} = 0 \\
\{P_K, Q_L\} &= -\delta_{KL} \quad P_\Omega\{P_J, R_{KL}\} = \delta_{JK} P_L - \delta_{JL} P_K \\
\{Q_K, Q_L\} &= R_{KL} \quad \{Q_J, R_{KL}\} = \delta_{JK} Q_L - \delta_{JL} Q_K \\
\{R_{IJ}, R_{KL}\} &= \delta_{IL} R_{JK} - \delta_{JL} R_{IK} + \delta_{JK} R_{IL} - \delta_{IK} R_{JL}.
\end{aligned}$$

It can be verified that this is isomorphic to the conformal algebra $\mathfrak{c}(2, 1)$. And also one may verify that the Lagrangian (C.106) is quasi-invariant under the transformations generated by the Noether charges $C_k \in \{P_\Omega, P_K, Q_K, R_{KL}\}$.

Ashtekar Variables

The previously described ADM approach to canonical GR prevailed up to the 1980's. In 1986, A. Ashtekar introduced what is now sometimes called "new variables", which make it possible to cast canonical GR in a form in which it more closely resembles canonical Yang-Mills theories. In my presentation I follow [316].

Basic entities for constructing the Ashtekar variables are "triads" h^i_a spanning the three-geometry similar to the tetrads describing the four-geometry. Thus the triads obey

$$h_{ij} h^i_a h^j_b = \delta_{ab}.$$

Furthermore, one can introduce related objects h^a_j "living" in the cotangent space. These fulfill $h^a_j h^i_a = \delta^j_i$ and $h^a_j h^j_b = \delta^a_b$. How to go from the tetrads e^I_μ and the normal n^μ to the $x^0 = $ const hypersurface to the triads is elucidated in [269]. Instead of the triads themselves consider their "densitized" variants

$$\tilde{h}^i_a := \sqrt{h} h^i_a \qquad \text{with} \qquad \sqrt{h} = \det(h^b_j).$$

It can be shown that the extrinsic curvature (C.91) expressed as

$$K^a_i := K_{ij} h^{ja}$$

is canonically conjugate to $\tilde{h}^i_a$. Further there is the identity

$$\mathcal{G}_a := \epsilon_{abc} K^a_i \tilde{h}^{ci} \equiv 0, \tag{C.107}$$

which in the Rosenfeld-Dirac-Bergmann algorithm become constraints (called the "Gauß-constraints"). These constraints generate $\mathbf{SO(3)}$-rotations, symmetries originating from the use of triads.

The canonical variables considered since the work of Ashtekar [16] are the

$$H^i_a = \frac{1}{8\pi\beta} \tilde{h}^i_a(x)$$

together with a mixture of K^a_i and $\tilde{h}^i_a(x)$ into a connection A^a_i defined by

$$G A^a_i(x) := \Gamma^a_i(x) + \beta K^a_i. \tag{C.108}$$

Newton's constant G needs to enter so that the objects GA^a_i acquire the dimension of an inverse length–like a Yang-Mills connection. Furthermore, β is an arbitrary non-vanishing complex number (the *Barbero-Immirzi parameter*). In (C.108) Γ^a_i is related to the $3D$ spin connection ω_{iab} as

$$\Gamma^a_i = -\frac{1}{2} \omega_{ibc} \epsilon^{abc}.$$

The configuration variables H^i_a and the momenta A^a_i are canonically conjugate:

$$\{H^j_b(x), A^a_i(y)\} = \delta^j_i \delta^b_a \delta(x, y) \qquad \{H^i_a, H^j_b\} = 0 = \{A^a_i, A^b_j\}.$$

How do the constraints look in the new variables? The Gauß constraints (C.107) are equivalent to

$$\hat{\mathcal{G}}_a = \partial_i H^i_a + G \epsilon_{abc} A^b_i H^{ci} := \mathcal{D}_i H^i_a \approx 0. \tag{C.109}$$

This is formally similar to the Gauß constraint (C.65) of Yang-Mills theory. From the commutator of two covariant derivatives, one calculates the curvature/field-strength associated to the connection/potential A^a_j as

$$F^a_{ij} = 2G\partial_{[i}A^a_{j]} + G^2\epsilon^a_{bc}A^b_i A^c_j.$$

The metric gravity constraints from (C.96) become proportional to

$$\hat{\mathcal{H}}_\perp = -(\det H^i_a)^{-1/2}\Big[\frac{\sigma}{2}\epsilon_{abc}H^{ai}H^{bj}F^c_{ij} + \frac{1-\beta^2\sigma}{\beta^2}H^i_{[a}H^j_{b]}(GA^a_i - \Gamma^a_i)(GA^b_j - \Gamma^b_j)\Big] + (GC) \tag{C.110a}$$

$$\hat{\mathcal{H}}_i = F^a_{ij}H^j_a + (GC) \tag{C.110b}$$

where (GC) stands for a linear combination of the Gauß constraints and where $\sigma = -1/+1$ for the Lorentzian/Euclidean case. Thus the second term in (C.110a) drops out for the choice $\beta = \pm i$ in the Lorentzian case and $\beta = \pm 1$ in the Euclidean case. Ashtekar originally used $\beta = i$, with the consequence that the connection (C.108) becomes complex. Thus one needed to introduce reality conditions; technicalities going beyond the scope of this book. Later J. F. Barbero introduced with $\beta = 1$ canonical variables similar to the ones by Ashtekar, but with the advantage of being real, and G. Immirzi pointed out, that any choice for the complex β is allowed. The price to be paid is a loss of simplicity, as can be seen at the Hamiltonian constraint (C.110a). Anyhow, being born into the world, the Barbero-Immirzi parameter β has given rise to many research activities. Among others came the observation that different Hamiltonian forms of GR can be derived from one and the same action by appropriate choices of a parameter [280]. This *Holst action* is an extension of the Hilbert-Palatini action derived from (7.66)

$$S_{\text{Holst}} = \int d^4x\, e\, e^\mu_I e^\nu_J \Big(R^{IJ}{}_{\mu\nu}(\omega) + \frac{\gamma}{2}\epsilon^{IJ}{}_{KL}R^{KL}{}_{\mu\nu}(\omega)\Big), \tag{C.111}$$

(here $2\kappa = 1$) in which, as indicated the tetrads and the spin connections are varied independently. For a constant parameter γ the second term does not influence the field equations since it is a boundary term[20]. In point of fact, the γ term seems to play a similar role as the θ term in non-Abelian gauge theories. The choices $\gamma = \{0, i, 1\}$ lead to the ADM, the Ashtekar, and the Barbero Hamiltonian for GR, respectively. For arbitrary choices of γ, the canonical formulation of the Holst action leads to a family of $SU(2)$ connection formulations of GR Hamiltonians.

Boundary Conditions and Surface Terms

In the sequel, I will deal solely with canonical general relativity in the ADM approach, although the procedures and findings described here apply *mutatis mutandis* to the

[20] The parameter γ is related to the Barbero-Immirzi parameter by $\gamma = -\frac{1}{\beta}$.

Ashtekar formulation (see [499]) and to canonical tetrad gravity (see [49] and references therein).

The central finding of ADM is the Hamiltonian (C.96)

$$H_{ADM}\left[N, N^i\right] = \int_\Sigma d^3x \, (N\mathcal{H}_\perp + N^i\mathcal{H}_i)$$

as a sum of constraints. Bryce DeWitt was the first to realize that this is only true if the three-manifold Σ is compact without a boundary [121]. For the asymptotically flat case, in particular, DeWitt added a surface term at infinity

$$E_{ADM}\left[h_{ij}\right] = \oint d^2s_k(h_{ik,i} - h_{ii,k}) \tag{C.112}$$

to the ADM-Hamiltonian so that

$$H_{DW} = H_{ADM} + E_{ADM}\left[h_{ij}\right]$$

coincides asymptotically with the known total energy expression from linearized GR. The role of surface terms was further clarified in pioneering work by T. Regge and C. Teitelboim [437]. They showed that the GR vacuum field equations do in fact not inevitably follow from the Hamiltonian H_{ADM}. The variation of this Hamiltonian is generically given by

$$\delta H_{ADM} = \int d^3x \, \left[A^{ij}(x)\delta h_{ij}(x) + B_{ij}(x)\delta p^{ij}(x)\right] + \oint d^2s_k \, C^k.$$

The explicit form of the C^k does not matter here; they are given as Eq. (2.6) in [437]. In any case, only if the surface term vanishes, are the functional derivatives of H_{ADM} defined, and only in this case are we allowed to conclude from Hamilton's variational principle that

$$A^{ij} =: \frac{\partial H_{ADM}}{\partial h_{ij}} = -\dot{p}^{ij} \qquad B_{ij} =: \frac{\partial H_{ADM}}{\partial p^{ij}} = \dot{h}_{ij}.$$

Notice that otherwise, we not only would find the wrong field equations, but we would get no consistent field equations at all! In the case of closed spaces, the surface integral vanishes, and thus this problem does not arise. For open spaces, however, this is not true in general. Regge and Teitelboim cured this insufficiency for asymptotically flat spaces by amending the ADM Hamiltonian through a surface term, that is by starting from a modified Hamiltonian

$$H\left[N, N^i\right] = H_{ADM}\left[N, N^i\right] + \oint_{\mathcal{S}} B(N, N^i)d\mathcal{S}$$

where $B(N, N^i)$ is linear in its arguments and the integral is understood as the limit $r \to \infty$ on spheres of radius r in the asymptotically flat region of Σ. They were not only able to show that there is a choice of B such that the modified Hamiltonian is functionally differentiable, but that by appropriate assumptions about the lapse function the Poincaré charges at spatial infinity can be defined–reproducing among other the ADM energy (C.112). Here are some more details: DeWitt already claimed that for asymptotically flat geometries, the falloff behavior of the three-metric for $r \to \infty$ must be $h_{ij} \sim \delta_{ij} + \mathcal{O}(r^{-1})$, and that the momenta p^{ij} behave as r^{-2}. Regge and Teitelboim argue that lapse and shift must behave as

$$N^\mu \overset{r\to\infty}{\sim} \alpha^\mu + \beta_i^\mu x^i$$

where α^μ describes space-time translations, $\beta_{ij} = -\beta_{ji}$ spatial rotations, and $\beta_i^\perp$ boosts. Given this falloff behavior the variation of the ADM Hamiltonian results as

$$\delta H_{ADM} = \int d^3x \, \left[A^{ij}(x)\delta h_{ij}(x) + B_{ij}(x)\delta p^{ij}(x)\right] + \alpha^\mu \delta \check{P}_\mu - \frac{1}{2}\beta_i^\mu \delta \check{M}_\mu^i$$

with

$$\begin{aligned}
\check{P}^\perp &= \oint_{r\to\infty} d^2s_k(h_{ik,i} - h_{ii,k}), \qquad \check{P}^i = -2\oint_{r\to\infty} d^2s_k \; p^{ki} \\
\check{M}^{ij} &= -2\oint_{r\to\infty} d^2s_k(x^i p^{kj} - x^j p^{ki}), \\
\check{M}_{\perp j} &= -2\oint_{r\to\infty} d^2s_k\left[x_j(h_{ik,i} - h_{ii,k}) - h_{jk} + h_{ii}\delta_{jk}\right].
\end{aligned}$$

These can be identified with the total energy, the total momentum, the total angular momentum and the generator of boosts respectively. Specifically $\check{P}^\perp \equiv E_{ADM}\left[h_{ij}\right]$. Now it is obvious how to arrive at a Hamiltonian that may serve to derive the GR Hamilton equations:

$$H_{RT} = H_{ADM} - \alpha^\mu \check{P}_\mu + \frac{1}{2}\beta_i^\mu \check{M}_\mu^i. \tag{C.113}$$

It is really striking that the quest for mathematical consistency–namely the existence of functional derivatives for the action–yields physically desirable results about conserved quantities at infinity!

The analysis of Regge and Teitelboim was improved by R. Beig and N. Ó Murchadha [41]: Amongst other things they derived falloff conditions for the lapse function and the shift vector from internal consistency (compatibility of the boundary conditions and the evolution equations) instead of the previous more or less *ad hoc* assumed behavior, and they corrected the expression for the boost generator $\check{M}_{\perp j}$. Furthermore, "Precise meaning is given to the statement that, as a result of these boundary conditions, the Poincaré group acts as a symmetry group on the phase space of general relativity" in a fairly abstract mathematical manner.

Arnowitt, Deser, and Misner had by laxity (common in those "ancient times") discarded any total derivative terms in deriving "their" Hamiltonian. Only during the last two decades has the role of boundary terms become recognized as pivotal. ADM even started with the wrong action: The Hilbert action does not lead to the GR vacuum field equations in the case of open geometries. Instead, as explained in Sect. 7.5.2, one has to consider the "Trace-K" action

$$S_K = \frac{1}{2\kappa}\int d^4x\ \sqrt{-g}R + \frac{1}{\kappa}\int_{\partial\Omega} d^3x\ \ \sqrt{\epsilon h}K$$

which is the sum of the Hilbert action and the Gibbons-Hawking-York boundary term (7.72). Curiosity provokes the question as to which Hamiltonian results if one starts with this action and keeps a record of all boundary terms arising in between (essentially from integration by parts). The gratifying answer was given in [258]: The resulting Hamiltonian is the same as the one established within the logic of Regge and Teitelboim in case of asymptotically flat spacetimes. The answer is even more satisfying in that S. Hawking and G. Horowitz could write down the gravitational Hamiltonian for more general boundary conditions.

The extension of the results of Regge and Teitelboim from asymptotic Minkowski to (A)dS spacetimes has been carried out by M. Henneaux and C. Teitelboim [271], and to Einstein-Cartan and teleparallel gravitational theories by M. Blagojević and I. Nicolić; see Chap. 6 in [49].

C.3.4 Tetrad Gravity

Constraints in Riemann-Cartan Gravity Theories

Instead of immediately treating the Hamiltonian formulation of GR coupled to Dirac fields, let me address the generic structure of the Hamiltonian and the algebra of constraints in theories with both curvature and torsion. Specifically for Poincaré gauge theories, the Hamiltonian version was tackled in a series of publications by M. Blagojević, I. A. Nicolić and M. Vasilić; their findings are summarized in [48], [49] and in the chapter "Hamiltonian Structure" of [50].

The gravitational part of the theory is given in terms of the vierbeins e^K_μ and the spin connections $\omega^{IJ}{}_{\mu}$. Their derivatives enter the Lagrangian $e\mathcal{L}_G(R^{IJ}{}_{\mu\nu}, T^I{}_{\mu\nu})$ only through the curvature[21] $R^{IJ}{}_{\mu\nu}$ and the torsion $T^I{}_{\mu\nu}$, and these derivatives appear antisymmetrized. Therefore the momenta canonically conjugate to e^K_0 and $\omega^{IJ}{}_0$ vanish. These are primary constraints

[21] Here, I refrain from indicating by a tilde that in the present context we are dealing with the Riemann-Cartan curvature.

$$\phi^0_K = \pi^0_K \approx 0 \qquad \phi_{IJ}^{\ \ 0} = \pi_{IJ}^{\ \ 0} \approx 0$$

which are always present independent of the specific form of the Lagrangian. They are called "sure" constraints by Blagojević *et.al.* to contrast them from "if" constraints which occur for certain parameter combinations in PG theories. The total Hamiltonian is thus given by $\mathcal{H}_T = \mathcal{H}_C + u^K_0 \phi^0_K + (1/2)u^{IJ}_{\ \ 0}\pi_{IJ}^{\ \ 0} + (u\cdot\phi)$ where the last term stands symbolically for the if-primaries. Now as it turns out, the canonical Hamiltonian can always be written as

$$\mathcal{H}_C = e^K_0 \mathcal{H}_K - \frac{1}{2}\omega^{IJ}_{\ \ 0}\mathcal{H}_{IJ} + \partial_i D^i = N\mathcal{H}_\perp + N^i\mathcal{H}_i - \frac{1}{2}\omega^{IJ}_{\ \ 0}\mathcal{H}_{IJ} + \partial_i D^i, \quad \text{(C.114)}$$

with the lapse function and the shift functions now expressed by the vierbeins and the normal n^K, so that $\mathcal{H}_\perp = n^K\mathcal{H}_K$ and $\mathcal{H}_i = e_i^K\mathcal{H}_K$. Any tangent space object, say V_K, is expressible by its parallel and its orthogonal components:

$$V_K = n_K V_\perp + V_{\bar{K}} \;\longleftrightarrow\; V_\perp = n_K V^K, \qquad V_{\bar{K}} = (\delta^I_K - n^I n_K)V_I.$$

The Hamiltonian (C.114) has a striking similarity with the ADM expression for metric gravity–aside from the $\mathcal{H}_{IJ}$-term. One gets immediately the sure secondary constraints

$$\mathcal{H}_\perp \approx 0, \qquad \mathcal{H}_i \approx 0, \qquad \mathcal{H}_{IJ} \approx 0. \qquad \text{(C.115)}$$

From the Lagrangian perspective, we know that the theory is invariant with respect to diffeomorphism and local Lorentz rotations. And being aware that a variational symmetry makes itself felt in the Hamiltonian by the existence of first-class constraints, we expect that the secondary constraints are first-class and will be part of a phase-space symmetry generator. This indeed turns out to be true.

It is possible to derive the explicit (phase-space) expressions of the constraints (C.115) for any gravitational Lagrangian. I renounce to write it down here, you find this in the literature cited above. A peculiarity was found for certain PGT's, in that the number and type of constraints may be different in different regions of phase space. In [90] the origin of this degenerate behavior was analyzed, with the prospect to get a viable PGT with fewer parameters.

Matter Constraints

In the literature cited before it is shown that the Hamiltonian for minimally coupled matter has the same structure as the gravitational part (C.114). With the conventions to rewrite the matter part of the Lagrangian

$$e\mathcal{L}_M(Q, \nabla_\mu Q) = e\overline{\mathcal{L}}_M(Q, \nabla_{\bar{K}}Q, \nabla_\perp Q; n^K)$$

and using the factorization of the determinant as $e = NJ$, where J does not depend on e_0^K, the matter field momenta become

$$P = \frac{\partial(e\mathcal{L}_M)}{\partial(\partial_0 Q)} = J\frac{\partial\overline{\mathcal{L}}_M}{\partial(\nabla_\perp Q)}.$$

Now the matter part of the canonical Hamiltonian is found to be similar to (C.114) with

$${}^M\mathcal{H}_\perp = P\nabla_\perp Q - J\overline{\mathcal{L}}_M, \qquad {}^M\mathcal{H}_i = P\nabla_i Q, \qquad {}^M\mathcal{H}_{IJ} = P\Sigma_{IJ} Q$$

where Σ_{IJ} is the spin matrix which enters because $\nabla_0 Q = N\nabla_\perp Q + N^i\nabla_i Q = \partial_0 Q + \frac{1}{2}\omega^{IJ}{}_0\Sigma_{IJ}Q$. Notice that only the constraint ${}^M\mathcal{H}_\perp$ depends on the Lagrangian, while the others do not. And be aware that the matter Lagrangian also may give rise to further primary constraints.

Einstein-Cartan-Dirac Interaction

Now we could turn to the example of coupling fermions to GR. However, in the context of the previous results, GR is not a natural theory, because its action corresponds to the Einstein-Cartan action plus a multiplier term taking account of the vanishing torsion. Therefore, "for the sake of simplicity" let us investigate the interaction of a Dirac spinor with Einstein-Cartan gravity.

The gravitational part is

$$\begin{aligned} S_{EC} &= \frac{1}{2\kappa}\int d^4x\sqrt{-g}R = -\frac{1}{8\kappa}\int d^4x\ \epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL}e^K_\rho e^L_\sigma R^{IJ}_{\mu\nu} \\ &= -\frac{1}{4\kappa}\int d^4x\ \epsilon^{\mu\nu\rho\sigma}_{IJKL}e^K_\rho e^L_\sigma(\partial_\mu\omega^{IJ}{}_\nu + \omega^I{}_{K\mu}\omega^{KJ}{}_\mu). \end{aligned}$$

with the abbreviation $\epsilon^{\mu\nu\rho\sigma}_{IJKL} := \epsilon^{\mu\nu\rho\sigma}\epsilon_{IJKL}$. The momenta are

$$\pi^\mu_K = \frac{\partial L_{EC}}{\partial(\partial_0 e^K_\mu)} = 0 \qquad \pi_{IJ}{}^\mu = \frac{\partial L_{EC}}{\partial(\partial_0\omega^{IJ}_\mu)} = \frac{1}{2\kappa}\epsilon^{0\mu\rho\sigma}_{IJKL}e^K_\rho e^L_\sigma.$$

Thus not only the π^0_K and $\pi_{IJ}{}^0$ do vanish (as derived before independently of any Lagrangian), but there are further primary constraints

$$\phi^i_K = \pi^i_K \approx 0 \qquad \phi_{IJ}{}^k = \pi_{IJ}{}^k - \frac{1}{2\kappa}\epsilon^{0k\rho\sigma}_{IJKL}e^K_\rho e^L_\sigma.$$

For the free EC theory, all together there are 40 primary constraints. This is already severe but is not yet the end, since we need to guarantee that the constraints are stable in time. This gives rise to 16 secondary constraints. It turns out that 10 of the primary constraints and 10 of the secondary constraints are first class. With the number $N_{FC} = 20$ and $N_{SC} = 36$ and bearing in mind that one gauge condition corresponds to each first-class constraint, one finds for the $N = 40$ field components (in configuration space) the number of degrees of freedom $2N - (2N_{FC} + N_{SC}) = 4$ in phase space. (For details see [48], [49].)

As for the matter part we have $eL_M = e(i\bar{\psi}\gamma^\mu \nabla_\mu \psi - m\bar{\psi}\psi)$. The momenta π and $\bar{\pi}$, canonically conjugate to ψ and $\bar{\psi}$, respectively are part of the primary constraints $\phi = \pi + (i/2)J\bar{\psi}\gamma^\perp \approx 0$ and $\bar{\phi} = \bar{\pi} \approx 0$. We met these constraints in (C.56) where (in the coupled Maxwell-Dirac case) they turned out to be second-class.

In using now the previous results for the gravitational part (with the specific primary constraint structure of the EC theory) and the results for the matter part (with the primary constraints resulting from the definition of the momenta canonically conjugate to the Dirac fields) you are able to write down the explicit form of the canonical Hamiltonian. Next you need to start the Rosenfeld-Dirac-Bergmann algorithm. In a first step, you need to investigate the stability of the 48 primary constraints. You will find that–incorporating the 8 extra constraints–the previous count remains valid, but that due to the presence of fermions the explicit form of the first- and second-class constraints is modified.

You see that the Hamiltonian analysis of the Einstein-Cartan-Dirac theory requires some ingenuity. This was already realized when people formulated this program for the tetrad version of GR. Here you might think of a first-order approach (related to the previous procedure) or a second-order approach where you express the spin connection by the tetrads and their derivatives. Both approaches, surprisingly, did not immediately lead to the same results. The comparison became even more problematic, if in the course of the calculations one took over findings from metric gravity. A résumé is given in [87]. One now understands that different procedures are interrelated by canonical transformations, and these are mediated by boundary terms in the action. But I am still not sure that all the diverse approaches for tetrad GR in their mutual relationships are fully understood.

C.3.5 Einstein-Dirac-Yang-Mills-Higgs Theory

According to present-day knowledge fundamental physics is characterized by a Lagrangian with gravitational, gauge boson, fermionic and Higgs fields. Thus this appendix may be expected to close with the Hamiltonian for today's "World Theory". Because of the presence of fermionic fields, the Hamiltonian has to be formulated with tetrads, and you have already realized in the previous subsection that things will become quite intricate. Therefore in this subsection I we will be far less ambitious than the title announces and only show a glimpse at what happens in the case of scalar and vector fields coupled to gravity.

Constraint Structure

As already mentioned, for any generally covariant theory, the algebra of constraints (C.98) is the same, although the explicit form of the constraints themselves will differ. This led C. Teitelboim to reflect about the "Hamiltonian Structure of Space-Time" [498].

Let us consider the case of a theory of integer-spin fields coupled to gravity. Denote by $\{q_\alpha, p^\alpha\}$ the set of all canonical phase-space variables. It is quite obvious that because the Lagrangian is a sum of a gravitational and a matter part, the ADM-type action has the generic form

$$S = \int d^4x \left[\dot{q}_\alpha p^\alpha - N\mathcal{H}_\perp - N^i\mathcal{H}_i - \lambda^I \mathcal{C}_I\right]$$

where the last term is a linear combination of constraints possibly arising from the local symmetry of the matter theory. If split into the gravitational and the matter part, this becomes

$$S = \int d^4x \left[\dot{h}_{ij} p^{ij} + \dot{Q}_a P^a - N({}^G\mathcal{H}_\perp + {}^M\mathcal{H}_\perp) - N^i({}^G\mathcal{H}_i + {}^M\mathcal{H}_i) - \lambda^I \mathcal{C}_I\right].$$

But it is less obvious that the constraints $\mathcal{H}_\perp$ and $\mathcal{H}_i$ acquire the generic form

$$\begin{aligned} \mathcal{H}_\perp(x) &= G^{\alpha\beta}(q)p_\alpha p_\beta + \sqrt{-h}\; U(q, \partial_i q) \\ \mathcal{H}_i(x) &= \mathcal{D}_{i\alpha}\; p^\alpha \end{aligned} \tag{C.116}$$

for both the gravitational and the matter part. Here, G is a nonsingular metric, and $\mathcal{D}$ a differential operator. For pure gravity, where the phase space variables are $\{h_{ij}, p^{ij}\}$, we identify

$$\begin{aligned} G^{\alpha\beta} &\leftrightarrow G^{ijkl} := \frac{1}{2\sqrt{-h}}(h^{ik}h^{jl} + h^{il}h^{jk} - h^{ij}h^{kl}) \\ U(q) &\leftrightarrow -\frac{1}{2\kappa}\,{}^{(3)}R \\ \mathcal{H}_i &\leftrightarrow -2p_i{}^j{}_{|j}. \end{aligned}$$

For a massive scalar field $Q \simeq \varphi,\;\; P \simeq \pi$:

$$\begin{aligned} {}^M\mathcal{H}_\perp &= \frac{1}{2\sqrt{-h}}\pi^2 + \frac{1}{2}\sqrt{-h}(h^{ij}\;\varphi_{,i}\varphi_{,j} + m^2\varphi^2) \\ {}^M\mathcal{H}_i &= \varphi_{,i}\pi. \end{aligned}$$

And for a Yang-Mills field with $Q_\alpha \simeq A^a_i,\ P^a \simeq \Pi^a_i$:

$$
\begin{aligned}
{}^M\mathcal{H}_\perp &= \frac{1}{2\sqrt{-h}}\, h^{ij}\, \gamma_{ab}\, \Pi^a_i \Pi^b_j + \frac{1}{2}\sqrt{-h}\, h^{ij} h^{kl}\, \gamma_{ab}\, F^a_{ik} F^b_{jl} \\
{}^M\mathcal{H}_i &= F^a_{ij}\, \Pi^j_a \\
\lambda^\alpha \mathcal{C}_\alpha &\leftrightarrow A^a_0\, (\Pi^i_{a\,|i} - f^c_{\ ab}\, A^b_i\, \Pi^i_c)
\end{aligned}
$$

where γ_{ab} is the Cartan-Killing metric of the internal gauge group.

Projectability and Mixed Symmetry Transformations

To safeguard the equivalence of the tangent space and the cotangent space description of singular systems not only on the level of dynamical equations but also with respect to symmetries, one must investigate the projectability of the tangent space symmetry transformations under the Legendre map. We saw in the case of pure gravity that the transformation parameter has to depend on the lapse and shift functions in the specific form (C.100). A peculiarity arises if a matter theory which itself exhibits local symmetries is coupled to the gravitational field.

Take for example the Yang-Mills Lagrangian

$$
\mathcal{L}_{YM} = -\frac{1}{4}\sqrt{|g|} F^a_{\mu\nu} F^b_{\rho\sigma} g^{\mu\rho} g^{\nu\sigma} \gamma_{ab}.
$$

which exhibits primary phase-space constraints

$$
\phi_a := \Pi^0_a = \frac{\partial \mathcal{L}_{YM}}{\partial \dot{A}^a_0} = \sqrt{|g|} F^b_{\mu\nu} g^{0\mu} g^{0\sigma} \gamma_{ab} \approx 0.
$$

With the zero eigenvectors of the Hessian $W^{\mu\nu}_{ab}$ found from

$$
\xi^b_{a\mu} = \frac{\partial \phi_a}{\partial \Pi^\mu_b} = \delta^b_a g_{0\mu}
$$

according to (C.14), the condition for Legendre projectability is controlled by the operator $\Gamma_a = \partial/\partial \dot{A}^a_0$. The Lagrangian $\mathcal{L}_{YM}$ is invariant with respect to the gauge transformations

$$
\delta_G(\theta) A^a_\mu = -\partial_\mu \theta^a - \theta^b A^c_\mu f^{bca}
$$

and this is clearly projectable: $\Gamma_a(\delta_G(\theta) A^b_\mu) = 0$.

Under diffeomorphism the Yang-Mills fields transform as

$$
\delta_D(\epsilon) A^a_\mu = -A^a_{\mu,\nu}\epsilon^\nu - A^a_\nu \epsilon^\nu_{\ ,\mu}.
$$

For projectability we now demand that this symmetry variation does not depend on $\dot{N}^\alpha$ and $\dot{A}_0^a$. This is true for the transformation of the spatial components, but not for the components A_0^a:

$$\delta_D(\epsilon)A_0^a = -\dot{A}_0^a\epsilon^0 - A_{0,k}^a\epsilon^k - A_0^a\dot{\epsilon}^0 - A_k^a\dot{\epsilon}^k.$$

This cannot be cured by a functional dependency of ϵ as in (C.100), nor by contemplating a dependency of ϵ on $\dot{A}_0^a$. But instead, it was found [462] that together with gauge transformations a modified diffeomorphism can be realized in the phase space. This modified diffeomorphism is a mixture of a diffeomorphism and a specific gauge transformation[22]:

$$\hat{\delta}(\epsilon) := \delta_D(\epsilon) - \delta_G(A_\nu^a\epsilon^\nu).$$

The modified transformation $\hat{\delta}(\epsilon)$ is projectable if the ϵ^ν are expressed by the "descriptors" ξ^ν according to (C.100); for further details see [425], also for the derivation of the symmetry generators. Let me remark here that it was already pointed out in [295] that the infinitesimal coordinate transformation of the gauge fields can be expressed as

$$\delta_D(\epsilon)A_\mu^a = F_{\mu,\nu}^a\epsilon^\nu - D_\mu(A_\nu^a\epsilon^\nu).$$

The second term is the same ϵ-dependent gauge transformation that is subtracted in the definition of $\hat{\delta}(\epsilon)$.

C.4 Alternative Approaches

- L. Faddeev and R. Jackiw proposed an alternative to the Dirac procedure in dealing with constraints. Indeed, the first sentence in [170] is unusually harsh: "It appears that some of our contemporary colleagues are unaware of modern, mathematically based, approaches to quantization, especially of constrained systems. They keep the prejudice that Dirac's method with its Dirac bracketing and categorization of constraints as first or second class, primary or secondary, 'is mandatory'". Faddeev and Jackiw start[23] from the canonical Hamiltonian H_C and the set of primary constraints $\phi^\rho = 0$ (remember that the latter are directly derived from the definition of the canonical momenta), and write the Lagrangian as

$$L = p_i\dot{q}^i - H_C(q, p) - u^\rho\phi_\rho \tag{C.117}$$

[22] A mixture of diffeomorphisms with gauge transformations seems to be necessary or at least advantageous in other instances too, see e. g. the introduction of group covariant Lie derivatives in Appendix F.4.

[23] I take the description of the FJ algorithm from [209].

with a set of multipliers u^ρ. Next, they eliminate as many variables as there are primary constraints by plugging these into the Lagrangian, resulting in a new Lagrangian $L' = a_s(x)\dot{x}^s - H(x)$, where x^s are the coordinates remaining after the reduction by the constraint solutions. Now according to a theorem by Darboux, it is always possible to change variables $x^s \to Q^r, P_r, Z^a$ such that the Lagrangian becomes

$$L' = P_r \dot{Q}^r - H'(Q^r, P_r, Z^a).$$

If H' does not depend on any of the Z^a (for which their canonical counterpart is missing), one is ready: The Q^r and P_r are the coordinates in the reduced phase space (which even has a genuine symplectic structure). Otherwise, it might be possible to eliminate some of the variables Z^a :

$$\frac{\partial H'}{\partial Z^a} = 0 \iff Z^{a_1} = f^{a_1}(Q^r, P_r, Z^{a_2}), \qquad f_{a_2}(Q^r, P_r) = 0$$

and one can argue that this gives rise to a Lagrangian

$$L'' = P_r \dot{Q}^r - H''(Q^r, P_r) - Z^{a_2} f_{a_2}(Q^r, P_r).$$

This has a structure similar to (C.117) and the algorithm can be restarted. The procedure terminates with the identification of the symplectic coordinates in the reduced phase. Notice that the Faddeev-Jackiw procedure in each of its steps aims at getting rid of unphysical variables. In contrast, the Rosenberg-Dirac-Bergmann procedure keeps all original variables and aims at a consistent picture of constraints that are conserved in time. And it preserves in its first-class constraints all variational symmetries of the original action. In contrast, the Faddeev-Jackiw method seems to lose the traces of local symmetries. However, nothing really gets lost. The two procedures were shown to be equivalent [209]–with a slight exception which is related to the Dirac conjecture. Nevertheless it is disputable which of the two, FJ or RDB, is to be preferred. It is true that the FJ method leads to the true degrees of freedom, but in order to disentangle them you need to solve constraints and to find the Darboux coordinates. Thus it is no surprise that the FJ procedure did not supersede the RDB method.

- The Faddeev-Jackiv method aims at reducing the original phase space in each of the algorithmic steps. In quite a different spirit a method inspired also by Faddeev was developed in a series of articles by I. A. Batalin, E. S. Fradkin, T. E. Fradkina and I. V. Tyutin (thereafter termed the BFFT formalism), designed for turning second-class constraints in a theory into first-class ones; the motivation being that first-class constraints have a more handy behavior both in canonical and in path-integral quantization. The central point in the BFFT procedure is to enlarge the phase space by as many canonical variables as there are second-class constraints (remember that their number is even in case of finite-dimensional bosonic systems) and to introduce new constraints as a power series in the second-class constraints.

Requiring these new constraints to be first class leads to consistency conditions for the expansion. At a first sight, this philosophy sounds a little weird. But the notion "original phase space" has no meaning at all. Might not the second-class constrained system be a gauge-fixed version of a larger gauge theory? For a review of the BFFT formalism in which you also find the reference to the original articles, see [23].

- Let me also mention an approach by J. Klauder [318], which more directly addresses quantization than the other procedures where a consistent classical canonical theory is established prior to quantization. Klauder formally constructs from the constraints a projection operator to the reduced phase space in the context of phase-space coherent states. With this, he can treat path integrals without introducing gauge constraints, the Faddeev-Popov determinant, or Dirac brackets. Thus far, only few physically relevant models could be treated explicitly.

C.5 Constraints and Presymplectic Geometry

In this section those changes will be addressed which arise in a geometry-based description for singular systems, as opposed to the description for regular systems in Subsect. 2.1.5.

C.5.1 Legendre Projectability

For singular systems, the Legendre map $\mathcal{F}L$ is no longer an isomorphism of the tangent and the cotangent bundles. There exists a nontrivial kernel for the pullback of $\mathcal{F}L^*$, spanned by the vector fields

$$\Gamma_\rho = \xi^k_\rho \frac{\partial}{\partial \dot{q}^k} \tag{C.118}$$

where the ξ_ρ are zero eigenvectors of the Hessian W; compare with (C.4). The pullback of a function g through $\mathcal{F}L$ is denoted as

$$\mathcal{F}L^* g(q, p) = g\big(q, \hat{p} = \frac{\partial L}{\partial \dot{q}}\big).$$

Since the image $\mathcal{F}L(TQ)$ is the primary constraint surface $\Gamma_P \in T^*Q$, we have $\mathcal{F}L^*(\phi_\rho) = \phi_\rho(q, \hat{p}) \equiv 0$. In differentiating the constraint $\phi_\rho(q, \hat{p}(q, \dot{q}))$ with respect to the velocities, one finds that the vector field components ξ^ρ_k in (C.118) are related to the cotangent bundle primary constraints ϕ_ρ by

$$\xi^i_\rho = \mathcal{F}L^*\big(\frac{\partial \phi_\rho}{\partial p_i}\big), \tag{C.119}$$

a little sloppily stated before as (C.14).

The vector fields Γ_ρ are instrumental in finding out which structures are *Legendre projectable*. For a function $f_L : T\mathbb{Q} \to \mathbb{R}$ the projectability condition reads

$$\Gamma_\rho \cdot \ f_L = 0 \qquad \text{for} \qquad \forall \rho \Leftrightarrow \exists f_H \qquad \text{such that} \qquad \mathcal{F}L^*(f_H) = f_L.$$

In taking for instance the Lagrangian energy E one finds immediately that this is Legendre projectable:

$$\Gamma_\rho \cdot E = \xi^k_\rho \frac{\partial}{\partial \dot{q}^k}\Big(\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L\Big) = \xi^k_\rho \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^k}\dot{q}^i + \frac{\partial L}{\partial \dot{q}^i}\delta^k_i - \frac{\partial L}{\partial \dot{q}^k}\right) = 0.$$

Therefore a Hamiltonian exists such that $\mathcal{F}L^*(H_C) = E$. This is the canonical Hamiltonian in the Rosenfeld-Dirac-Bergmann algorithm. But also the so-called total Hamiltonian $H_T = H_C + u^\rho \phi_\rho$ obeys $\mathcal{F}L^*(H_C) = E$ because of $\mathcal{F}L^*(\phi_\rho) = 0$.

A useful and convenient vector field is the evolution operator K which transforms a cotangent-space function $f(q, p)$ into its time derivative in the tangent space:

$$K \cdot f = \dot{q}^i \ \mathcal{F}L^*\Big(\frac{\partial f}{\partial q^i}\Big) + \frac{\partial L}{\partial q^i}\mathcal{F}L^*\Big(\frac{\partial f}{\partial p_i}\Big) + \mathcal{F}L^*\Big(\frac{\partial f}{\partial t}\Big). \tag{C.120}$$

Since the two-form ω_L (defined by (2.30)) is only presymplectic (closed, but degenerate) the solution of $i_X\omega_L = dE$ is no longer unique. It is true that solutions have the form[24]

$$X_E = \dot{q}^i \frac{\partial}{\partial q^i} + a^i(q, \dot{q})\frac{\partial}{\partial \dot{q}^i}$$

but, contrary to the regular case with (2.28), the functions a^i are not completely determined. Namely if X_E is a solution, so is $X_E + Y$, where the vector field Y belongs to the kernel of ω_L, that is $i_Y\omega_L = 0$. In [424] one finds a derivation that this kernel is provided by Γ_ρ and by

$$\Delta_{\bar{\mu}} = \xi^k_{\bar{\mu}} \frac{\partial}{\partial q^k} + \left(K\big(\frac{\partial \phi_{\bar{\mu}}}{\partial p_k}\big) - \mathcal{F}L^*\big(\frac{\partial \phi'_{\bar{\mu}}}{\partial p_k}\big)\right)\frac{\partial}{\partial \dot{q}^k}, \tag{C.121}$$

where this expression is defined for all indices $\bar{\mu}$ for which the constraints $\phi_{\bar{\mu}}$ are first class. The $\phi'_{\bar{\mu}}$ are secondary constraints: $\phi'_{\bar{\mu}} := \{\phi_{\bar{\mu}}, H\}$.

As demonstrated in [33], the Lagrangian and Hamiltonian dynamics relate to each other by

$$\dot{q}^k = \mathcal{F}L^*\{q^k, H_C\} + v^\mu \mathcal{F}L^*\{q^k, \phi_\mu\} \tag{C.122a}$$

$$\frac{\partial L}{\partial q^k} = \mathcal{F}L^*\{p_k, H_C\} + v^\mu \mathcal{F}L^*\{p_k, \phi_\mu\}. \tag{C.122b}$$

[24] Again assuming the absence of any explicit time dependence.

On applying the operators Γ_ρ to (C.122a), we find that because of (C.119), and due to the fact that $\Gamma_\rho \circ \mathcal{F}L^* = 0$, the functions v^μ are not projectable, but rather $\Gamma_\rho v^\mu = \delta^\mu_\rho$. With the aid of (C.122), the defining equation for the operator K can be written

$$K \cdot f = \mathcal{F}L^*\{f, H_C\} + v^\mu \mathcal{F}L^*\{f, \phi_\mu\} + \mathcal{F}L^*\Big(\frac{\partial f}{\partial t}\Big). \tag{C.123}$$

C.5.2 Symmetry Transformations in the Tangent and Cotangent Bundle

We saw that for regular systems, a constant of motion associated with a Noether symmetry can be understood as a generator of the symmetry in phase space. This ceases to be true in general, since a Noether symmetry transformation may not be projectable onto phase space.

The following exposition essentially follows [34], [421] and [423]. The Noether invariance identity in (2.50) was split as

$$V_i \delta q^i + \dot{q}^i \frac{\partial J_L}{\partial q^i} + \frac{\partial J_L}{\partial t} \equiv 0 \tag{C.124a}$$

$$\frac{\partial J_L}{\partial \dot{q}^i} - W_{ik} \delta q^k \equiv 0. \tag{C.124b}$$

Here, I forego indicating the symmetry transformation by the letter S on δq^k since the following refers to symmetry transformations only. (Also, I assume that the variation is the active one.) Instead, the current is written with a subscript L to indicate that it is a function in $\mathcal{F}(T\mathbb{Q})$. For regular systems, the identity (C.124b) can directly be solved to yield explicit expressions of the infinitesimal symmetry transformations. In the singular case, after multiplying (C.124b) with the zero eigenvectors ξ^ρ one obtains the result that

$$\Gamma_\rho \cdot J_L = 0,$$

and therefore a G exists such that $J_L = \mathcal{F}L^*(G)$. This makes it possible to rewrite (C.124b) as

$$W_{ik}\Big(\mathcal{F}L^*\Big(\frac{\partial G}{\partial p_k}\Big) - \delta q^k\Big) = 0,$$

revealing that the factor multiplying the Hessian must be a linear combination of zero eigenvectors, or

$$\delta q^k = \mathcal{F}L^*\Big(\frac{\partial G}{\partial p_k}\Big) + \xi^k_\sigma\, r^\sigma(q,\dot q,t) = \mathcal{F}L^*\{q^k, G\} + r^\sigma\, \mathcal{F}L^*\{q^k,\phi_\sigma\} \quad \text{(C.125)}$$

with functions $r^\rho(q,\dot q,t)$. We observe that the transformations δq^k are projectable only if $\Gamma_\rho r^\sigma = 0$. If the functions r^σ are not projectable, there is a variational symmetry in the tangent bundle which is not the pull-back of a Hamiltonian symmetry transformation. In the sequel, to keep things simple, only Noether projectable symmetry transformations will be treated; for the general case see [421].

Plugging the expression for δq^k into (C.124a) results in

$$\begin{aligned}\Big(\frac{\partial L}{\partial q^i} - \dot q^k \frac{\partial \hat p_i}{\partial q^k}\Big)\mathcal{F}L^*\Big(\frac{\partial G}{\partial p_i}\Big) + V_i \xi^i_\sigma r^\sigma + \dot q^i \frac{\partial J_L}{\partial q^i} + \frac{\partial J_L}{\partial t}&\\ = \frac{\partial L}{\partial q^i}\mathcal{F}L^*\Big(\frac{\partial G}{\partial p_i}\Big) + \dot q^i \mathcal{F}L^*\Big(\frac{\partial G}{\partial q^i}\Big) + \mathcal{F}L^*\Big(\frac{\partial G}{\partial t}\Big) + r^\sigma \chi_\sigma&\\ = K\cdot G + r^\sigma \chi_\sigma = 0.&\end{aligned}$$

In this expression, the

$$\chi_\sigma := \xi^k_\sigma V_k = K\cdot\phi_\sigma$$

are the primary Lagrange constraints. This can be seen from (C.119) and from

$$0 = \frac{\partial \phi_\mu}{\partial q^i} + \frac{\partial \phi_\mu}{\partial \hat p^k}\frac{\partial \hat p^k}{\partial q^i}, \qquad \text{which is} \qquad \mathcal{F}L^*(\frac{\partial\phi_\mu}{\partial q^i}) = -\xi^k_\mu\, \mathcal{F}L^*(\frac{\partial^2 L}{\partial q^i \partial \dot q^k});$$

i.e.

$$\begin{aligned}K\cdot\phi_\mu &= \dot q^i\ \mathcal{F}L^*\Big(\frac{\partial\phi_\mu}{\partial q^i}\Big) + \frac{\partial L}{\partial q^i}\mathcal{F}L^*\Big(\frac{\partial\phi_\mu}{\partial p_i}\Big)\\ &= \xi^k_\mu\Big(-(\frac{\partial^2 L}{\partial q^i\partial \dot q^k}) + \frac{\partial L}{\partial q^i}\Big) = \xi^k_\mu V_k = \chi_\mu.\end{aligned}$$

Now, in the case of projectability, the condition derived for G to be the generator of a symmetry transformation in phase space can be written as

$$K\cdot(G + r^\sigma\phi_\sigma) = 0. \quad \text{(C.126)}$$

The transformations induced in the tangent bundle can be expressed as

$$\delta q^k = \mathcal{F}L^*\{q^k, \tilde G\} \qquad \text{with} \qquad \tilde G := G + r^\sigma_H \phi_\sigma.$$

Furthermore, it can be shown, that

$$\delta \hat{p}_k = V_{kj}\delta q^j + W_{kj}\delta \dot{q}^j = \mathcal{F}L^*\{q_k, \tilde{G}\} + [L]_j \frac{\partial \delta q^j}{\partial \dot{q}^k};$$

and thus the phase-space induced transformations for the momenta are (in general) "correct" on solutions only. We saw this previously in the example of the relativistic particle. A similar term, depending on the functional dependence of the symmetry transformation was shown to arise in the regular case; see Sect. 2.2.3.
Expressing the action of K by (C.123), the condition (C.126) becomes

$$\mathcal{F}L^*\{\tilde{G}, H_C\} + v^\mu \mathcal{F}L^*\{\tilde{G}, \phi_\mu\} + \mathcal{F}L^*\left(\frac{\partial \tilde{G}}{\partial t}\right) = 0.$$

Acting on this with Γ_ρ, yields for projectable transformations

$$\mathcal{F}L^*\{\tilde{G}, H_C\} + \mathcal{F}L^*\left(\frac{\partial \tilde{G}}{\partial t}\right) = 0, \qquad \mathcal{F}L^*\{\tilde{G}, \phi_\mu\} = 0.$$

This is equivalent to

$$\{G, H_C\} + \frac{\partial G}{\partial t} = PC \qquad \{G, \phi_\mu\} = PC.$$

These conditions for relating Lagrangian to Hamiltonian symmetries hold both for global and local transformations. For local transformations they lead to detailed conditions for the coefficients G_k in the expansion $G(\epsilon) = \sum \epsilon^{(k)} G_k$; compare (C.38).

I should mention that a related geometry-based relationship of Noether symmetry transformations and infinitesimal canonical transformations by first-class constraints is derived in [338].

The Algebra of Symmetry Transformations

Next, let us compare the commutator algebra of transformations in configuration space with the Poisson bracket algebra of generators in phase space. Based on previous findings the transformations are related to each other by

$$\delta^L q^k = \mathcal{F}L^*(\delta^H q^k) \qquad\qquad \delta^L \hat{p}_k = \mathcal{F}L^*(\delta^H p_k) + [L]_j \frac{\partial \delta q^j}{\partial \dot{q}^k}.$$

For the commutator of two Hamiltonian transformations the Jacobi identity implies

$$[\delta_1^H, \delta_2^H] q^k = \{q^k\{G_2, G_1\}\}.$$

And the commutator of two Lagrangian transformations is calculated to be

$$[\delta_1^L, \delta_2^L]q^k = \mathcal{F}L^*(\{q^k\{G_2, G_1\}\}) - [L]_i W_{jl} \left(\frac{\partial^2 G_1}{\partial p_i \partial p_l} \frac{\partial^2 G_2}{\partial p_k \partial p_j} - \frac{\partial^2 G_2}{\partial p_i \partial p_l} \frac{\partial^2 G_1}{\partial p_k \partial p_j} \right).$$

The extra term is an antisymmetric combination of the equations of motion, called "trivial" transformation in Sect. C.1.4. This term vanishes off-shell only if (one of) the generators is linear in the momenta[25]. Typically this is the case for Yang-Mills type theories, but in reparametrization-invariant theories the generators always contain the Hamiltonian constraint $\mathcal{H}_\perp$, which is quadratic in the momenta; see (C.116) and the expressions immediately following.

C.5.3 Constraint Stabilization

The Rosenfeld-Dirac-Bergmann algorithm was formulated in purely geometric terms by M. J. Gotay, J. M. Nester and G. Hinds [234], for its essence see also [74], [491]. They start from the observation that the Legendre map $\mathcal{F}L$ is no longer a diffeomorphism, but a map into $T^*\mathbb{Q}$. The image $\mathcal{F}L(T\mathbb{Q})$ of $T\mathbb{Q}$ under $\mathcal{F}L$ may be assumed to be a submanifold $\Gamma_P \in T^*\mathbb{Q}$. However, in general it is not a symplectic manifold but merely presymplectic (closed, but degenerate). The two-form ω_P inherited from the inclusion map $j : \mathcal{F}L(T\mathbb{Q}) \to T^*\mathbb{Q}$ as $j^*\omega = \omega_P$ gives rise to the equation

$$(i_X \omega_P - dH_C)\big|_{\Gamma_P} = 0.$$

This looks like a Hamilton equation, but does not have a unique solution for the vector field X. It serves as the starting point of an algorithm in which successively the constraint surfaces are refined.

Is there an inherent meaning of primary/secondary, of first-/second-class constraint classification and of the Dirac bracket? As demonstrated in [421], the first- and second-class classification distinguishes projectable and non-projectable primary Lagrange constraints. As for the Dirac bracket: Assume that the second-class constraints define locally a submanifold $\bar{\Gamma}$ of the full phase space $T^*\mathbb{Q}$. The inclusion map $\bar{j} : \bar{\Gamma} \to T^*\mathbb{Q}$ induces on $\bar{\Gamma}$ a two-form $\bar{\omega} = \bar{j}^*\omega$. It can be shown that $\bar{\omega}$ is a symplectic form on $\bar{\Gamma}$ and that the Dirac bracket is the Poisson bracket with respect to this symplectic form.

[25] This result was also noted in [374].

Appendix D
*Symmetries in Path-Integral and BRST Quantization

The path-integral formulation of quantum theories goes back to ideas of P.A.M. Dirac in the 1930's, and it was elaborated by R. Feynman in the 1940's within his doctoral thesis, see [180]. At first, path-integral techniques looked like just another approach to quantum physics, but today, specifically when it comes to quantum field theory, they are judged to be the most appropriate means to tackle conceptual issues. Path-integral methods allow one to directly determine Lorentz covariant Feynman rules of a theory from its Lagrangian. This became feasible also for non-AbeIian gauge theories and general relativity after the work of L. Faddeev and V. Popov in the 1960's. And G. 't Hooft's derivation of the Feynman rules for spontaneously broken gauge theories was likewise carried out in the path-integral formulation[1]. Today, with the better understanding of the BRST symmetries, any theory based on a classical or a pseudo-classical action (involving also fermionic variables in the sense of Appendix B.2.4) can be treated by path integrals.

The intuitive ideas on path integrals do come from Feynman's thought experiments about interference patterns in double-slit configurations. These led to postulates as follows

- The probability of a particle moving from point a to point b is given by the absolute square of the complex transition function $K(b, a)$: $P(b, a) = |K(b, a)|^2$.
- The transition function obeys $K(c, a) = \sum K(c, b)K(b, a)$ where the sum is taken over all path connecting a and c, as well all intermediate points b.
- The transition function is given by a sum of phase factors as

$$K(b, a) = \sum_{paths(ab)} \gamma e^{iS(path)/\hbar} \tag{D.1}$$

  where S is the classical action, evaluated along a path, and the sum extends over all paths connecting a and b. The normalization constant γ can be determined from the second postulate.

[1] See [581] for a topical point of view.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

We expect a transition from quantum mechanics to classical mechanics by taking the limit $\hbar \to 0$. All those paths for which the action is large are suppressed by large fluctuations of the phase. The paths that dominate the sum are those for which the action is small. In the limit $\hbar \to 0$ the path with the smallest value of S becomes relevant. But this is just the classical trajectory. In this sense the "sum-over-histories" approach offers an explanation why nature (classically) obeys extremum principles.

D.1 Basics

D.1.1 Path-Integral Formulation of Quantum Mechanics

We begin by calculating the probability amplitude for a process in which a particle starts from the position q' at time t' and is found at q'' at some later time t''. In the Schrödinger picture, the time evolution of states is described by $|q, t\rangle = e^{iHt}|q\rangle$. Thus

$$\langle q'', t''|q', t'\rangle = \langle q''|e^{-iH(t''-t')}|q'\rangle.$$

Let us now see in which sense this probability amplitude is evaluated by summing over all possible paths that connect (q', t') and (q'', t''). First, divide the time interval $(t'' - t')$ into equidistant segments:

$$\delta t = t_{n+1} - t_n = \epsilon \qquad t_0 = t' \qquad t_N = t''.$$

Sandwich the completeness relation $\int dq \ |q, t\rangle\langle q, t| = 1$ $(N\text{-}1)$ times into the amplitude:

$$\begin{aligned}\langle q'', t''|q', t'\rangle =& \int dq_1 \dots dq_{N-1} \langle q'', t''|q_{N-1}, t_{N-1}\rangle \dots \langle q_2, t_2|q_1, t_1\rangle\langle q_1, t_1|q', t'\rangle \\ =& \prod_{n=1}^{N-1} \int dq_n \ \langle q''|e^{-iH\epsilon}|q_(N-1)\rangle \ \dots \ \langle q_2|e^{-iH\epsilon}|q_1\rangle\langle q_1|e^{-iH\epsilon}|q'\rangle.\end{aligned} \tag{D.2}$$

Now take the most simple Hamiltonian imaginable, namely the one for a free relativistic particle: $H = \hat{p}^2/2\ m$. Consider any one of the inner products $\langle q_{n+1}| \exp\{-i\epsilon\hat{p}^2/2\ m\}|q_n\rangle$; insert another complete set of states according to $\int \frac{dp}{2\pi}\ |p\rangle\langle p| = 1$ to arrive at

$$\begin{aligned}\langle q_{n+1}|\exp\{-i\epsilon\frac{\hat{p}^2}{2\,m}\}|q_n\rangle &= \int \frac{dp}{2\pi}\langle q_{n+1}|\exp\{-i\epsilon\frac{\hat{p}^2}{2\,m}\}|p\rangle\langle p|q_n\rangle \\ &= \int \frac{dp}{2\pi}\exp\{-i\epsilon\frac{p^2}{2\,m}\}\langle q_{n+1}|p\rangle\langle p|q_n\rangle \\ &= \int \frac{dp}{2\pi}\exp\{-i\epsilon\frac{p^2}{2\,m}\}\exp\{ip(q_{n+1}-q_n)\}.\end{aligned} \tag{D.3}$$

The last step in this manipulation made use of $\langle q|p\rangle = e^{ipq}$. The resulting expression for each of the inner products is a Gaussian integral with respect to p.

Let me next give the Gaussian integral formula in a quite general form: For a real, symmetric $N \times N$ matrix A_{ij} with positive eigenvalues, and for a vector x_i

$$\prod_{k=1}^{N}\int_{-\infty}^{+\infty} dx_k\; e^{-\frac{1}{2}x\cdot A\cdot x + J\cdot x} = \left(\frac{(2\pi)^N}{\det A}\right)^{1/2} e^{\frac{1}{2}J\cdot A^{-1}\cdot J}. \tag{D.4}$$

Assume that this Gauss integration holds also for imaginary A and J, and formally replace $A \to iA$, $J \to iJ$. Then

$$\prod_{k=1}^{N}\int_{-\infty}^{+\infty} dx_k\; e^{-\frac{i}{2}x\cdot A\cdot x + iJ\cdot x} = \left(\frac{(-2\pi i)^N}{\det A}\right)^{1/2} e^{-\frac{1}{2}J\cdot A^{-1}\cdot J}. \tag{D.5}$$

This assumption has to be taken with caution. One may be more careful by either taking formally $(i\epsilon)$ to be real, or by putting in by hand a factor $e^{-\zeta p^2}$ into (D.3), with the option $\zeta \to 0$ at the end of the calculation. In any case, identifying in (D.3) $A = \epsilon/m$ and $J = (q_{n+1} - q_n)$ the p integration yields with (D.5)

$$\langle q_{n+1}|\exp\{-i\epsilon\frac{\hat{p}^2}{2\,m}\}|q_n\rangle = \frac{1}{2\pi}\left(\frac{2\pi m}{i\epsilon}\right)^{1/2}\exp\{\frac{im}{2\epsilon}(q_{n+1}-q_n)^2\}.$$

Inserting these results into (D.2), we finally arrive at

$$\langle q''|e^{-iH(t''-t')}|q'\rangle = \left(\frac{-im}{2\pi\epsilon}\right)^{N/2}\prod_{n=1}^{N-1}\int dq_n\;\exp\{i\epsilon(m/2)\sum_{k=1}^{N}[(q_{k+1}-q_k)/\epsilon]^2\}.$$

Letting $\epsilon = \delta t \to 0$ we can replace

$$[(q_{k+1}-q_k)^2/\delta t]^2 \;\to\; \dot{q}^2 \qquad \delta t\sum_{k=1}^{N-1} \;\to\; \int_{t'}^{t''} dt.$$

Furthermore, define the integral over paths as

$$\int \mathcal{D}q = \lim_{N\to\infty} \Big(\frac{-imN}{2\pi(t''-t')}\Big)^{N/2} \Big(\prod_{n=1}^{N-1} \int dq_n\Big), \tag{D.6}$$

so that the previous result can be written

$$\langle q''|e^{-iH(t''-t')}|q'\rangle = \int \mathcal{D}q \exp\Big\{i\int_{t'}^{t''} dt\ \frac{1}{2}m\dot{q}^2\Big\}.$$

The expression (D.6) for the measure $\mathcal{D}q$ needs some getting used to. Observe that it is an integral over functions $q_n(t)$, therefore instead of the physically motivated "path integral", one is mathematically dealing with a "functional integral".
If instead of the Hamiltonian for the free particle, one takes the more general $H = \hat{p}^2/2\,m + V(\hat{q})$, the transition amplitude would become

$$\langle q''|e^{-iH(t''-t')}|q'\rangle = \int \mathcal{D}q \exp\Big\{i\int_{t'}^{t''} dt\ [\frac{1}{2}m\dot{q}^2 - V(q)]\Big\} = \int \mathcal{D}q\ e^{iS(q,\dot{q})}. \tag{D.7}$$

This then demonstrates explicitly what is meant by the sum over all paths in (D.1). Compared to (D.1), there is a factor $1/\hbar$ missing in the exponential. But not only must it be there for dimensional reasons, we also can recover it from the calculations by noticing that the correct normalization condition is $\langle q|p\rangle = e^{ipq/\hbar}$ entailing the completeness relation $\int \frac{dp}{2\pi\hbar}\ |p\rangle\langle p| = 1$; compare with (4.15).

The expression (D.7) contains the action $S(q,\dot{q})$ in its Lagrangian form. But we see more or less immediately that the transition to the Hamiltonian form $S(q, p)$ is simply mediated by a path integral in momentum space, namely

$$\begin{aligned}\int \mathcal{D}q e^{iS} &= \int \mathcal{D}q \exp\Big\{i\int_{t'}^{t''} dt\ [\frac{1}{2}m\dot{q}^2 - V(q)]\Big\}\\ &= \int \mathcal{D}q\mathcal{D}p \exp\Big\{i\int_{t'}^{t''} dt\ [p\dot{q} - \frac{p^2}{2\,m} - V(q)]\Big\}\\ &= \int \mathcal{D}q\mathcal{D}p \exp\Big\{i\int_{t'}^{t''} dt\ [p\dot{q}(q,p) - H(p,q)]\Big\}.\end{aligned}$$

The exact procedure is to insert into the configuration space integral a sequence of N Gaussian integrations over p_k by using (D.4) and to reabsorb the constants in the limit $N\to\infty$ in the definition of $\mathcal{D}p$, similar to (D.6). Is it not intriguing to recover the Legendre transformation as a p-integration in the path integral? (This, however, only holds for Hamiltonians quadratic in the momenta.)

Now consider instead of the previous transition amplitude the more general $\langle q'', t''|\hat{q}(t_1)|q', t'\rangle$ (where $t' < t_1 < t''$):

$$\langle q'', t''|\hat{q}(t_1)|q', t'\rangle = \langle q'', t''|e^{-iH(t'-t_1)}\hat{q}e^{-iH(t_1-t'')}|q', t'\rangle = \int \mathcal{D}q\mathcal{D}p\, q(t_1)e^{iS[q,p]}.$$

Moreover, one finds that

$$\int \mathcal{D}q\mathcal{D}p\ q(t_1)q(t_2)e^{iS[q,p]} = \langle q'', t''|P\hat{q}(t_1)\hat{q}(t_2)|q', t'\rangle,$$

where P denotes time-ordering:

$$P(\hat{q}(t_k)\hat{q}(t_l)) = \begin{cases} \hat{q}(t_k)\hat{q}(t_l) & \text{for} \quad t_k < t_l, \\ \hat{q}(t_l)\hat{q}(t_k) & \text{for} \quad t_k > t_l. \end{cases}$$

A further notion is the *generating functional*

$$\langle q'', t''|q', t'\rangle_J = \int \mathcal{D}q\ \exp\left\{i \int dt\ [L(q, \dot{q}) + Jq]\right\}$$

with its functional derivatives

$$\frac{\delta}{\delta J(t_1)}\langle q'', t''|q', t'\rangle_J = i \int \mathcal{D}q\mathcal{D}p\ q(t_1) \exp\left\{i \int dt\ [L(q, \dot{q}) + Jq]\right\}$$

$$\frac{\delta}{\delta J(t_1)} \frac{\delta}{\delta J(t_2)}\langle q'', t''|q', t'\rangle_J = (i)^2 \int \mathcal{D}q\mathcal{D}p\, q(t_1)q(t_2) \exp\left\{i \int dt\, [L(q, \dot{q}) + Jq]\right\}.$$

From this, one observes that the generating functional yields

$$\langle q'', t''|P(\hat{q}(t_1)\ldots\hat{q}(t_n))|q', t'\rangle = (-i)^n \frac{\delta^n}{\delta J(t_1)\ldots\delta J(t_n)}\langle q'', t''|q', t'\rangle_J \bigg|_{J=0}. \tag{D.8}$$

In general, one might be interested in the transition of an initial state $|I\rangle$ (not necessarily a position) to a final state $|F\rangle$. This is

$$\langle F, t''|I, t'\rangle = \int dq''\ dq'\ \langle F|q''\rangle\langle q''|e^{-iH(t''-t')}|q'\rangle\langle q'|I\rangle.$$

A special name is given to the transition amplitude in case $|I\rangle$ and $|F\rangle$ are the ground state and t' and t'' refer to the infinite past and infinite future, respectively:

$$\langle 0|0\rangle = \lim_{t'\to-\infty}\ \lim_{t''\to+\infty} \langle 0, t''|0, t'\rangle.$$

This cumbersome expression can be avoided by replacing H by $(1 - i\epsilon)H$ (see e.g. [481], Sect. 6; [431], Sect. 2.3) Define

$$Z[J] := \int \mathcal{D}q\mathcal{D}p \exp\left\{i \int_{-\infty}^{+\infty} dt\ [p\dot{q} - (1 - i\epsilon)H(p, q) + Jq]\right\};$$

then $\langle 0|0\rangle = Z[0]$. In subsequent expressions the $(i\epsilon)$ will often be omitted unless it is needed for explicit calculations.

Example: Harmonic Oscillator

Let me state some results for the harmonic oscillator[2] which exhibits very many features of functional integrals for relativistic field theories. The Hamiltonian is

$$H(p,q) = \frac{1}{2\,m}p^2 + \frac{1}{2}m\omega^2 q^2.$$

Here the (1-$i\epsilon$) prescription amounts to the replacement $m \to (1+i\epsilon)m$ and $m\omega^2 \to (1-i\epsilon)m\omega^2$, and thus, after switching to the Lagrangian:

$$Z[J] = \int \mathcal{D}q \exp\Big\{i\int_{-\infty}^{+\infty} dt\ [\frac{1}{2}(1+i\epsilon)m\dot{q}^2 - \frac{1}{2}(1-i\epsilon)m\omega^2 q^2 + Jq]\Big\}.$$

Here, we directly see the physical meaning of the Jq term as an external force on the oscillator. In going over to the Fourier components, e.g.

$$\tilde{J}(E) := \frac{1}{\sqrt{2\pi}}\int dt\ J(t)e^{-iEt},$$

the generating functional can be expressed as (from now on $m = 1$)

$$\begin{aligned} Z[J] &= Z(0)\exp\Big\{-\frac{i}{2}\int_{-\infty}^{+\infty} dE\ \frac{\tilde{J}(E)\tilde{J}(-E)}{E^2-\omega^2+i\epsilon}\Big\} \\ &= Z(0)\exp\Big\{-\frac{i}{2}\int_{-\infty}^{+\infty} dt dt'\ J(t)G(t-t')J(t')\Big\}. \end{aligned}$$

Here,

$$G(t-t') = \int_{-\infty}^{+\infty}\frac{dE}{2\pi}\ \frac{e^{-i(t-t')E}}{E^2-\omega^2+i\epsilon} = \frac{1}{2i\omega}[\theta(t)e^{-i\omega t} + \theta(-t)e^{i\omega t}]$$

is the Green's function for the operator $\left(\frac{d^2}{dt^2}+\omega^2\right)$:

$$\Big(\frac{d^2}{dt^2}+\omega^2\Big)G(t) = -\delta(t).$$

[2] Details in e.g. [481], Sect. 7.

We may also verify the relations between the Green's function and the time ordered products (D.8). Consider for instance the two-point function

$$\begin{aligned}\langle 0|P(\hat{q}(t_1)\hat{q}(t_2)|0\rangle &= \frac{1}{i}\frac{\delta}{\delta J(t_1)}\frac{1}{i}\frac{\delta}{\delta J(t_2)}\langle 0|0\rangle_J\Big|_{J=0}\\ &= \frac{1}{i}\frac{\delta}{\delta J(t_1)}\left[\int_{-\infty}^{+\infty} dt dt'\, G(t_2-t')J(t')\right]Z[J]\Big|_{J=0}\\ &= \left[\frac{1}{i}G(t_2-t_1) + (\text{terms in } J)\right]Z[J]\Big|_{J=0} = -iG(t_2-t_1).\end{aligned}$$

This can be generalized to express the n-point function in terms of the Green's function.

D.1.2 Functional Integrals in Field Theory

The path integral can be immediately extended to more than just one degree of freedom and it can be written in a characteristic way as

$$\int \mathcal{D}q e^{iS(q,\dot{q})} = \int \Big(\prod_{k=1}^{N} \mathcal{D}q_k\Big) e^{iS(q_k,\dot{q}_k)},$$

where again $S(q_k, \dot{q_k})$ is the classical action and the path integral extends over all degrees of freedom.

The next step is to generalize the path-integral description to a field theory. The fields in fundamental physics are either bosonic (with spin 0, 1, 2) or fermionic (with spin $\frac{1}{2}$, $\frac{3}{2}$).

Bosonic Fields

Let us begin with the free scalar field theory described by the Lagrangian

$$\mathcal{L}_f = \frac{1}{2}(\partial_\mu\varphi\partial^\mu\varphi - m^2\varphi^2).$$

The generating functional is defined as

$$Z_f[J] = \int \mathcal{D}\varphi\ \exp i\int d^4x[\mathcal{L}_f(\varphi,\partial\varphi) + J\varphi], \tag{D.9}$$

where the $J(x)$ denotes an arbitrary classical source field. The symbol $\mathcal{D}\varphi$ indicates that at each point of spacetime, one integrates over all possible values of the fields $\varphi(x)$. By defining the Fourier components as

$$\tilde{J}(k) := \frac{1}{(2\pi)^2} \int d^4x \; J(x) e^{-ikx}$$

one can essentially mimic the calculations of the one-dimensional harmonic oscillator to arrive at

$$Z_f[J] = Z_f(0) \exp\Big\{ -\frac{i}{2} \int_{-\infty}^{+\infty} dx dy \; J(x) \triangle_F(x-y) J(y) \Big\}. \tag{D.10}$$

Here,

$$\triangle_F(x-y) = \int_{-\infty}^{+\infty} \frac{d^4k}{(2\pi)^4} \frac{e^{-ik(x-y)}}{k^2 - m^2 + i\epsilon}$$

is the Green's function for the Klein-Gordon operator

$$-(\partial_\mu \partial^\mu + m^2) \triangle_F(x) = \delta^4(x). \tag{D.11}$$

This result can be obtained formally also in the following way: Write the derivative term in the Lagrangian as $(\partial\varphi)^2 = \partial(\varphi\partial\varphi) - \varphi\, \partial^2\varphi$. Then (D.9) becomes–up to a boundary term–

$$Z_f[J] = \int \mathcal{D}\varphi \; \exp i \int d^4x [-\frac{1}{2} \varphi(\partial_\mu \partial^\mu + m^2)\varphi + J\varphi].$$

This is a Gaussian integral (D.5) with the identification $A = -(\partial_\mu \partial^\mu + m^2)$. If we baldly extend the Gaussian integral from the case of finite discrete integration to the continuous case, we need to identify the inverse of the Klein-Gordon operator A. But, according to (D.11) this is none other than $\triangle_F(x)$, and we arrive directly back at (D.10). Observe that the normalization of Z is formally identical to

$$Z[0] = [\det(\partial^2 + m^2)]^{-1/2},$$

a quantity which is not well-defined, but also not needed in explicit calculations of physically-measurable quantities such as cross sections and decay rates. This is because we always calculate vacuum expectation values

$$\langle \hat{O} \rangle := \frac{\langle 0 | \hat{O} | 0 \rangle}{\langle 0 | 0 \rangle},$$

and here, the dependence on $\langle 0|0\rangle = Z[0]$ cancels out because it is in both the numerator and the denominator.

Similarly to the case of the harmonic oscillator

$$\langle 0|P(\hat{\varphi}(x_1)(\hat{\varphi}(x_2))|0\rangle = -i\triangle(x_2 - x_1),$$

and (D.8) generalizes to

$$\langle 0|P(\hat{\varphi}(x_1)\ldots\hat{\varphi}(x_n))|0\rangle = (-i)^n \frac{\delta^n}{\delta J(x_1)\ldots\delta J(x_n)} Z[J]\Big|_{J=0}. \tag{D.12}$$

This reproduces the techniques developed in the early 1950's by J. Schwinger of "probing" a quantum field theory with a varying external source.

Extending the free scalar theory by an interaction Lagrangian term to the action $S = \int d^4x\ \mathcal{L}_f + \int d^4x\ \mathcal{L}_{int}$, we may write

$$\exp i(S_f + S_{int}) = e^{iS_f} \sum_{n=0}^{\infty} \frac{i^n}{n!} (\mathcal{L}_{int})^n.$$

After expanding $\mathcal{L}_{int}$ in terms of φ, the path integral for the full theory is a sum of terms proportional to

$$\int \mathcal{D}\varphi\ e^{iS_f} \varphi(x_1)\ldots\varphi(x_k).$$

These are related to the k-point functions which can be symbolized by Feynman diagrams and associated rules. For details, see again any book on QFT.

The previous considerations for a scalar field can in principle be extended to massive spin-1 and spin-2 fields (let them be called generically Φ) by rewriting the free Lagrangian in the form $\Phi\mathcal{O}\Phi$ with a differential operator $\mathcal{O}$, then performing a Gaussian integration with respect to Φ, identifying the inverse of $\mathcal{O}$ with its Green's function. For instance for the Proca theory with

$$\begin{aligned}\mathcal{L}_P &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu = \frac{1}{2}A_\mu\big[(\partial_\lambda\partial^\lambda + m^2)\eta^{\mu\nu} - \partial^\mu\partial^\nu\big]A_\nu + b.t.\\ &= A_\mu \mathcal{O}^{\mu\nu} A_\nu + b.t.\end{aligned}$$

the inverse of $\mathcal{O}^{\mu\nu}(x)$ is given by

$$D_{\nu\lambda}(x) = \int \frac{d^4k}{(2\pi)^2} D_{\nu\lambda}(k) e^{ikx} \quad \text{and} \quad D_{\nu\lambda}(k) = \frac{-\eta_{\nu\lambda} + k_\nu k_\lambda/m^2}{k^2 - m^2}. \tag{D.13}$$

Observe that for the Maxwell theory with $m^2 = 0$, the inverse of $[(\partial\partial)\ \eta^{\mu\nu} - \partial^\mu\partial^\nu]$ does not exist. Since the spin-1 and spin-2 fields of fundamental physics are gauge bosons–and as such massless–they need a special treatment, for instance by the Faddeev-Popov method. Before giving an example of this, a few remarks about fermionic fields and functional integrals are in order.

Fermionic Fields

For fermionic fields $\psi(x)$, the generating functional can be defined by using Grassmann variables and be manipulated by integration over Grassmann variables (see Appendix B.2):

$$Z_F[\eta, \overline{\eta}] = N \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{i\{\mathcal{L}_F(x)+\overline{\eta}\psi+\bar{\psi}\eta\}}$$

with the external currents η and $\overline{\eta}$, where $\mathcal{L}_F$ might for instance be the Dirac Lagrangian $\mathcal{L}_D = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi$. This can be integrated formally by using (B.41), resulting in

$$Z_D[\eta, \overline{\eta}] = \exp\left[i \int d^4x d^4y\ \overline{\eta}(x) S(x-y)\eta(y)\right],$$

where $S(x)$ is the inverse of the Dirac operator, that is $(i\gamma^\mu\partial_\mu - m)S(x) = \delta^4(x)$, which yields the Feynman propagator for the Dirac field:

$$S_{\alpha\beta}(x-y) = i\langle 0|P\psi_\alpha(x)\bar{\psi}_\beta(y)|0\rangle.$$

D.1.3 Faddeev-Popov Ghost Fields in Theories with Local Symmetries

Yang-Mills Theories

Consider a Yang-Mills theory gauge theory of fermions, starting from

$$Z = \int \mathcal{D}A\ \mathcal{D}\bar{\psi}\ \mathcal{D}\psi e^{i\{S_{YM}[A]+S_F[A,\psi]\}} \tag{D.14}$$

with the Yang-Mills action and the minimally-coupled fermion action

$$S_{YM}[A] = -\frac{1}{4\,g^2}\int d^4x F^a_{\mu\nu}F^{a\mu\nu}, \quad S_F[A,\psi] = i\bar{\psi}\gamma^\mu(\partial_\mu + iA^a_\mu T_a)\psi - m\bar{\psi}\psi.$$

Here, $\mathcal{D}[A]$ symbolizes the rather cumbersome measure $\prod_{\mu,a,x} dA^a_\mu(x)$. As it stands, the expression (D.14) for Z is infinite, because in integrating over all fields one overcounts by picking contributions from all those fields which are related to each other by a gauge transformation

$$A^\Omega_\mu = \Omega A_\mu \Omega^{-1} + i(\partial_\mu\Omega)\Omega^{-1};$$

compare (5.61). Hence, in a first step one must "divide out" this infinite factor stemming from the gauge-orbit volume. A formal method developed by L. Faddeev and V. Popov [171] is based on path-integral adapted gauge fixing. Write the gauge fixing conditions generically as $G[A] = 0$ (here I completely dispense with indices, both for the gauge fields and for enumerating the gauge fixing conditions). Off course it is possible to insert a delta-function $\delta(G[A])$ into the path integral, but this would change the measure of integration (in the sense of (D.16)). Instead, define the *Faddeev-Popov determinant* $\triangle^G_{FP}(A)$ implicitly by

$$1 = \triangle^G_{FP}(A) \int \mathcal{D}\Omega \; \delta(G[A^\Omega]). \tag{D.15}$$

Here $\mathcal{D}\Omega = \prod_x d\Omega(x)$ is the invariant group measure obeying $\mathcal{D}\Omega = \mathcal{D}(\Omega'\Omega)$. This so-called Haar measure exists for compact groups, see e.g. [254]. The Faddeev-Popov determinant depends apparently on the gauge condition, but actually, it is gauge invariant: Calculate $\triangle^G_{FP}$ for a gauge-shifted field as

$$(\triangle^G_{FP})^{-1}(A^{\Omega'}) = \int \mathcal{D}\Omega' \, \delta(G[A^{\Omega'\Omega}]) = \int \mathcal{D}(\Omega'\Omega) \, \delta(G[A^{\Omega'\Omega}]) = (\triangle^G_{FP})^{-1}(A).$$

Insert the "1" defined in terms of (D.15) into (D.14):

$$Z = \int \mathcal{D}A \; \mathcal{D}\bar{\psi} \; \mathcal{D}\psi \Big(\triangle^G_{FP} \int \mathcal{D}\Omega \; \delta(G[A^\Omega]) \Big) e^{i\{S_{YM}[A] + S_F[A,\psi]\}}.$$

Now perform a gauge transformation $A \to A^{\Omega^{-1}}$ on this path integral. At first one might ask whether the measure is invariant. Under an infinitesimal gauge transformation $(A^a_\mu)_\epsilon = A^a_\mu + \partial_\mu \epsilon^a - f^{abc} A^c_\mu \epsilon^b$, the measure with respect to the vector fields changes as

$$(\mathcal{D}A)_\epsilon = \prod_{\mu,a,x} (dA^a_\mu)_\epsilon = \mathcal{J}_\epsilon \prod_{\mu,a,x} dA^a_\mu$$

where $\mathcal{J}_\epsilon$ is the determinant of the Jacobian

$$\frac{\delta (A^a_\mu)_\epsilon}{\delta A^b_\nu} = \delta^4(x-y) \delta^\nu_\mu \big[\delta^a_b - f^a{}_{bc} \epsilon^c \big].$$

To first order in ϵ, we have $\det \mathcal{J}_\epsilon = 1 - \mathrm{tr}(g f^a{}_{bc} \epsilon^c)$, and this is unity because of the antisymmetry of the structure constants. Let us assume that also the fermionic part $\mathcal{D}\bar{\psi}\mathcal{D}\psi$ is invariant, although some intricacies are encountered in the verification. These will be dealt with in Sect. D.2.4. The part involving the action is of course invariant, and so is the Faddeev-Popov determinant. The factor $G[A^\Omega]$ simply changes into $G[A]$. Therefore, in

$$Z = \left(\int \mathcal{D}\Omega \right) \int \mathcal{D}A \ \mathcal{D}\bar{\psi} \ \mathcal{D}\psi \triangle_{FP}^{G} \ \delta(G[A]) e^{i\{S_{YM}[A]+S_F[A,\psi]\}}$$

the (infinite) group volume $\int \mathcal{D}\Omega$ can be factored out. Since we already saw that a factor which does not depend on the fields cancels out from vacuum expectation values, we may simply omit it and there is no longer the problem of overcounting. Observe that the gauge condition is embedded in the path integral in form of a delta function, but that the Faddeev-Popov procedure provides the correct measure. Thus far, however, the Faddeev-Popov determinant is only defined implicitly by (D.15). In this expression, $\delta(G[A^\Omega])$ appears. Remember that the delta function relation for a function $f(x)$ in one variable with $f(a) = 0$ (assume that there is only one such a) obeys

$$\delta(f(x)) = \frac{\delta(x-a)}{f'(x)}.$$

This can be generalized to the functionals $G[A^\Omega]$ that are zero for $\Omega = \Omega_0$, as:

$$\delta(G[A^\Omega]) = \delta(\Omega - \Omega_0) \det \left| \frac{\delta G[A^\Omega](x)}{\delta\Omega(x')} \right|^{-1}. \tag{D.16}$$

Write the gauge-field transformations in their infinitesimal form $(\Omega = 1 + i\epsilon^a T^a)$ and expand $G[A^\Omega]$ to give:

$$G[A^\Omega(x)] = G[A(x)] + \int d^4y \ [M^G(x,y)]\epsilon(y) + \mathcal{O}(\epsilon^2). \tag{D.17}$$

The derivative of this functional with respect to the $\epsilon's$ is simply the determinant of M^G. Therefore

$$\triangle_{\text{FP}}^{G} = \det M^G = \int \mathcal{D}\eta \ \mathcal{D}\tilde{\eta} \ \exp i \Big(\int d^4x \ d^4y \ \tilde{\eta}(x) M^G(x,y) \eta(y) \Big), \tag{D.18}$$

where in the last step the *Faddeev-Popov ghost fields*[3] η and $\tilde{\eta}$ were introduced in order to convert the Faddeev-Popov determinant into a Gaussian integral. Finally, also the gauge fixing term $\delta(G[A])$ can be promoted to an exponential, with the result

$$Z = \int \mathcal{D}A \ \mathcal{D}\bar{\psi} \ \mathcal{D}\psi \ \mathcal{D}\gamma \ \mathcal{D}\eta \ \mathcal{D}\tilde{\eta} e^{i\{S_{YM}[A]+S_F[A,\psi]+S_{\text{gf}}+S_{\text{ghost}}\}}, \tag{D.19}$$

[3] These are called ghosts since they are Grassmann even fields with the wrong, namely Fermi-Dirac, statistics. They were introduced into theoretical physics already by R. Feynman [182] and B. DeWitt [121].

and with the gauge fixing and the ghost action terms

$$S_{\text{gf}} = \int d^4x \;\; \gamma^a G^a[A] \tag{D.20a}$$

$$S_{\text{ghost}} = \int d^4x \;\; d^4y \;\; \tilde{\eta}^a(x) M^G_{ab}(x, y)\eta^b(y). \tag{D.20b}$$

To get the final message after all of these formal manipulations: The generating functional (D.19) defines a quantum field theory with fermions ψ, gauge bosons A, ghosts η, $\tilde{\eta}$ and the fields γ. The "rules of the game" allow us to derive the Feynman graphs in which all of these entities participate, order by order. The diagrams and the calculational recipes involved with them depend on the chosen gauge. For instance, in the Lorenz gauge $G^a \cong L^a = \partial^\mu A^a_\mu$ the Faddeev-Popov matrix is calculated starting from (D.17)

$$\begin{aligned}
L^a[A(\epsilon)] &= \partial^\mu A^a_\mu(\epsilon) = \partial^\mu(A^a_\mu + \partial_\mu \epsilon^a - f^{abc} A^b_\mu \epsilon^b) \\
&= \partial^\mu A^a_\mu + (\delta^a_b \partial^\mu \partial_\mu - f^{abc} A^c_\mu \partial^\mu)\epsilon^b \\
&= L^a[A] + (M^L)^a{}_b \epsilon^b = L^a[A] + \int d^4y \; (M^L(y))^a{}_b \delta^4(x - y)\epsilon^b,
\end{aligned}$$

from which

$$(M^L(x, y))^a{}_b = \partial^\mu(\partial_\mu \delta^{ab} - f^{abc} A^c_\mu)\delta^4(x - y),$$

such that in this gauge, the ghost part of the action becomes

$$\begin{aligned}
S^{(L)}_{\text{ghost}} &= -\int d^4x \;\; \partial^\mu \tilde{\eta}^a(x)\left\{\partial_\mu \eta^a(x) - g f^{abc} A^c_\mu(x)\eta^b(x)\right\} \\
&= -\int d^4x \;\; \partial^\mu \tilde{\eta}^a(x) D_\mu \eta^a.
\end{aligned} \tag{D.21}$$

In the Abelian case, there is no coupling between the ghost and the gauge fields, and the contribution of the ghosts to the path integral is just an (irrelevant) constant. This is the reason why the Feynman rules for QED could be derived before Faddeev and Popov entered the scene. But it should be mentioned that it is not the non-Abelian character of the theory which necessitates non-trivial ghosts. For instance, in the axial gauge the ghosts can be integrated away both in the Abelian and non-Abelian gauge theory; e.g. [431], (see Sect. 8.1).

In most of the literature, you will find the gauge condition implemented in another way then with the fields γ^a in the action part S_{gf} above. This approach is based on the extension of the Lagrangian by terms with auxiliary fields that may represent the gauge fixing. In Sect. 3.3.4, this was explained for the Lorenz gauge in case of electrodynamics, with the Lagrangian

$$\mathcal{L}_\xi = -\frac{1}{2}\partial^\mu A^\nu(\partial_\mu A_\nu - \partial_\nu A_\mu) + B(\partial^\mu A_\mu) + \frac{\xi}{2}B^2 + A_\mu j^\mu.$$

Extending (D.14) by auxiliary B^a terms, we also introduce a further integration over these fields:

$$Z = \int \mathcal{D}A\,\mathcal{D}B\,\mathcal{D}\bar{\psi}\,\mathcal{D}\psi e^{i\{S_{YM}[A]+\int d^4x\,\left[\frac{\xi}{2}B^2+BG\right]+S_F[A,\psi]\}}.$$

But the B integrations can immediately be executed with (D.4) to yield a term proportional to

$$S'_{\mathrm{gf}} = -\frac{1}{2\xi}\int d^4x\; G^a G^a.$$

Since the most common gauge fixings are linear in the gauge fields (or derivatives thereof), this trick provides a further quadratic term in A^a_μ to the effective action. We furthermore find the link $\gamma^a = -1/(2\xi)G^a$.

The above derivation of the Faddeev-Popov integral can be rendered more geometric by introducing a metric and an affine connection in the space of gauge fields. This has the merit that one directly arrives at the gauge invariant and gauge fixing independent effective action (this term being explained in D.2.2), first advocated by G.A. Vilkovisky; see e.g. [329].

Gravity: Summing Over Geometries

For gravity the first detailed investigation with functional integrals was that of Misner (1957). For the metric field, the path integral formally becomes

$$Z[g_{\mu\nu}] = \int \mathcal{D}g\; e^{iS[g_{\mu\nu}]}.$$

As in any other theory, the measure $\mathcal{D}g = \prod_{\mu,\nu,x} dg_{\mu\nu}(x)$ needs to be defined. This amounts to a summation over all geometries (and maybe also all topologies). But again, as in the Yang-Mills case, one must be aware of the problem of overcounting. In this case, the naive sum over geometries embraces all metrics which are related by diffeomorphism. The Faddeev-Popov method may be employed, but since the diffeomorphism group is not compact, the notion of an invariant group measure is dubious, to put it mildly [38]. In any case, there are various techniques by which functional-integral methods can lead to quantum gravity results, even despite the fact that gravitation is not renormalizable according to dimensional arguments, see e.g. [316].

D.2 Noether and Functional Integrals

The title of this section is obviously nonsense, since E. Noether lived and worked at least one generation before R. Feynman. But what I would like to stress here is how the Noether theorems make their appearance in functional integrals.

D.2.1 Noether Currents and Ward-Takahashi-Slavnov-Taylor Identities

The functional integral is formulated in terms of the classical action. Since Noether symmetries are defined with reference to an action, we expect that these symmetries can be traced more or less directly into the quantum level. Let us assume that we are dealing with a theory with fields Φ_α, defined by its action $S[\Phi]$. Let us first investigate the transformation of the path integral under infinitesimal transformations $\Phi'(x) = \Phi(x) + \delta\Phi$ which we later assume to be symmetry transformations of the classical theory. We begin with the path integral

$$Z[J] = \int \mathcal{D}\Phi \; e^{i(S+J\Phi)} \tag{D.22}$$

where $J\Phi = \int d^4x \; J_\alpha \Phi_\alpha(x)$, and the sum extends over all indices that are necessary to fully specify Φ_α with its Lorentz or internal group transformation behavior. The value of $Z[J]$ does not change when changing the integration variable from Φ to Φ', provided that the measure is invariant ($\mathcal{D}\Phi' = \mathcal{D}\Phi$), which is true in most cases. Thus

$$0 = \delta Z[J] = i \int \mathcal{D}\Phi \; e^{i(S+J\Phi)} \int d^4x \; \big(\frac{\delta S}{\delta \Phi_\alpha} + J_\alpha\big) \delta\Phi_\alpha(x). \tag{D.23}$$

If we take the functional derivative of this expression with respect to $J_\beta(x')$ and then set $J = 0$, we find

$$0 = \int \mathcal{D}\Phi \; e^{iS} \int d^4x \; \left[i \frac{\delta S}{\delta \Phi_\alpha(x)} \Phi_\beta(x') + \Phi_\beta(x') \delta_{\alpha\beta} \delta^4(x,x') \right] \delta\Phi_\alpha(x).$$

In the case that $\delta\Phi$ is a symmetry transformation $\delta_S\Phi$, it has an associated current J^μ_S, and we have from (3.45) $\frac{\delta S}{\delta \Phi_\alpha} \delta_S \Phi_\alpha = -\partial_\mu J^\mu_S$, which can be inserted in the first term of the previous expression

$$0 = \int \mathcal{D}\Phi \; e^{iS} \int d^4x \; \left[-i \partial_\mu J^\mu_S(x) \; \Phi_\beta(x') + \Phi_\alpha(x') \delta^4(x,x') \delta_S \Phi_\alpha(x) \right],$$

or, written in terms of vacuum expectation values

$$i\partial_\mu \langle 0|P J_S^\mu(x)\Phi_\beta(x')|0\rangle = \langle 0|P\Phi_\alpha(x')\delta^4(x,x')\delta_S\Phi_\alpha(x)|0\rangle.$$

In taking not only one but n functional derivatives of (D.23) with respect to $J_{\alpha_k}(x_k)$, one accordingly arrives at

$$\begin{aligned} &i\partial_\mu \langle 0|P J_S^\mu(x)\Phi_{\alpha_1}(x_1)\ldots\Phi_{\alpha_n}(x_n)|0\rangle \\ &\quad= \sum_{k=1}^n \delta^4(x-x_k)\langle 0|P\Phi_{\alpha_1}(x_1)\ldots\delta_S\Phi_{\alpha_k}(x)\ldots\Phi_{\alpha_n}(x_n)|0\rangle. \end{aligned}$$

In QED these identities between correlation functions are called Ward-Takahashi identities, because their consequences were already known from work of J.C. Ward and Y. Takahashi of the early 1950's concerning renormalizability issues. Their non-Abelian generalization goes under the name Slavnov-Taylor identities. Generically, they show that in the quantum field theory the Noether currents, located inside of correlation functions, are conserved up to terms depending on the detailed form of the symmetry transformations $\delta_S\Phi$.

D.2.2 Quantum Action and its Symmetries

Definition and Properties of the Quantum Action

A pivotal role in this book about symmetries is played by the classical action. Only if a theory can be derived from minimizing an action can one make use of all the powerful consequences of variational symmetries. The corresponding object on the quantum level is the *quantum action*, also called the "(quantum) effective action" (and sometimes even "effective" action, although the latter could be confused with the action of an effective field theory).

Again, we start of with the generic path integral (D.22). Define the vacuum expectation value of the $\hat{\Phi}^\alpha$ in the presence of the current J as ϕ_J^α :

$$\phi_J^\alpha(x) = \frac{\langle 0|\hat{\Phi}^\alpha(x)0\rangle_J}{\langle 0|0\rangle_J} = -\frac{i}{Z[J]}\frac{\delta}{\delta J_\alpha(x)}Z[J] = \frac{\delta}{\delta J_\alpha(x)}W[J]. \tag{D.24}$$

Here the W-functional is related to $Z[J]$ by

$$Z[J] = e^{iW[J]}, \qquad W[J] = -i\ln Z[J]. \tag{D.25}$$

In every textbook on quantum field theory it is shown that while $Z[J]$ enfolds all Feynman graphs, the functional $W[J]$ generates all connected graphs. The relation

(D.24) can be inverted to read

$$\phi_J^\alpha(x) = \phi^\alpha \qquad \text{if} \qquad J_\alpha(x) = J_{\phi^\alpha}(x).$$

The quantum action is defined as

$$\Gamma[\phi] = W[J_\phi] - \int d^4x \; J_{\phi^\alpha}(x)\phi^\alpha(x). \tag{D.26}$$

This is a functional of the mean field $\phi^\alpha(x)$ through the implicit dependence of J on the ϕ's. The variational derivative of the quantum action is

$$\frac{\delta\Gamma[\phi]}{\delta\phi^\alpha(x)} = -J_{\phi^\alpha}(x). \tag{D.27}$$

This justifies calling $\Gamma[\phi]$ an "effective action": The values of the $\phi^\alpha(x)$ in the absence of a current J are given by the stationary points of Γ

$$\frac{\delta\Gamma[\phi]}{\delta\phi^\alpha(x)} = 0.$$

This is reminiscent to the derivation of the classical field equations from the action as $\delta S/\delta\phi_c^\alpha = 0$. And indeed, the classical action is the leading term in an ${}^1\!/\!_\hbar$ expansion of the quantum action:

$$\Gamma[\phi] = \frac{1}{\hbar}S[\phi] + \sum_{k=1}^{\infty}(\hbar)^{(k-1)}\Gamma_k[\phi].$$

Symmetries of the Quantum Action

As it turns out, for establishing the renormalizability of a theory, one needs to understand how symmetries of the classical action relate to the symmetries of the quantum action. Hereafter, I will follow the argumentation of [536], Sect. 16.4: Denote again all fields in the theory by Φ and consider infinitesimal transformations $\Phi^\alpha \to \phi^\alpha + \delta\Phi^\alpha[x;\Phi]$. The generating function (D.22) transforms as

$$Z[J] \to \int \mathcal{D}(\Phi + \delta\Phi) \; \exp i\{S[\Phi + \delta\Phi] + J(\Phi + \delta\Phi)\}.$$

If we assume that both the action and the measure are invariant, the previous expression becomes

$$\begin{aligned} Z[J] &\to \int \mathcal{D}\Phi \; e^{\{iS[\Phi] + iJ(\Phi+\delta\Phi)\}} \\ &= \; Z[J] + i\epsilon \int \mathcal{D}\Phi \int d^4y \; \delta\Phi^\alpha[y;\Phi] J_\alpha(y) \; e^{\{iS[\Phi]+iJ\Phi\}}. \end{aligned}$$

From the invariance of $Z[J]$, we obtain the condition

$$\int d^4y \; \langle \delta\Phi^\alpha(y) \rangle_J \, J_\alpha(y) = 0$$

where

$$\langle \delta\Phi^\alpha(y) \rangle_J := \frac{1}{Z[J]} \int \mathcal{D}\Phi \; \int d^4y \; \delta\Phi^\alpha[y;\Phi] \; e^{\{iS[\Phi]+iJ\Phi\}}.$$

Due to (D.27), this condition can be written as

$$\int d^4y \; \langle \delta\Phi^\alpha(y) \rangle_{J_\Phi} \frac{\delta\Gamma[\Phi]}{\delta\Phi^\alpha(y)} = 0, \tag{D.28}$$

revealing that the quantum action is invariant under the infinitesimal transformation $\Phi^\alpha \to \Phi^\alpha + \langle \delta\Phi^\alpha(y) \rangle_{J_\Phi}$. In general, this is not the same invariance transformation as for the classical action. An important exception are linear transformations

$$\delta\Phi^\alpha[x;\phi] = s^\alpha(x) + \int d^4y \; t^\alpha{}_\beta(x,y)\Phi^\beta(y)$$

for which, because of $\langle \Phi^\beta(y) \rangle_J = \Phi^\beta(y)$, we find $\langle \delta\Phi^\alpha(x) \rangle_{J_\Phi} = \delta\Phi^\alpha[x;\phi]$. For these transformations, the quantum action is invariant (provided that both the classical action and the measure are invariant). And indeed the linear transformations are those we encounter in Yang-Mills theories and gravitational theories where

$$\delta\Phi^\alpha = \int d^4y \big[\mathcal{A}^\alpha_r \delta(x-y) + \mathcal{B}^{\alpha\mu}_r \partial_\mu \delta(x-y) + \ldots\big] \cdot \epsilon^r$$

and the coefficients in front of the delta functions are at most linear in the fields.

D.2.3 BRST Symmetries

Symmetries Out of Nowhere?

A major technical advance in dealing with Yang-Mills theories and gravitational theories came with the observation that after gauge fixing, the path integral exhibits a global symmetry. This was discovered by C. Becchi, A. Rouet, R. Stora, and independently by I. V. Tyutin, shortly after the Faddeev-Popov recipe was published. These authors observed that the path integral (D.19) is invariant under the BRST transformations mediated by a constant odd Grassmann number χ

$$\begin{aligned}
\delta_\chi A^a_\mu &= \chi D_\mu \eta^a \\
\delta_\chi \psi &= \chi (T^a \eta^a)\psi \\
\delta_\chi \eta^a &= -\frac{\chi}{2} f^{abc} \eta^b \eta^c \\
\delta_\chi \tilde{\eta}^a &= -\chi \gamma^a \\
\delta_\chi \gamma^a &= 0.
\end{aligned}$$

At first sight these transformations seem a bit patchy, but they do not really come out of nowhere: Acting on the vector fields and the spinors, the transformations δ_χ are recognized as "ordinary" gauge transformation $A^a_\mu \to A^a_\mu + D_\mu \theta^a$ and $\psi \to \psi + (T^a \theta^a)\psi$ with $\theta^a = \chi \eta^a$. The form of the transformations on the other fields is related to the nilpotency of the BRST transformations.

Nilpotency of BRST Transformations

Remarkably, the δ_χ-transformation is nilpotent: $\delta_\chi \delta_{\chi'} X = 0$, where X is any product of the fields from the set $\Psi = \{A^a_\mu, \psi, \eta^a, \tilde{\eta}, \gamma\}$. This can be proved in two steps. Start by investigating the action of a second BRST transformation on the fields themselves. In writing $\delta_\chi \Psi =: \chi s \Psi$ we show that $\delta_\chi s \Psi = 0$. This is obvious for γ and $\tilde{\eta}^a$. For η^a with $s\eta^a = -\frac{\theta}{2} f^{abc} \eta^b \eta^c$:

$$\delta_\chi s \eta^a \propto f^{abc} \chi \{ f^{bde} \eta^d \eta^e \eta^c + f^{cde} \eta^b \eta^d \eta^e \} \equiv 0,$$

because of the Grassmann odd nature of the $\eta' s$ and the Jacobi identities of the structure coefficients. For $s\psi$:

$$\begin{aligned}
\delta_\chi s \psi = \delta_\chi ((T^a \eta^a)\psi) &= (T^a \delta_\chi \eta^a)\psi - (T^a \eta^a)\delta_\chi \psi \\
&= -(T^a \frac{\chi}{2} f^{abc} \eta^b \eta^c)\psi - (T^a \eta^a)\chi (T^b \eta^b)\psi \\
&= \chi \big(T^a T^b \eta^a \eta^b - \frac{1}{2} T^a f^{abc} \eta^b \eta^c \big)\psi.
\end{aligned}$$

Since the first term in the bracket is anti-symmetric under the exchange of η^a and η^b it may be replaced by $(1/2)\ [T^a, T^b]\ \eta^a \eta^b = (1/2) f^{abc}\ T^c\ \eta^a \eta^b$, and then it cancels against the second term. Finally, for $s A^a_\mu$:

$$\begin{aligned}
\delta_\chi s A^a_\mu =& \delta_\chi \{ \partial_\mu \eta^a - f^{abc} A^c_\mu \eta^b \} \\
=& -\frac{\chi}{2} \partial_\mu (f^{abc} \eta^b \eta^c) - \chi f^{abc} (D_\mu \eta^c)\eta^b + \frac{\chi}{2} f^{abc} A^c_\mu f^{bde} \eta^d \eta^e \\
=& -\frac{\chi}{2} f^{abc} \Big\{ \partial_\mu (\eta^b \eta^c) + 2(\partial_\mu \eta^c)\eta^b \Big\} \\
& + \frac{\chi}{2} f^{abc} \Big\{ 2\, f^{cde} A^e_\mu \eta^d \eta^b + f^{bde} A^c_\mu \eta^d \eta^e \Big\} \equiv 0.
\end{aligned}$$

The first term vanishes because of the anti-symmetry of the structure constants in the last two indices, the second vanishes due to the Jacobi identity. This concludes the demonstration that the BRST transformation is nilpotent on any of the fields from the set Ψ. Next, take a product of two fields. We find:

$$\begin{aligned}\delta_\chi(\Psi\Psi') &= (\delta_\chi\Psi)\Psi' + \Psi(\delta_\chi\Psi')\\ &= \chi\{(s\Psi)\Psi' + (-1)^{|\Psi|}\Psi(s\Psi')\} = \chi s(\Psi\Psi').\end{aligned}$$

Making use of $\delta_\chi(s\Psi) = 0 = \delta_\chi(s\Psi')$, we derive

$$\begin{aligned}\delta_\chi s(\Psi\Psi') &= (s\Psi)(\delta_\chi\Psi') + (-1)^{|\Psi|}(\delta_\chi\Psi)(s\Psi')\\ &= (s\Psi)(\chi s\Psi') + (-1)^{|\Psi|}(\chi s\Psi)(s\Psi')\\ &= \chi\{(-1)^{|s\Psi|}(s\Psi)(s\Psi') + (-1)^{|\Psi|}(s\Psi)(s\Psi')\} = 0,\end{aligned}$$

since $|s\Psi| = |\Psi| + 1$. This extends to products of more than two of the Ψ fields and finalizes the proof of the nilpotency of the BRST transformations. The argumentation can be reversed, in the sense that demanding the nilpotency of BRST transformations on the Ψ fields leads to the BRST transformation of the ghost fields (see [481], Sect. 74).

Now let us verify the invariance of (D.19) under the BRST transformations. It was mentioned that the transformation δ_χ, acting on the gauge boson fields and the fermion fields, can be understood as a gauge transformation. Therefore the Yang-Mills action S_{YM} and the fermion part S_F is obviously invariant. What about the (quasi)-invariance of the gauge fixing and the ghost Lagrangians? For the sake of simplicity, let us investigate this for the case of the Lorenz gauge: We have

$$\begin{aligned}\delta_\chi\mathcal{L}_{\text{gf}} &= \delta_\chi(\gamma^a\partial^\mu A^a_\mu) = \gamma^a\partial^\mu\delta_\chi A^a_\mu = \gamma^a\partial^\mu(\chi D_\mu\eta^a)\\ \delta_\chi\mathcal{L}_{\text{ghost}} &= -\delta_\chi(\partial^\mu\tilde{\eta}^a D_\mu\eta^a) = -\delta_\chi(\partial^\mu\tilde{\eta}^a)D_\mu\eta^a - \partial^\mu\tilde{\eta}^a\delta_\chi(D_\mu\eta^a)\\ &= \chi(\partial^\mu\gamma^a)D_\mu\eta^a,\end{aligned}$$

where the term $\delta_\chi(D_\mu\eta^a)$ drops out because of the nilpotency of the BRST transformations. Therefore, at the end of the day,

$$\delta_\chi(\mathcal{L}_{\text{gf}} + \mathcal{L}_{\text{ghost}}) = \partial^\mu(\chi\gamma^a D_\mu\eta^a)$$

which is a total derivative. From this we realize that in the Lorenz gauge the gauge fixing condition and the ghost contribution are interwoven in such a way that $(\mathcal{L}_{\text{gf}} + \mathcal{L}_{\text{ghost}})$ is quasi-invariant. This holds true for any gauge choice; see again [536], Sect. 15.7).

Last but not least, the BRST-invariance proof is complete only if the measure in (D.19) is invariant. This may be verified by computing the Jacobian of the BRST transformations on the set of fields Ψ near the identity:

$$\mathcal{J}_\chi = \mathrm{sdet}\big[\delta_\alpha^\beta + \frac{\partial(\delta_\chi \Psi_\alpha)}{\partial \Psi_\beta}\big] = 1 + \mathrm{str}\big[\frac{\partial(\delta_\chi \Psi_\alpha)}{\partial \Psi_\beta}\big];$$

(for the definition and properties of the supertrace and the superdeterminant see Appendix B.2.2.) The transformations on the $\tilde{\eta}^a$ and the γ^a fields do not contribute to the trace, which means that the measures $\mathcal{D}\tilde{\eta}^a$ and $\mathcal{D}\gamma^a$ are separately BRST invariant. The remaining terms read

$$\mathcal{J}_\chi = 1 + \mathrm{str}\big[\frac{\partial(\delta_\chi A_\mu^a)}{\partial A_\nu^d} + \frac{\partial(\delta_\chi \eta^a)}{\partial \eta^d}\big] = 1 + \chi \;\; \mathrm{tr}\big[f^{adc}\delta_{\mu\nu}\eta^c - f^{adc}\eta^c\big] = 1.$$

The BRST transformations are global transformations. They give rise to Noether currents

$$J_\chi^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Psi^\alpha)}\delta_\chi \Psi^\alpha =: \chi j_B^\mu,$$

and thus to the BRST charge $Q = \int j_B^0 d^3x$ which because of $\delta\chi\delta\chi = 0$ satisfies the nilpotency condition

$$Q^2 = 0.$$

The BRST charge is a sublime device to identify the physical states among the Fock-space states $|\Omega\rangle$ generated in the usual way. As will be shown below, the condition $Q|\Omega\rangle = 0$ properly selects the physical states of the system. Therefore, BRST quantization has established itself as a mathematically rigorous approach to quantizing field theories with a local symmetry.

BRST Symmetries as the Quantum Equivalent to Classical Symmetries

Although the BRST symmetry was discovered in the Faddeev-Popov path integral formulation of non-Abelian gauge theories, it developed an independent existence as it became possible to introduce the symmetries and their associated fields already on the level of the classical theory.

Let me outline in which sense "two quantum symmetries are associated with a classical symmetry" [5]; also [36]. In the condensed DeWitt notation, the symmetry transformations on the fields $\Phi = \{A, \psi\}$ were written in Sect. 3.3.5 as

$$\delta\Phi^\alpha = \mathcal{R}_r^\alpha \cdot \epsilon^r. \tag{D.29}$$

To simplify matters, assume that there are only bosonic symmetries, such that the symmetry parameters ϵ are even Grassmann variables. Further, assume that the algebra closes on-shell; that is, we are dealing with a closed algebra (not necessarily a Lie algebra). Then the conditions (3.90) read

$$\mathcal{R}^{\alpha}_{r,\beta}\mathcal{R}^{\beta}_{s} - \mathcal{R}^{\alpha}_{s,\beta}\mathcal{R}^{\beta}_{r} = -\mathcal{R}^{\alpha}_{t} f^{t}{}_{rs}.$$

Now introduce for each of the ϵ^r a partner c^r with Grassmann parity opposite to ϵ^r. Thus in this assumed case, the c^r are Grassmann odd. Then the previous closure relations and the Jacobi identity become

$$\mathcal{R}^{\alpha}_{s,\beta}\mathcal{R}^{\beta}_{r}c^r c^s = \frac{1}{2}\mathcal{R}^{\alpha}_{t} f^{t}{}_{rs}c^r c^s \qquad (f^{r}{}_{st,\alpha}\mathcal{R}^{\alpha}_{p} - f^{r}{}_{qt} f^{q}{}_{ps})c^p c^s c^t = 0. \tag{D.30}$$

These relations even hold if in the set of infinitesimal parameters there were Grassmann odd symmetry parameters ϵ^r (together with their Grassmann even partner c^r)–this is one of the strengths of introducing the c^r. Furthermore, in the case of a (super)-Lie algebra with $f^{r}{}_{st,\alpha} = 0$ the second relation implies the (graded) Jacobi-identity for the structure constants.

But the trick of using the c^r as a bookkeeping device is only the beginning of a fascinating story of defining the quantum symmetries. For this purpose, consider pairs of independent ghost fields c^r and $\tilde{c}^r$. (Very often in the literature the $\tilde{c}^r$ are called anti-ghosts. This could be confusing since they are in no sense the anti-particles of the ghosts c^r.) Define the BRST and anti-BRST transformations (denoted by s and $\bar{s}$)

$$s\Phi^{\alpha} = \mathcal{R}^{\alpha}_{r}c^r \qquad \bar{s}\Phi^{\alpha} = \mathcal{R}^{\alpha}_{r}\tilde{c}^r \tag{D.31a}$$

$$sc^r = -\frac{1}{2}f^{r}{}_{st}c^s c^t \qquad \bar{s}\tilde{c}^r = -\frac{1}{2}f^{r}{}_{st}\tilde{c}^s\tilde{c}^t \tag{D.31b}$$

$$s\tilde{c}^r + \bar{s}c^r = -\frac{1}{2}f^{r}{}_{st}c^s\tilde{c}^t. \tag{D.31c}$$

If one introduces an additional field b^r (corresponding to the γ-fields in the gauge fixing action part (D.20a)) having the same Grassmann parity as the ϵ^r, the previous set of transformation is completed to

$$\begin{aligned} s\tilde{c}^r &= b^r & \bar{s}c^r &= -b^r - f^{r}{}_{st}\tilde{c}^s c^t \\ sb^r &= 0 & \bar{s}b^r &= -f^{r}{}_{st}\tilde{c}^s b^t - \frac{1}{2}\mathcal{R}^{\alpha}_{q} f^{r}{}_{st,\alpha}c^q\tilde{c}^s c^t. \end{aligned}$$

One finds that the transformations s and $\bar{s}$ fulfill the nilpotency relations

$$s^2 = s\bar{s} + \bar{s}s = \bar{s}^2 = 0$$

as a consequence of the closure property and the Jacobi condition in the form (D.30). The nilpotency of the generators is indeed equivalent to the closure relation when they act on the fields Φ, while it is equivalent to the Jacobi identity when they act on ghosts and anti-ghosts. Thus, one can define the symmetry transformations either by equations (D.29) completed by the closure and Jacobi relations, or by the set of BRST and anti-BRST transformations together with the nilpotency relations. Observe the course of events: One started from a theory which is invariant under local symmetry transformations and thus has a redundant set of fields. Instead of resolving this redundancy one introduces more fields. In this way, the local symmetry disappears, there is no redundancy any longer; instead there is a global symmetry with a Noether charge that can be used directly to define the Hilbert space of the quantized theory. This equivalence between the local gauge invariance in the original theory and the global nilpotent symmetries remarkably holds in case of closed algebras and therefore not only for Yang-Mills type gauge theories but also for general relativity. Another remarkable result concerns the path integral measure

$$\mathcal{D}\Phi := \mathcal{D}\Phi^\alpha \mathcal{D}c^r \mathcal{D}\tilde{c}^r \mathcal{D}b^r .$$

It turns out that $\mathcal{D}\Phi$ is invariant under s- and $\bar{s}$-transformations iff ${f^r}_{rt} = 0$, and this is true if the symmetry group is compact. This reflects factoring out of the Haar measure in the Faddeev-Popov procedure.

A BRST charge can be established for any gauge symmetry with a Lie algebra $[\Gamma_r, \Gamma_s] = {f^t}_{rs}\Gamma_t$ by a simple recipe: Introduce two sets of anticommuting operators c^r and $\tilde{c}_r$ which obey the commutation relation $\{c^r, \tilde{c}_s\} = \delta^r_s$. (If interpreted as a Poisson bracket, the c and $\bar{c}$ are canonically conjugate variables.) Then

$$Q = c^r(\Gamma_r - \frac{1}{2}{f^t}_{rs}c^s\tilde{c}_t)$$

is nilpotent because of the Jacobi identity of the structure constants and the oddness of the ghosts and anti-ghosts. As charges, the Γ_r generate symmetry transformations: $\delta\Psi^\alpha = [\Gamma_r, \Psi^\alpha]\,\epsilon^r = \mathcal{R}^\alpha_r\epsilon^r$. The transformations mediated by Q

$$\begin{aligned}
\delta_Q\Psi^\alpha &:= \{Q, \Psi^\alpha\} = \left[\Gamma_r, \Psi^\alpha\right]c^r = s\Psi^\alpha \\
\delta_Q c^r &:= \{Q, c^r\} = -\frac{1}{2}{f^r}_{st}c^s c^t = sc^r \\
\delta_Q\tilde{c}_r &:= \{Q, \tilde{c}_r\} = \Gamma_r - {f^t}_{rs}c^s\tilde{c}_t
\end{aligned}$$

reproduce previous results as in (D.31). Define the ghost number operator

$$N_g = \sum_{r=1}^{\dim G} c^r\tilde{c}_r .$$

Its commutator with the BRST charge is $\left[N_g, Q\right] = Q$. All those states $|\Phi\rangle$ of a Hilbert space are called BRST invariant for which $Q|\Phi\rangle = 0$. Because of the nilpotency of Q, one traces two kinds of BRST-invariant states: (1) any state of the form $|\Phi\rangle = Q|\Omega\rangle$. The states $|\Phi\rangle$ and $|\Omega\rangle$ differ by one unit in ghost number. The state $|\Phi\rangle$ has zero norm because $\langle\Omega|Q^\dagger Q|\Omega\rangle = \langle\Omega|Q^2|\Omega\rangle = 0$. (2) states $|\Phi\rangle$ with

$$Q|\Phi\rangle = 0 \qquad |\Phi\rangle \neq Q|\Omega\rangle \qquad \text{for all } \Omega \text{ of the Hilbert space;}$$

these are the physically relevant ones. It also makes sense to term all those states $|\Phi\rangle$ and $|\Phi'\rangle$ equivalent for which $|\Phi'\rangle = |\Phi\rangle + Q|\Omega\rangle$. The equivalence classes are called BRST cohomology classes[4]. Now consider specifically all those states $|\Phi\rangle^0$ which do have zero ghost charge. These states are annihilated by the $\tilde{c}_r$ and we have $Q|\Phi\rangle^0 = c^r\Gamma_r|\Phi\rangle^0 = 0$. Therefore these BRST-invariant states are also invariant under the action of the Lie algebra.

Antibracket, Antifields and Gauge Theory Quantization

The title of this (admittedly small) subsection is identical with that of a review article [232] on what is also known as the Batalin-Vilkovisky quantization–treated likewise in the textbook [536]. It can be thought of as a further generalization of the previously sketched BRST formalism for gauge theories with closed algebras to those with an open algebra or with so-called reducible gauge symmetries. Thus, it seems to be the most general quantization procedure for field theories invariant under those symmetry transformations (D.29) that close as a group. It has its merits specifically in gravity and supergravity.

D.2.4 Fujikawa: Fermionic Path Integrals and Anomalies

In the previous sections, we came several times to the question of whether the path-integral measure with respect to the fermions is invariant under certain transformations. The answer depends on the type of transformations. This sounds obvious, but was not really deeply investigated before the work of K. Fujikawa in 1979 [204]. In his textbook, A. Zee tells us that some time ago, you could find people that were not fond of the path-integral approach, and that they felt vindicated in that the path integrals seemingly could not exhibit the breaking of chiral invariance; [578], p. 278. Fujikawa however realized that a chiral transformation gives rise to a change in the measure of the path integral over the fermion fields, and that the Jacobian yields exactly the term historically found from triangle diagrams. In deriving this, Fujikawa

---

[4] Indeed there is a large similarity of the BRST operator Q and the de Rham operator d. Therefore, the previous considerations mimic the distinction of closed and of exact forms; see Appendix E.2.4.

worked with the Euclidean path integral[5], because in four Euclidean dimensions the Dirac operator is Hermitean, and this allows for a rigorous derivation of the anomaly. For an elaborate account of anomalies in quantum field theory see [39].

Consider a unitary transformation on fermionic fields

$$\psi(x) \to U(x)\psi(x) \qquad \bar{\psi}(x) \to \bar{\psi}(x)\ (\gamma^0 U^\dagger(x)\gamma^0).$$

Since for fermionic variables, the measure transforms with the inverse of the transformation matrix (see Appendix B.2.3.), we get

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} \to (\text{Det}\,\mathcal{U}\ \text{Det}\,\bar{\mathcal{U}})^{-1}\mathcal{D}\psi\mathcal{D}\bar{\psi}$$

where

$$\mathcal{U}_{xn,ym} := U(x)_{nm}\delta^4(x-y) \qquad \overline{\mathcal{U}}_{xn,ym} := [\gamma^0 U^\dagger(x)\gamma^0]_{nm}\delta^4(x-y).$$

Here the indices (n, m) run over Dirac spin indices and possible species ("flavor") indices. In case that $U(x)$ is a unitary non-chiral transformation $\mathcal{U}(x) = \exp\{i\lambda(x)T\}$ with T a Hermitean matrix not involving γ^5, we have $\overline{\mathcal{U}}\mathcal{U} = 1$. This entails $\text{Det}\,\mathcal{U}\ \text{Det}\,\overline{\mathcal{U}} = 1$, and therefore for non-chiral transformations, the measure is invariant. This is true in particular for gauge transformations where T is any one of the gauge group generators T^a. For chiral transformations, we write

$$\mathcal{U}(x) = \exp\{i\gamma^5\lambda(x)T\}.$$

Since γ^5 anti-commutes with γ^0, the matrix $\mathcal{U}$ is (pseudo)-Hermitean:

$$\overline{\mathcal{U}} = \mathcal{U}$$

and the measure picks up a factor $(\text{Det}\,\mathcal{U})^{-2}$: $\mathcal{D}\psi\mathcal{D}\bar{\psi} \to (\text{Det}\,\mathcal{U})^{-2}\mathcal{D}\psi\mathcal{D}\bar{\psi}$. We may calculate

$$\text{Det}\,\mathcal{U} = \exp\text{Tr}\ \{\ln\mathcal{U}\} = \exp\text{Tr}\ \{\ln(1 + (\mathcal{U} - 1))\} \sim \exp\text{Tr}(\mathcal{U} - 1),$$

which holds for infinitesimal chiral transformations

$$[\mathcal{U} - 1]_{xn,ym} \propto i\lambda(x)[\gamma^5\, T]_{nm}\delta^4(x-y).$$

Therefore,

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} \to e^{i\int d^4x\ \lambda(x)\mathcal{A}(x)}\mathcal{D}\psi\mathcal{D}\bar{\psi}$$

---

[5] I did not mention the technical trick of a "Wick rotation" $t \leftrightarrow i\tau$ which serves to provide a path integral with a better behavior, but is not completely well defined because of the question of whether an analytical continuation always exists.

with the anomaly function

$$\mathcal{A}(x) = -2\mathrm{Tr}_{(n,m)}\,[\gamma^5\, T]\delta^4(x-y)$$

and with the trace to be taken both over Dirac and flavor indices.

The variation of the generating function is

$$\mathcal{D}\psi\mathcal{D}\bar{\psi}\; e^{i\int d^4x\;\mathcal{L}(x)} \\ \to \mathcal{D}\psi\mathcal{D}\bar{\psi} \times e^{i\int d^4x\;\lambda(x)\mathcal{A}(x)} \times \exp\Big\{i\int d^4x\;[\mathcal{L}(x) - \lambda(x)\partial_\mu J^\mu_A]\Big\},$$

where $J^\mu_A = \bar{\psi}\gamma^\mu\gamma^5\psi$ is the classical axial current (6.35). In order that the generating function is invariant one must choose $\partial_\mu J^\mu_A = \mathcal{A}(x)$.

Now of course the tricky point is to calculate the anomaly function $\mathcal{A}(x)$. As a matter of fact, it is not well defined by the previous expression, since the trace vanishes and the delta function is infinite. Therefore it needs to be regularized. Since this is not intended to become a book on quantum field theory, I will not replicate the regularization procedure here (see the original article by Fujikawa or Sect. 22.2 in Weinberg's book) but only state the result

$$\mathcal{A}(x) = -\frac{1}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}_a F^{\rho\sigma}_b\;\mathrm{Tr}(T^a T^b T)$$

where 'Tr' is the trace over the flavor indices.

Appendix E
*Differential Geometry

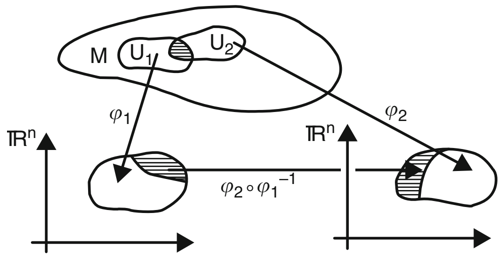
This appendix introduces more or less informally and rather sketchy the differential geometric notions that can help to understand gauge and gravity theories on an abstract level. This demonstrates anew the "unreasonable effectiveness of mathematics in the natural sciences" (E. Wigner). For further details, and especially for proofs, I refer to [324] (by many regarded as the most comprehensive description of the field on an advanced level), [146] (describing in a compact way the full set of tools needed in mathematical physics), [192] (maybe more mathematics- than physics-oriented), [292] (high-level with detailed proofs, but with many examples, also from physics), [470] (a strongly physics-oriented overview), [511] (an excellent, but rather dense presentation of fibre bundles in fundamental physics). A rich arsenal of material, especially if it comes to the use of differential forms in contemporary theoretical physics is found in [265]. A good introductory chapter on differential geometry, topology and fibre bundles can be found in [39], and the relations between notions in physics and those of topology and geometry are exemplified in [379]. In most of this literature you find topics that are not treated in this appendix, especially the astounding relations which exist between topological invariants and the local geometry of manifolds.

But before plunging into manifolds and fibre bundles, let me motivate shortly why in physics these mathematical concepts are required. The arena for talking about physics is (still) space and time. It seems natural to represent spacetime by a set of 'spacetime points'.[1] But in physics we need more structure, since we want to talk about space-time regions, and about points that are in a certain neighborhood, or even ordered. This necessitates to give spacetime a topology. Since the days of Newton, physics became successful in describing dynamical evolution in terms of differential equations in spacetime. Therefore an additional structure (differentiability) needs to be imposed on the mathematical spaces involved. In order to identify events in spacetime we use coordinates, but of course these are only a technical means. The dynamics of a physical system is independent of coordinates, and the description in

[1] Quotation marks are used here, because general relativity reveals that this reading is to naive, and that importance has to be given to relations between events; see e.g. [451].

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

Fig. E.1 Manifold



terms of coordinates may even hide certain structural features of the system or pretend features which are pure coordinate effects. Adding coordinate independence to the previous structures immediately leads to the mathematical notion of differentiable manifolds. In present-day fundamental physics we meet further structures like gauge groups, connections and curvatures. And as will be pointed out in this appendix, the appropriate arena to deal with these structures are fibre bundles.

E.1 Differentiable Manifolds

E.1.1 From Topological Spaces to Differentable Manifolds

It is assumed that the reader is acquainted with the notion of a topological space as a structure on which one can define a neighborhood and continuous functions. A *homeomorphism* between two topological spaces is a 1-1 map $\varphi : X \to Y$ for which both φ and its inverse φ^{-1} are continuous. If φ and φ^{-1} are continuously differentiable then φ is called a *diffeomorphism*.

A D-dimensional manifold M^D is a topological space[2] that locally has the properties of a D-dimensional Euclidean space $\mathbb{R}^D$: A neighborhood of a point in M^D can continuously be mapped in a one-to-one way to the neighborhood of a point in $\mathbb{R}^D$. To be more precise, introduce a *chart* $(U_\alpha, \varphi_\alpha)$ as a homeomorphism φ_α from an open set $U_\alpha \subset M^D$ into an open set $R_\alpha \subset \mathbb{R}^D$. Two charts are compatible if the overlap maps are diffeomorphisms $(\varphi_1 \cdot \varphi_2^{-1} \in C^\infty,\ \varphi_2 \cdot \varphi_1^{-1} \in C^\infty)$ unless $U_1 \cap U_2 = \emptyset$; see Fig. E.1. A set of compatible charts covering M^D is called an atlas. In every chart the manifold can be equipped with a coordinate system: for $x \in M^D$ the coordinates are $x^\mu = \varphi(x) \in \mathbb{R}^D$.
The naming makes it clear what one is aiming at. For instance the surface of a sphere, although not being homeomorphic to a plane, locally has enough smoothness to be mapped into an atlas. One chart is not sufficient since there will always be a point on the sphere that cannot be projected to the plane.

[2] More rigourosly, it is assumed that it is a connected, Hausdorff topological space. But since the following presentation is without proofs, these technicalities are mentioned here only for purists.

In this appendix I will only treat finite-dimensional manifolds. One possibility of extending the notion of manifolds to infinite dimensions is to consider Banach manifolds modeled on Banach spaces. It is also assumed that we are throughout this appendix dealing with C^∞ manifolds (as in the previous definition). In certain contexts it might suffice that the charts are C^k-related. (In the sequel all maps in C^∞ manifolds will be called smooth.) Also complex manifolds are investigated in mathematics and applied to modern theoretical physics (catchword: Kähler manifolds). In these the transition functions are required to be analytic.

E.1.2 Tensor Bundles

On a manifold one can erect tensor bundles as "superstructures" by starting with defining the tangent and cotangent spaces of a manifold.

Tangent Bundle and Vector Fields

We are interested in the notion of vectors on a manifold M (henceforth I will mostly drop the index for the dimension of the manifold and for the Euclidean space). The idea is to introduce these as tangent vectors of curves 'through' $x \in M$: A curve through[3] a point x is a smooth mapping of an interval $I = [0, 1] \subset \mathbb{R}$ to the manifold:

$$C : \mathbb{I} \to M \qquad t \mapsto C(t) \qquad \text{with} \qquad C(0) = x.$$

The coordinates of this curve are $x^\mu(C(t))$, and the tangent vector to the curve is

$$\frac{d}{dt} x^\mu(C(t)).$$

Since one can have more then one curve with $C(0) = x$, the proper definition is: *A tangent vector* at $x \in M$ is an equivalence class of curves in M, where the equivalence relation between two curves is that they are tangent at the point x.
Another–equivalent–definition is to understand a tangent vector as a directional derivative: Consider functions $f \in \mathcal{F}M$, that is $f : M \to \mathbb{R}$. The change of f along a curve is given by

$$\frac{d}{dt} f(C(t)), \qquad \text{locally} \qquad \frac{\partial}{\partial x^\mu} f \, \frac{dx^\mu(C(t))}{dt}.$$

[3] Observe that a curve is a map, thus what is meant, is the image of the curve through x.

In defining

$$X = (X^\mu \partial_\mu) \qquad \text{with} \qquad X^\mu = \frac{dx^\mu(C(t))}{dt}$$

we can write $\frac{d}{dt} f(C(t)) = Xf$. For every point along the curve we take this expression to define the differential operator X_x as the tangent vector to the manifold in $x \in M$. All tangent vectors at a point in the manifold can be shown to build a vector space $\mathfrak{X}_x M$ isomorphic to $\mathbb{R}^D$. The natural basis in $\mathfrak{X}_x M$ is the coordinate or *holonomic* basis $\{\partial_\mu\}$. But of course any other (*anholonomic*) basis $\{e_I\}$ can be chosen where $\partial_\mu = e^I_\mu \, e_I$ with an invertible $D \times D$ matrix e^I_μ–in physical contexts called a D-bein. Every element $X_x \in T_x M$ can be written as

$$X_x = (X^\mu \partial_\mu)_x = (X^I e_I)_x \qquad \text{with} \qquad X^\mu = X^I \, e^\mu_I .$$

The X^μ correspond to what in the component language is called a covariant vector. This can directly be verified by calculating, how the vector changes from one coordinate chart to another overlapping one.

The space $\mathfrak{X}_x M$ can be given a Lie algebra structure:

$$\begin{aligned} &[,] : \mathfrak{X}_x M \otimes \mathfrak{X}_x M \to \mathfrak{X}_x M \qquad \text{with} \\ &\qquad [X_x, Y_x] = Z_x = (Z^\nu \partial_\nu)_x = \Big[(X^\mu Y^\nu_{,\mu} - Y^\mu X^\nu_{,\mu}) \partial_\nu \Big]_x . \end{aligned} \tag{E.1}$$

This algebra is infinite-dimensional and closely related to the group of diffeomorphisms of the manifold. The algebra (E.1), evaluated on the basis vectors, results in $[\partial_\mu, \partial_\nu] = 0$ for the holonomic basis, but in any anholonomic basis we have $[e_I, e_J] = C^K{}_{IJ} \, e_K$ instead.

Finally define the *tangent bundle* $TM = \cup_x T_x M$ in which at each point in the (base) manifold the tangent space at this point is attached as a fibre. (More on fibre bundles in the last part of this appendix.) Then the space of all vector fields $\mathfrak{X} M$ is interpreted as a section $s : M \to TM$ in this bundle, where s is a C^∞ map.

Cotangent Bundle and One-Forms

The vector space $T^*_x M$, called the cotangent space to the manifold at one of its points x, is defined as the vector space dual to $T_x M$. That is, if as a basis in $T_x M$ one chooses the holonomic $(\partial_\mu)_x$ and calls $(dx^\mu)_x$ the dual basis in $T^*_x M$ we have by definition $dx^\mu(\partial_\nu) = \langle dx^\mu, \partial_\nu \rangle = \delta^\mu_\nu$. For generic elements $X \in T_x M$ and $\omega \in T^*_x M$ holds

$$\alpha(X) = \langle \alpha, X \rangle = \alpha_\mu X^\nu \langle dx^\mu, \partial_\nu \rangle = \alpha_\mu X^\mu .$$

This result is independent of the basis chosen. For anholonomic bases ϑ^I in T_x^*M holds $\langle \vartheta^I, e_J\rangle = \delta^I_J$. In this case the anholonomic and holonomic bases are related by $\vartheta^I = e^I_\mu dx^\mu$.

In analogy to previous definitions one obviously can build the vector space $\mathfrak{X}^*M$. The bundle formed from a manifold and its pointwise attached cotangent spaces is called the *cotangent bundle*. Sections of the cotangent bundle are called *one-forms*. The components of a one-form are identified with those of a contravariant vector[4]. Any cotangent bundle is a symplectic manifold; see Appendix E.1.4.

Tensor Fields

From the tangent and cotangent space of a manifold M at a point x (for short written as $T = T_x M$ and $T^* = T_x^* M$) one can *via* multilinear mappings

$$\underbrace{T^* \times T^* \times \ldots T^*}_{\text{r}} \times \underbrace{T \times T \times \ldots T}_{\text{s}} \to \mathbb{R}$$

define r-times covariant and s-times contravariant tensor spaces of the manifold M at a point x ; the pair (r, s) defining the order of the tensor space. In taking together tensor spaces of a given order from all points of the manifold one obtains the tensor bundle $T^r_s M$. Sections in this bundle are *tensor fields*.

As common in most of the mathematical literature, a tensor field is understood to take vector fields and one-forms as arguments to deliver (locally) its components: For $t^r_s \in T^r_s M$

$$t^r_s = t^{I_1 \cdot\cdot I_r}_{J_1 \cdot\cdot J_s}\, \vartheta^{J_1} \otimes \cdot \cdot \otimes \vartheta^{J_s} \otimes e_{I_1} \otimes \cdot \cdot \otimes e_{I_r}, \qquad t^r_s(X_1, \cdot\cdot, X_r, \omega_1, \cdot\cdot, \omega_s) = t^{I_1 \cdot\cdot I_r}_{J_1 \cdot\cdot J_s}.$$

In the majority of considerations in this book, the notions of a tangent and a cotangent bundle suffice. Furthermore, a prominent role is played by bundles $\Omega_p M \subset T^0_p M$. The bundles $\Omega_p M$ contain all totally antisymmetric p-maps $T^* \times T^* \times \ldots T^* \to \mathbb{R}$. Sections in $\Omega_p M$ are called (differential) p-forms. More about differential forms below in E.2.

Pull-Back and Push-Forward

Frequently we are asking for the transformation of vector fields or forms (or tensor fields in general) of a manifold M with respect to diffeomorphisms $\varphi : M \to N$. The mappings

$$\varphi_* : T^r_s M \to T^r_s N \qquad \text{and} \qquad \varphi^* : T^r_s N \to T^r_s M$$

---

4 … that is, a contravariant vector is not a vector in the language and domain of smooth manifolds.

are called *push-forward* and *pull-back*, respectively. Instead of defining them generically here, let me state what they are in case of vector fields and one-forms:
For $\varphi_* : T_x M \to T_{\varphi(x)} N$ define for $X \in T_x M$ the push-forward $\varphi_* X$ by

$$\varphi_* X\,[f] = X\,[f \circ \varphi] \qquad \text{in coordinates} \qquad \varphi_* X^\nu(y) = X^\mu \frac{\partial y^\nu}{\partial x^\nu}(x).$$

For $\varphi^* : T^*_{\varphi(x)} N \to T^*_x M$ define the pull-back in such a way that for an arbitrary vector field $X = X^\nu \partial_\nu \in T_x M$ and a one-form $\alpha = \alpha_\mu dx^\mu \in T^*_x M$ holds $\langle \varphi^* \alpha, X \rangle = \langle \alpha, \varphi_* X \rangle$. By comparing $\langle \varphi^* \alpha, X \rangle = \varphi^* \alpha_\mu(x) X^\mu(x)$ with $\langle \alpha, \varphi_* X \rangle = \alpha_\nu(y)$ the pull-back for the components is

$$\varphi^* \alpha_\mu(x) = \alpha_\nu(y(x)) \frac{\partial y^\nu}{\partial x^\mu}.$$

This can readily be generalized to p-forms. The pull-back on forms has the properties

$$d(\varphi^* \alpha) = \varphi^* d\alpha\ , \quad \varphi^*(\alpha_p \wedge \beta_q) = \varphi^* \alpha_p \wedge \varphi^* \beta_q, \quad (\varphi \cdot \varphi')^* = \varphi^{'*} \cdot \varphi^*. \tag{E.2}$$

E.1.3 Flows and the Lie Derivative

Vector fields on a manifold generate flows: Let C be a curve on the manifold parametrized as $C(t) \in M$ and let the vector $X(C(t))$ be tangent to the curve, that is $\dot{C} = X(C(t))$. Then C is called the integral curve of the vector field X. This obviously characterizes dynamical systems in which an integral curve is called a trajectory. Given a vector field it is always possible to find locally a unique integral curve.

Flows are also called one-parameter groups of diffeomorphisms: Let Φ_t be (for every t) a diffeomorphism $\Phi_t : M \to M$ with $\Phi_0 = \text{id},\ \Phi_t \cdot \Phi_{t'} = \Phi_{t+t'}$. The vector field associated to this diffeomorphism is

$$\frac{d}{dt} \Phi_t(x) \Big|_{t=0} = X^\Phi_x. \tag{E.3}$$

A vector field X on a manifold induces a differential operator $\pounds_X$–called the *Lie derivative*–and having the meaning of the derivative in the direction of the vector field. It is defined to operate on all tensor fields on the manifold M and transforms a geometric object of a certain kind (in terms of its covariant and contravariant indices) into an object of the same kind.

In the simplest case it acts on functions $f \in \mathcal{F} M$:

$$(\pounds_X f)_x = \frac{d}{dt} \Big(\big(f(\Phi_t(x)) \big) \Big|_{t=0}$$

or in coordinates/charts for which $X = X^\mu \partial_\mu$

$$\pounds_X f = X^\mu \frac{\partial f}{\partial x^\mu}.$$

The differential operator $\pounds_X$ has all properties of a derivative, that is (i) $\pounds_X(f_1 + f_2) = \pounds_X f_1 + \pounds_X f_2$ (linearity) (ii) $\pounds_X(f_1 \cdot f_2) = f_1 \pounds_X \cdot f_2 + f_2 \cdot \pounds_X f_1$ (Leibniz rule).

The Lie derivative of a vector field Y with respect to a vector field X is defined to be the vector field Z by

$$Z = \pounds_X Y = \pounds_X \pounds_Y - \pounds_Y \pounds_X := [X, Y]\,.$$

In a chart with $X = X^\mu \partial_\mu,\ Y = Y^\nu \partial_\nu$

$$[X, Y]\ f = \left[X^\mu \frac{\partial Y^\nu}{\partial x^\mu} - Y^\mu \frac{\partial X^\nu}{\partial x^\mu} \right] \frac{\partial}{\partial x^\nu}\ f.$$

This $[\cdot, \cdot]$ composition is identical with the one in (E.1).

The Lie derivative on a p-form is defined in the next section.

The invariance of a tensor field $T \in T^r_s M$ under a flow mediated by a vector field X is synonymous with $\pounds_X T = 0$.

E.1.4 Symplectic Manifolds

Definition: A symplectic manifold (M, ω) is a smooth manifold M with a symplectic form ω, that is a non-degenerate closed two-form.

Although differential forms are presented in more details in the next section, the terms 'two-form', 'closed' and 'non-degenerate' can already be explained here: A two-form is related to an antisymmetric tensor field in $T^*M \otimes T^*M$. In a coordinate basis it has the generic form

$$\omega = \frac{1}{2}\omega_{\mu\nu}(dx^\mu dx^\nu - dx^\nu dx^\mu) =: \frac{1}{2}\omega_{\mu\nu}\, dx^\mu \wedge dx^\nu\,.$$

The two form is *closed* if its exterior derivative (explained below) vanishes. In components the exterior derivative of ω is: $d\omega = \frac{1}{3!}\omega_{\mu\nu,\lambda}\ dx^\lambda \wedge dx^\mu \wedge dx^\nu$. Therefore closedness is synonomous with $\omega_{[\mu\nu,\lambda]} = 0$ (in all coordinate systems). In a coordinate-free language, a two-form takes two vectors as input and outputs a real number: $\omega(X, Y) \mapsto \mathbb{R}$. The two-form is defined as *non-degenerate* if $\omega(X, Y) = 0$ for $\forall\ Y$ implies $X = 0$. In the local coordinates used before, non-degeneracy is synonymous to $\det(\omega_{\mu\nu}) \neq 0$.

Prominent examples of symplectic manifolds are the cotangent bundles T^*M of a manifold. In order to keep contact to applications in physics, we select especially T^*Q, the phase space to a configuration space Q in classical mechanics. In coordinates (q^k, p_k) on T^*Q, the canonical symplectic form is $\omega = dp_k \wedge dq^k$.

Local symmetries in fundamental physics typically imply that the dynamics happens on a subset of the phase space. This subset is in general not a symplectic manifold, but only a pre-symplectic one, characterized by a two-form which is closed but degenerate.

E.2 Cartan Calculus

E.2.1 Differential Forms

Differential p-forms ω on a D-dimensional manifold are elements of $\Omega_p M \subset T_p^* M$: If ϑ^I is chosen as the basis in the cotangent space T^*M, a basis in $\Omega_p M$ is provided by

$$\vartheta^{I_1 I_2 \ldots I_p} := \vartheta^{[I_1} \otimes \vartheta^{I_2} \otimes \ldots \otimes \vartheta^{I_p]}$$

where the [] denotes complete anti-symmetrization including a factorial, e.g.

$$\vartheta^{IJ} = \frac{1}{2!}(\vartheta^I \otimes \vartheta^J - \vartheta^J \otimes \vartheta^I).$$

A p-form thus has the generic appearance

$$\alpha_p = \frac{1}{p!}\hat{\alpha}_{I_1 I_2 \ldots I_p}\vartheta^{I_1 I_2 \ldots I_p}.$$

with $\hat{\alpha}_{I_1 I_2 \ldots I_p} \in \mathcal{F}M$. (If it is clear from the context neither the subscript indicating the form degree nor the hat symbol on the components will be mentioned.) In the space

$$\Omega M := \bigoplus_{p=0}^{D} \Omega_p M$$

one defines several operations

- Exterior product $\wedge : \ \Omega_p M \times \Omega_q M \to \ \Omega_{p+q} M$, defined on the basis as

$$\vartheta^{I_1 I_2 \ldots I_p} \wedge \vartheta^{J_1 J_2 \ldots J_q} := \vartheta^{I_1 I_2 \ldots I_p J_1 J_2 \ldots J_q}$$

from which

$$\alpha_p \wedge \beta_q = (-1)^{pq} \alpha_q \wedge \beta_p.$$

- Exterior derivative $d : \Omega_p M \rightarrow \Omega_{p+1} M$ with

$$df = f_{,\mu} dx^\mu$$

and the properties

$$\begin{aligned} d(\alpha_p \wedge \beta_q) &= d\alpha_p \wedge \beta_q + (-1)^p \alpha_p \wedge d\beta_q \\ dd\alpha &= 0 \\ d\alpha_D &= 0. \end{aligned}$$

- Contraction with a vector field $i_Y : \mathfrak{X} \times \Omega_p M \rightarrow \Omega_{p-1} M$
Let the vector field $Y \in \mathfrak{X}$ locally be given by $V = Y^I e_I$. Then the actions on functions and basic one-forms defined by

$$i_Y f = 0 \qquad i_Y \vartheta^I = Y^I$$

allow to calculate the action of i_Y on any differential form. One easily proves that

$$\begin{aligned} i_Y(\alpha_p \wedge \beta_q) &= i_Y \alpha_p \wedge \beta_q + (-1)^p \alpha_p \wedge i_Y \beta_q \\ i_X i_Y + i_Y i_X &= 0 \\ i_{fY} &= f i_Y. \end{aligned}$$

In denoting $i_I := i_{e_I}$ one gets the useful formulae

$$\begin{aligned} i_I \vartheta^J &= \delta_I^J \\ \vartheta^I \wedge i_I \alpha_p &= p\ \alpha_p \\ i_I(\vartheta^I \wedge \alpha_p) &= (D-p)\alpha_p \\ i_{I_p} i_{I_{p-1}} \ldots i_{I_1}\ \alpha_p &= \hat{\alpha}_{I_1 I_2 \ldots I_p}. \end{aligned}$$

- Lie derivative with respect to a vector field $£_Y : \mathfrak{X} \times \Omega_p M \rightarrow \Omega_p M$

$$\begin{aligned} £_Y \alpha &:= (i_Y d + d i_Y)\alpha \\ £_Y(\alpha_p \wedge \beta_q) &= £_Y \alpha_p \wedge \beta_q + \alpha_p \wedge £_Y \beta_q \end{aligned}$$

from which

$$£_Y d = d £_Y \qquad £_Y i_Y = i_Y £_Y \qquad i_{[X,Y]} = [i_X, £_Y].$$

E.2.2 Differentiation with Respect to a Form

Let F be a function of a form α, then we take the convention

$$\delta F = \delta\alpha \wedge \frac{\partial F}{\partial \alpha},$$

that is the variation is always to the left. This gives for instance

$$F = \alpha_p \wedge \beta_q : \qquad \frac{\partial F}{\partial \alpha_p} = \beta_q \qquad \frac{\partial F}{\partial \beta_q} = (-1)^{pq}\alpha_p .$$

This convention has also to be obeyed for the chain rule:

$$F(\alpha) = f(g(\alpha)) : \qquad \frac{\partial F}{\partial \alpha} = \frac{\partial g}{\partial \alpha} \wedge \frac{\partial f}{\partial g}.$$

The maps $\mathrm{Der} := \{d, i_Y, £_Y\}$ act on a function of forms as

$$\mathrm{Der}\;\; F(\alpha) = \mathrm{Der}\;(\alpha) \wedge \frac{\partial F}{\partial \alpha}.$$

E.2.3 Hodge Duality*

The Hodge duality operation can be defined for orientable (semi)-Riemannian manifolds. A manifold M^D is *orientable*, if there exists a global non-vanishing D-form

$$\alpha_D = \frac{1}{D!}\hat{\alpha}_{I_1 I_2 \dots I_D}\vartheta^{I_1 I_2 \dots I_D}.$$

A metric $g = g_{IJ}\vartheta^I \otimes \vartheta^J$ is an element of $T_2^* M$. On an orientable semi-Riemannian manifold the canonical D-form (volume form) is

$$\eta := \sqrt{|\,g\,|}\vartheta^1 \wedge \vartheta^2 \dots \wedge \vartheta^D. \tag{E.4}$$

The Hodge $*$ operation is a mapping $* : \Omega_p M \to \Omega_{D-p} M$ defined on the chain of forms (occasionally called *Trautmann forms*)

$$\eta = *1, \quad \eta^I = *\vartheta^I, \quad \dots, \quad \eta^{IJK\dots} = *\vartheta^{IJK\dots}, \quad \dots, \quad \eta^{I_1 I_2 \dots I_D} = *\vartheta^{I_1 I_2 \dots I_D}.$$

Sometimes it is more useful to work with the forms $\eta_{IJK\ldots}$. They are successively derived from the D-form η as $\eta_I = i_I\eta$, $\eta_{IK} = i_K i_I \eta, \ldots$,

$$i_I \eta_{I_1 I_2 \ldots I_p} = \eta_{I_1 I_2 \ldots I_p I}$$

such that $\eta_{I_1 I_2 \ldots I_D}$ is the Levi-Civita tensor in D dimensions. The set of η's may serve as bases of D, D-1, .. forms. They satisfy the identities

$$\vartheta^I \wedge \eta_{I_1 I_2 \ldots I_p} = p\ \delta^I_{[I_p} \eta_{I_1 I_2 \ldots I_{p-1}]};$$

specifically (and especially in four dimensions) one often needs

$$\begin{aligned}
\vartheta^I \wedge \eta_J &= \delta^I_J \eta \\
\vartheta^I \wedge \eta_{JK} &= \delta^I_K\ \eta_J - \delta^I_J\ \eta_K \\
\vartheta^I \wedge \eta_{JKL} &= \delta^I_J\ \eta_{KL} + \delta^I_K\ \eta_{LJ} + \delta^I_L\ \eta_{JK} \\
\vartheta^I \wedge \eta_{JKLM} &= \delta^I_M\ \eta_{JKL} - \delta^I_L\ \eta_{JKM} + \delta^I_K\ \eta_{JLM} - \delta^I_J\ \eta_{KLM}.
\end{aligned}$$

Without proof I state the following relations (σ being the signature of the metric):

$$*(*\alpha_p) = (-1)^{p(D-p)+\sigma}\ \alpha_p \tag{E.5a}$$
$$\alpha_p \wedge *\beta_p = \beta_p \wedge *\alpha_p \tag{E.5b}$$
$$i_I * \alpha = (-1)^p * (\vartheta_I \wedge \alpha) \tag{E.5c}$$
$$\vartheta_I \wedge *\alpha = (-1)^{p+1} * i_I \alpha. \tag{E.5d}$$

Some Useful Variations

In dealing with Lagrangians and symmetries one typically performs variations. We need

$$\delta\eta_{I_1 I_2 \ldots I_p} = \delta\vartheta^I \wedge \eta_{I_1 I_2 \ldots I_p I} = \delta\vartheta^I \wedge i_I \eta_{I_1 I_2 \ldots I_p} \tag{E.6a}$$
$$\begin{aligned}\delta(\alpha \wedge *\beta) = {}& \delta\alpha \wedge *\beta + \alpha \wedge *\delta\beta \\ & + \alpha \wedge (\delta\vartheta^I \wedge i_I * \beta) - \alpha \wedge *(\delta\vartheta^I \wedge i_I \beta)\end{aligned} \tag{E.6b}$$
$$\delta(i_I\alpha) = x i_I \delta\alpha - (i_I \delta e^J) \wedge i_J \alpha. \tag{E.6c}$$

One often needs (E.6b) for expressions $\delta(\alpha \wedge *\alpha)$. By the use of (E.5b) this becomes

$$\delta(\alpha \wedge *\alpha) = 2\delta\alpha \wedge *\alpha + \delta\vartheta^I \wedge [i_I(\alpha \wedge *\alpha) - 2 i_I \alpha \wedge *\alpha]. \tag{E.7}$$

E.2.4 Integration of Differential Forms and Stokes's Theorem

Integration requires an integration measure: This must transform in an appropriate way under a change of integration variables: For example a volume element dx^1dx^2 in two dimensions transforms under a change $x^i \to y^k(x)$ as $dx^1dx^2 = \det\left(\frac{\partial x}{\partial y}\right)dy^1dy^2$, that is with the Jacobi determinant. In the exterior calculus, the canonical D-form (E.4) transforms correctly as a volume element as can be seen by writing it in local coordinates as $\eta = dx^1 \wedge dx^2 \cdots \wedge dx^D = \frac{1}{D}\epsilon_{\mu_1\cdots\mu_D}dx^{\mu_1}\cdots dx^{\mu_D}$. This allows one to define for any D-form $\omega = \hat{\omega}\eta$ the integral in the domain of a chart $(U_\alpha, \varphi_\alpha)$ with coordinates $(x^1, \cdots, x^D)$ by the ordinary integral:

$$\int_{U_\alpha} \hat{\omega}\eta = \int_{\varphi_\alpha(U_\alpha)} \hat{\omega}(\varphi_\alpha^{-1}(x))\, dx^1 \cdots dx^D = \int_{R_\alpha} \hat{\omega} dV.$$

The integral of ω over the whole manifold is finally given by a finite sum of integrations from different charts:

$$\int_M \omega = \sum_\alpha \int_{U_\alpha} \hat{\omega}_\alpha \eta.$$

Stokes' Theorem

Let M be a D-dimensional orientable compact manifold with a non-empty boundary ∂M and let α be a $(D-1)$ form. Then

$$\int_M d\alpha = \int_{\partial M} \alpha. \tag{E.8}$$

This generalizes integration theorems we are acquainted with from one, two, and three dimensions:

In the 1-dimensional case take $M = [a, b]$, $\partial M = \{a, b\}$. Now $\alpha(x)$ is a scalar function with $d\alpha(x) = \frac{d\alpha}{dx}dx$ and Stokes' theorem is nothing but ordinary integration:

$$\int_M d\alpha = \int_a^b \frac{d\alpha}{dx}dx = \alpha(b) - \alpha(a) = \int_{\partial M} \alpha.$$

In the 2-dimensional case take for M the disk $M = \{(x_1, x_2)|x_1^2 + x_2^2 \leqslant 1\}$, such that $\partial M = \{(x_1, x_2)|x_1^2 + x_2^2 = 1\}$ is a circle. For the one-form $\alpha = \alpha_1(x)dx^1 + \alpha_2(x)dx^2 := \vec{A}(x)d\vec{x}$ we have generically

$$d\alpha = \frac{\partial \alpha_j}{\partial x^i}dx^i dx^j := \frac{1}{2}\epsilon_{ijk}B_k dx^i dx^j = \vec{B}d\vec{f}$$

with $B_k = \epsilon_{kij}\partial_i A_j$, $df_k = \frac{1}{2}\epsilon_{kij} dx^i dx^j$. From this, we regain what one learns as Stokes' theorem in electrodynamics

$$\int_M d\alpha = \int_{\text{disk}} (\vec{\nabla} \times \vec{A}) d\vec{f} = \int_{\text{circle}} \vec{A}(x) d\vec{x} = \int_{\partial M} \alpha$$

in terms of the vector potential $\vec{A}$ and the magnetic field $\vec{B}$.

Quite in analogy one recovers in three dimensions

$$\int_M d\alpha = \int_{\text{ball}} (\vec{\nabla}\vec{E}) dV = \int_{\text{sphere}} \vec{E}(x) d\vec{f} = \int_{\partial M} \alpha$$

with a two form $\alpha = \frac{1}{2}\alpha_{ik}(x) dx^i dx^j := \vec{E}(x) d\vec{f}$ and its derivative $d\alpha = (\partial_k E_k) dV$. If $\vec{E}$ is interpreted as the electric field, this becomes what in electrodynamics is called Gauss' law.

Coderivative or Adjoint Exterior Derivative

In the space of p-forms one can define an inner product by

$$(\alpha_p, \beta_p) := \int_M \alpha_p \wedge *\beta_p.$$

Because of (E.5b), the inner product is symmetric, $(\alpha_p, \beta_p) = (\beta_p, \alpha_p)$, and positive definite: $(\alpha_p, \alpha_p) \geqslant 0$, with equality only for $\alpha_p = 0$. Now consider

$$(d\alpha_{p-1}, \beta_p) = \int_M d\alpha_{p-1} \wedge *\beta_p = \int_M d(\alpha_{p-1} \wedge *\beta_p) + (-1)^p \int_M \alpha_{p-1} \wedge d * \beta_p.$$

The first term vanishes due to Stokes' theorem (E.8) in the case of a manifold without boundary ($\partial M = 0$). The coderivative of a form is defined as a map $d^\dagger : \Omega_p M \to \Omega_{p-1} M$, where

$$d^\dagger \alpha_p := (-1)^{D(p+1)+\sigma+1} * d * \alpha_p \tag{E.9}$$

such that for a manifold without boundary by (E.5a), $(d\alpha_{p-1}, \beta_p) = (\alpha_{p-1}, d^\dagger \beta_p)$. From the external derivative and the coderivative one defines the Laplace-deRham operator as a map $d : \Omega_p M \to \Omega_p M$ by

$$\triangle := dd^\dagger + d^\dagger d.$$

One easily shows that it commutes with the Hodge star, the external derivative, and the coderivative. Furthermore–and this is important for many applications–it is a positive operator:

$$(\alpha_p, \triangle\alpha_p) = (\alpha_p, dd^\dagger\alpha_p) + (\alpha_p, d^\dagger d\alpha_p) = (d^\dagger\alpha_p, d^\dagger\alpha_p) + (d\alpha_p, d\alpha_p) \geqslant 0.$$

From this follows that $\triangle\alpha_p = 0$ is equivalent to $d\alpha_p = 0$ together with $d^\dagger\alpha_p = 0$.

E.2.5 Poincaré Lemma and de Rham Cohomology

Harmonic, Closed, Coclosed and Exact Forms

Definitions: A p-form α is called *harmonic* if $\triangle\alpha_p = 0$, *closed* if $d\alpha = 0$, and *coclosed* if $d^\dagger\alpha = 0$. Furthermore α_p is called *exact*, if there exists globally a form β_{p-1} such that $\alpha_p = d\beta_{p-1}$. Because of $d^2 = 0$ any exact form is a closed form. In analogy, α_p is called *coexact*, if there exists globally a form β_{p+1} such that $\alpha_p = d^\dagger\beta_{p+1}$.

Due to a theorem by W. V. A. Hopf, any p-form on a compact orientable manifold without boundary can be uniquely decomposed as

$$\alpha_p = d\beta_{p-1} + d^\dagger\gamma_{p+1} + \delta_p,$$

where δ_p is harmonic.

The question of whether a closed form is exact depends on the topology of the domain in question: There is a lemma by Poincaré stating that any closed form is locally exact, and that on $\mathbb{R}^D$ (or on contractible manifolds) every closed form with $p > 0$ is globally exact; for a proof see e.g. [192].

Take as an example the non-contractible manifold $M = \mathbb{R}^2\backslash\{0\}$ and a one-form $\alpha = \alpha_i dx^i$ with

$$\alpha_1 = -\frac{x^2}{r^2}, \qquad \alpha_2 = \frac{x^1}{r^2} \qquad r^2 = (x^1)^2 + (x^2)^2.$$

The form α is closed since

$$d\alpha = \Big(\frac{\partial}{\partial x^k}\frac{x^k}{r^2}\Big)dx^1 dx^2 = 0.$$

We can write $\alpha = d\beta = d(\arctan\frac{x^2}{x^1})$, but $\beta(x^1, x^2)$ is not differentiable on the whole manifold M. Thus α is not exact on M. Exactness can however be achieved if one cuts the plane–for instance along the negative x^1 axis. Therefore on $M' = \mathbb{R}^2\backslash\{(x^1, x^2)|x^1 \leqslant 0, x^2 = 0\}$ the one-form α is exact.

De Ram Cohomology

The question under which circumstances a closed form is exact, arises in various physical contexts. So for instance gauge equivalent gauge potentials differ by an exact form. Another example is a boundary term in an action which, although it does not influence the dynamical equations, nevertheless relates to the boundary behavior of the fields and to the canonical structure of the theory. Therefore an understanding, whether and "how much" closed-ness differs from exact-ness is relevant to questions of invariance and symmetries. The appropriate mathematics is the de Rham cohomology. At first some definitions:
Define as the p-cocycles of a manifold M all its closed forms, and as p-coboundaries its exact forms:

$$Z^p(M) = \{\alpha \in \Omega_p M |\, d\alpha = 0\} \qquad\qquad B^p(M) = \{\alpha \in \Omega_p M | \alpha = d\beta\}.$$

Since the exterior derivative is a mapping $d : \Omega_p \to \Omega_{p+1}$ we observe that $Z^p(M)$ is the kernel and $B^p(M)$ is the image of d with:

$$\text{Im } (d : \Omega_{p-1} \to \Omega_p) \ \subset \ \text{Ker } (d : \Omega_p \to \Omega_{p+1}).$$

Obviously $B^p(M) \subseteq Z^p(M)$ because every exact p-form is closed. But when are the sets of closed and exact p-forms on a manifold the same? A measure for the deviation of exact-ness from closed-ness is the *de Rham cohomology group*, defined as the quotient

$$H^p(M) := Z^p(M)/B^p(M).$$

What does this amount to in specific examples? (1) At first consider $H^0(M)$. Since for functions $f \in \Omega_0$, only sense can be given to $B^0(M) = \{0\}$. On the other hand, a function is closed iff $df = \frac{\partial f}{\partial x^k} dx^k = 0$, that iff its partial derivatives vanish in any coordinate system. If M is connected, a closed form can only be a constant, i.e. a real number. Therefore with $Z^0(M) \cong \mathbb{R}$ we get $H^0(M) \cong \mathbb{R}$ for all connected manifolds. If M has k connected components, results $H^0(M) \cong \mathbb{R}^k$. (2) Due to the Poincaré lemma, $Z^p(\mathbb{R}^D) \cong B^p(\mathbb{R}^D)$ for $p > 0$, it follows that $H^p(\mathbb{R}^D) \cong \{0\}$ (except for $H^0(\mathbb{R}^D) \cong \mathbb{R}$). (3) Taking instead of $\mathbb{R}$ the manifold S one can prove that $H^1(S) \cong \mathbb{R}$, exhibiting the other topology of the circle compared to the real line. Remarkably, although defined in terms of the differentiable structure on the manifold, the de Rham cohomology groups solely depend on the topological structure. The dimension of $H^p(M)$ is called the p-th Betti number of M. The Euler characteristic of a manifold, a pure topological quantity depending on the number of vertices, edges and faces of a polyhedron homeomorphic to M can be expressed as the alternating sum of Betti numbers; for more details see e.g. [192].

E.3 Manifolds with Connection

Connections, covariant derivatives, curvature, and torsion for Riemann-Cartan geometries are treated in Subsect. 7.3. But these notions can be broadened to have a meaning for more general manifolds and for fibre bundles. The most general definitions are stated here in an axiomatic style. But I also indicate that these lead to the notions as commonly used in Riemann-Cartan geometries.

Definition: Linear Connection

A linear connection ∇ on a differentiable manifold M is a mapping

$$\nabla : T^q_p M \to T^q_{p+1} M$$

which (1) is linear: $\nabla(t^q_p + s^q_p) = \nabla\ t^q_p + \nabla\ s^q_p$; (2) behaves as a derivation: $\nabla(t^q_p \otimes s^{q'}_{p'}) = \nabla t^q_p \otimes s^{q'}_{p'} + t^q_p \otimes \nabla s^{q'}_{p'}$; (3) acts on functions f as $\nabla f = df$; (4) commutes with contractions: $(\nabla\langle X, \omega\rangle)(Y) = (\nabla X)(\omega, Y) + (\nabla\omega)(X, Y)$.

E.3.1 Linear Connection on Tensor Fields

… on Vector Fields

The definition of a linear connection if applied to a vector field X results at first in

$$\nabla X = \nabla(X^I e_I) = dX^I \otimes e_I + X^I \nabla e_I .$$

Now ∇e_I must be a T^1_1 tensor:

$$\nabla e_I = \Gamma^K{}_{JI}\ \vartheta^J \otimes e_K = \omega^K{}_I\ \otimes e_K$$

with the connection one-form ω for which $\omega^K{}_I(e_J) = \Gamma^K{}_{JI}$. The connection on a vector field can then be written as

$$\nabla X = (dX^I + X^K \Gamma^I{}_{JK}\ \vartheta^J) \otimes e_I =: (\nabla_I X^K)\vartheta^I \otimes e_K ,$$

with $\nabla_I X^K = e_I \cdot X^K + X^J \Gamma^K{}_{IJ}$.

… on Covector Fields aka One-Forms

Because of condition (4) in the definition of a linear connection, one gets on the basic one-form fields: $(\nabla\vartheta^I)(e_J, e_K) = -(\nabla e_J)(\vartheta^I, e_K) = -\Gamma^I{}_{KJ}$, or

$$\nabla\vartheta^I = -\Gamma^I{}_{KJ}\, \vartheta^K \otimes \vartheta^J = -\vartheta^K \otimes \omega^I{}_K.$$

Then, for any one-form $\omega = \omega_I \vartheta^I$ one obtains from $\nabla\omega = d\omega_I \otimes \vartheta^I + \omega_I \nabla\vartheta^I$

$$\nabla\omega = (\nabla_J \omega_I)\vartheta^J \otimes \vartheta^I = (e_J \cdot \omega_I - \omega_K \Gamma^K{}_{IJ})\, \vartheta^J \otimes \vartheta^I.$$

… on Higher Tensor Fields

On arbitrary higher tensor fields, the linear connection can be derived from ∇e_I and $\nabla\vartheta^I$ by applying the product rule (2) in the definition of the linear connection.

E.3.2 Covariant Derivative

The covariant derivative ∇_Y along a vector field Y is a mapping $\nabla_Y : \mathfrak{X}M \times T^q_p M \to T^q_p M$ which (1) is linear; (2) behaves as a derivation; (3) acts on functions as $\nabla_Y f = Y \cdot f = (df)(Y)$; (4) obeys $\nabla_{fY} = f \nabla_Y$; (5) commutes with contractions: $\nabla_Y(\langle X, \omega\rangle) = \langle \nabla_Y X, \omega\rangle + \langle X, \nabla_Y \omega\rangle$.
Defining

$$\nabla_{e_I} e_J = \Gamma^K{}_{IJ}\, e_K = \omega^K{}_J(e_I)\, e_K$$

one can–due to the defining properties of the covariant derivative–deduce its action of any tensor fields

$$\nabla_Y X = Y^I (\nabla_{e_I} X^J)\, e_J \qquad \nabla_Y \omega = -Y^I (\nabla_{e_I} \omega_J)\, \vartheta_J$$
$$(\nabla_Y t)(X_1, \cdot\cdot, X_q, \omega_1, \cdot\cdot, \omega_p) = Y \cdot t(X_1, \cdot\cdot, X_q, \omega_1, \cdot\cdot, \omega_p)$$
$$-t(\nabla_Y X_1, \cdot\cdot, X_q, \omega_1, \cdot\cdot, \omega_p) - \cdots - t(X_1, \cdot\cdot, X_q, \nabla_Y \omega_1, \cdot\cdot, \omega_p) - \cdots.$$

E.3.3 Torsion and Curvature

Torsion

The torsion is defined by

$$T(X, Y) := (\nabla_X Y - \nabla_Y X) - [X, Y]. \tag{E.10}$$

This a mapping $T\ :\ \mathfrak{X}M\times\mathfrak{X}M\ \rightarrow\ \mathfrak{X}M$. From its definition, the torsion (1) is linear; (2) is antisymmetric; (3) obeys $T(fX, gY) = fgT(X, Y)$. Both $(\nabla_X Y - \nabla_Y X)$ and $[X, Y]$ are vector fields, but only the specific combination (E.10) has the property (3)

For a basis $\{e_K\}$ in $\mathfrak{X}M$, the torsion becomes due to previous results

$$T(e_I, e_J) := \nabla_{e_I} e_J - \nabla_{e_J} e_I - [e_I, e_J] = \Gamma^K_{IJ} e_K - \Gamma^K_{JI}\, e_K - C^K_{IJ}\, e_K =: T^K_{IJ}\, e_K,$$

and for instance in a holonomic basis $\{\partial_\lambda\}$, one verifies $T^\lambda{}_{\mu\nu} = \Gamma^\lambda{}_{\mu\nu} - \Gamma^\lambda{}_{\nu\mu}$ as the expression for the torsion in a Riemann-Cartan space.

The torsion can also be interpreted as a vector-valued 2-form $T \in \Omega_2(M, \mathfrak{X}M)$, that is

$$\begin{aligned} T(X, Y) &= \Theta^K(X, Y)\, e_K \\ \Theta^K &= \Theta^K(e_I, e_J)\vartheta^I \otimes \vartheta^J = \Theta^K(e_I, e_J)\, \vartheta^I \wedge \vartheta^J = \frac{1}{2} T^K{}_{IJ}\, \vartheta^I \wedge \vartheta^J . \end{aligned}$$

In reformulating the defining equation for the torsion by the one-forms ϑ and ω one obtains its differential form expression (E.11) below. The essential steps for this derivation are as follows:

$$\begin{aligned} T(X, Y) &= T(X^I e_I, Y^J e_J) \\ &= \Big[\nabla_X Y^K - \nabla_Y X^K - [X, Y]^K\Big] e_K + \Big[Y^K \nabla_X - X^K \nabla_Y\Big] e_K . \end{aligned}$$

Now, with $\nabla_X Y^K = X \cdot \vartheta^K(Y)$ and $\nabla_X e_K = \omega^L{}_K(X) e_L$ the previous expression becomes

$$\begin{aligned} T(X, Y) = \Big[X \cdot \vartheta^K(Y) - Y \cdot \vartheta^K(X) - \vartheta^K([X, Y])\Big]\, e_K \\ + \Big[\vartheta^K(Y)\omega^L{}_K(X) - \vartheta^K(X)\omega^L{}_K(Y)\Big]\, e_L . \end{aligned}$$

The first term can be identified with $(d\vartheta^K)(X, Y)e_K$ and the second one with $(-(\vartheta^K \wedge \omega^L{}_K)(X, Y))e_L$, and therefore indeed

$$T(X, Y) = \Big[d\vartheta^K - \vartheta^L \wedge \omega^K{}_L\Big](X, Y)\, e_K .$$

Curvature

The curvature is defined by

$$R(X, Y, Z) := \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]} Z .$$

Thus it is a mapping $R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$. From its definition, the curvature (1) is linear; (2) is antisymmetric in its first two arguments; (3) obeys $R(fX, gY, hZ) = fgh\ R(X, Y, Z)$. The name curvature may be justified in that $R(\partial_\mu, \partial_\nu, \partial_\rho) = R^\sigma{}_{\mu\nu\rho}\partial_\sigma$. Like for the torsion also the curvature can be interpreted as a vector-valued two-form:

$$R(X, Y, e_I) = \Omega_I^K(X, Y)e_K \qquad \Omega_I^K = \Omega_I^K(e_J, e_L)\vartheta^J \wedge \vartheta^L = \frac{1}{2}R^K_{IJL}\vartheta^J \wedge \vartheta^L.$$

and again, a half-page calculation then gives the expression of the curvature in terms of differential forms, namely (E.12).

Bianchi Identities

Torsion and curvature are defined in terms of the D-beins ϑ^I and spin connections $\omega^I{}_J$ through the Cartan structural equations

$$\Theta^I = d\vartheta^I + \omega^I{}_J \wedge \vartheta^J = \frac{1}{2}T^I{}_{KJ}\vartheta^{KJ} \tag{E.11}$$

$$\Omega^I{}_J = d\omega^I{}_J + \omega^I{}_K \wedge \omega^K{}_J = \frac{1}{2}R^I{}_{JKL}\vartheta^{KL}. \tag{E.12}$$

Taking the exterior derivative of these equations one arrives at the *Bianchi identities*

$$\begin{aligned}
d\Theta^I &= d\omega^I{}_J \wedge \vartheta^J - \omega^I{}_J \wedge d\vartheta^J \\
&= d\omega^I{}_J \wedge \vartheta^J - \omega^I{}_J \wedge (\Theta^J - \omega^J{}_K \wedge \vartheta^K) \\
&= \Omega^J{}_J \wedge \vartheta^J - \omega^I{}_J \wedge \Theta^J \\
d\Omega^I{}_J &= d\omega^I{}_K \wedge \omega^K{}_J - \omega^I{}_K \wedge d\omega^K{}_J \\
&= (\Omega^I{}_K - \omega^I{}_L \wedge \omega^L{}_K) \wedge \omega^K{}_J - \omega^I{}_K \wedge (\Omega^K{}_J - \omega^K{}_L \wedge \omega^L{}_J) \\
&= \omega^K{}_J \wedge \Omega^I{}_K - \omega^I{}_K \wedge \Omega^K{}_J.
\end{aligned}$$

The covariant derivative D with respect to the spin-connection is

$$\begin{aligned}
Df^{I_1..I_k}_{J_1,..,J_l} = df^{I_1,..,I_k}_{J_1,..,J_l} &+ \omega^{I_1}_L \wedge f^{L,..,I_k}_{J_1,..,J_l} + .. + \omega^{I_k}_L \wedge f^{I_1,..,L}_{J_1,..,J_l} \\
&- \omega^L_{J_1} \wedge f^{I_1,..,I_k}_{L,..,J_l} - .. - \omega^L_{J_k} \wedge f^{I_1,..,I_k}_{J_1,..,L}.
\end{aligned}$$

Applying the covariant derivative twice we obtain for a form with an upper Lorentz index

$$
\begin{aligned}
DDf^I &= D(df^I + \omega^I{}_J \wedge f^J) \\
&= \omega^I{}_J \wedge df^J + d(\omega^I{}_J \wedge f^J) + \omega^I{}_K \wedge (\omega^K{}_J \wedge f^J) \\
&= d\omega^I{}_J \wedge f^J + \omega^I{}_K \wedge \omega^K{}_J \wedge f^J + \omega^I{}_J \wedge df^J - \omega^I{}_J \wedge df^J \\
&= \Omega^I{}_J \wedge f^J.
\end{aligned}
$$

Similarly one derives:

$$
\begin{aligned}
DDf^I &= \Omega^I{}_J \wedge f^J, \\
DDg_K &= -\Omega^J{}_K \wedge g_J, \\
DDh^I{}_K &= \Omega^I{}_J \wedge h^J{}_K - \Omega^L{}_K \wedge h^I{}_L.
\end{aligned}
\tag{E.13}
$$

In terms of the covariant derivative the Bianchi identities become

$$
D\Theta^I = \Omega^I{}_J \wedge \vartheta^J \qquad D\Omega^I{}_J = 0. \tag{E.14}
$$

The torsion itself can be written as $\Theta^I = D\vartheta^I$, but observe that $\Omega^I{}_J \neq D\omega^I{}_J$.

E.4 Lie Groups

A D-dimensional Lie group[5] **G** has both properties of a D-dimensional manifold and a D-parameter group with a compatibility of the differentiability in the manifold with the group composition: The group operation induces a C^∞ map of the manifold into itself. All the previously introduced definitions for manifolds can be taken over, but the surplus group structure allows to sharpen and refine the terminology. Points in the manifold G are now the group elements $g \in G$, and as we will see, the Lie algebra generators X_a are the basis of a vector space isomorphic to T_eG, that is the tangent space of the Lie group at the group identity e.

E.4.1 Lie Algebra

In Appendix A.2, the Lie algebra associated to a Lie group **G** is introduced by expanding group elements near the group identity e in a Taylor series in local coordinates, and by relating the group structure to generators. Here it will be shown that the relation of a Lie group to its associated Lie algebra and the representation issues can be understood in generic geometrical terms.

[5] In both this and the next subsection I deviate from the notation for Lie groups in bold face font used elsewhere in this book. This not only supports readability but may indeed be justified by accentuating the manifold facet of Lie groups.

Any element g of G maps $h \in G$ to either $h \mapsto gh$ (*left translation by* g) or to $h \mapsto hg$ (*right translation by* g). For left translations[6]:

$$L_g : G \to G \qquad L_g h = gh.$$

These maps induce push-forward maps in tangent spaces: $L_{g^*} : T_h G \to T_{gh} G$. A vector field $\bar{X}$ on G is called left-invariant if L_{g^*} maps $\bar{X}$ at h to $\bar{X}$ at gh:

$$L_{g^*} : \bar{X}_h = \bar{X}_{L_g h} = \bar{X}_{gh} \qquad \forall \;\; g, h \in G.$$

A vector $A \in T_e G$ generates a unique left-invariant vector field on the manifold G by $(X_A)_g = L_{g^*} A$.

The left-invariant vector fields on G form a D-dimensional real vector space. Furthermore, the Lie bracket of two left-invariant vector fields is left-invariant, because of

$$L_{g^*}[\bar{X}, \bar{Y}]_h = L_{g^*}(\bar{X}_h \bar{Y}_h - \bar{Y}_h \bar{X}_h) = [L_{g^*}\bar{X}_h, L_{g^*}\bar{Y}_h] = [\bar{X}_{gh}, \bar{Y}_{gh}] = [\bar{X}, \bar{Y}]_{gh}.$$

The Lie algebra $\mathfrak{g}$ associated to the Lie group G is defined by the Lie algebra of the left-invariant vector fields. Since any left-invariant vector field is uniquely defined by its value at the group unit element $(\bar{X}_g = L_{g^*}\bar{X}_e)$, the Lie algebra $\mathfrak{g}$ is as a vector space isomorphic to the tangent space $T_e G$ to the group manifold G at the unit element.

Let $X_a \quad (a = 1, \ldots, D)$ be a basis in $\mathfrak{g}$. Then holds $[X_a, X_b] = f^c{}_{ab} X_c$ with structure constants $f^c{}_{ab}$ which completely characterize the Lie algebra (and also the group in the neighborhood of the identity). The algebra $\mathfrak{g}$ inherits the properties of the group G near the identity element.

It is also possible to define left-invariant forms of G: The form $\bar{\omega} \in \Omega_p G$ is called left-invariant if

$$L_g^* \, \bar{\omega}_{gh} = \bar{\omega}_h \qquad L_g^* : T_{gh}^* G \to T_h^* G.$$

The vector space $\mathfrak{g}^*$ of left-invariant one-forms is dual to $\mathfrak{g}$: If $\bar{X} \in \mathfrak{g}$ and $\bar{\omega} \in \mathfrak{g}^*$, the internal product $\bar{\omega}(\bar{X}) = \langle \bar{\omega}, \bar{X} \rangle$ is constant on G. Let ω^a be a basis in $\mathfrak{g}^*$. Then by the duality property $\langle \omega^a, X_b \rangle = \delta^a_b$. Now consider the exterior derivative of the two-form $d\omega$ evaluated in the X^a basis:

$$d\omega^a(X_b, X_c) = X_b(\langle \omega^a, X_c \rangle) - X_c(\langle \omega^a, X_b \rangle) - \langle \omega^a, [X_b, X_c] \rangle.$$

The first two terms vanish because the arguments are constants, the last term can be expressed via the structure constants, with the result $d\omega^a(X_b, X_c) = -f^a{}_{bc} X_c$. On the other hand, $(\omega^d \wedge \omega^e)(X_b, X_c) = \delta^d_b \delta^e_c - \delta^d_c \delta^e_b$. Combining this with the previous

[6] In the following only those structures defined by left translations will be treated. The ones for right translations are defined in analogy.

expression results in the Maurer-Cartan (structure) equation

$$d\omega^a + \frac{1}{2} f^a{}_{bc}\, \omega^b \wedge \omega^c = 0 \tag{E.15}$$

holding for any left-invariant one-form. This equation is to be interpreted as the counterpart in $\mathfrak{g}^*$ of $[X_a, X_b] = f^c{}_{ab} X_c$ which are valid in $\mathfrak{g}$.
The canonical one-form or *Maurer-Cartan form* $\hat{\Omega}$ on **G** is a left-invariant Lie algebra valued one-form, defined by

$$\hat{\Omega} = X_a \otimes \theta^a$$

where $\{\theta^a\}$ is a basis in $T^*_g G$. The Maurer-Cartan form moves a tangent vector $X_g \in T_g G$ back to the unit: $\hat{\Omega}_g(X_g) = L_{g^{-1}*} X_g = X_e$. For this the Maurer-Cartan equation simply reads $d\hat{\Omega} + \frac{1}{2}\hat{\Omega} \wedge \hat{\Omega} = 0$. The Maurer-Cartan form plays an important role in advanced calculations of Faddeev-Popov ghosts and the BRST quantization procedure of theories with local symmetries.

E.4.2 Group Covariant Derivative

Some of the next definitions and terms can potentially only be fully appreciated in terms of the fibre bundle understanding of gauge theories.

In physics we deal with "fields" Ψ^α which transform according to (irreducible) representations of G, that is $\Psi^\alpha \in \Omega(M, V_d)$, where V_d is the representation space (and $\alpha = 1, \ldots, d$). Under infinitesimal group transformations

$$g = 1 + \gamma = 1 + \gamma^a \otimes X_a$$

the Ψ transform as

$$\delta_\gamma\, \Psi = \hat{\gamma}\Psi, \qquad \text{i.e.} \qquad \delta_\gamma\, \Psi^\alpha = (\hat{\gamma})^\alpha{}_\beta\, \Psi^\beta = \gamma^a (\hat{X}_a)^\alpha{}_\beta\, \Psi^\beta, \tag{E.16}$$

where $(\hat{X}_a)^\alpha{}_\beta$ are d-dimensional representation matrices for the generators X_a.

Since for spacetime-dependent parameter $\gamma^a(x)$ the derivative $d\Psi^\alpha$ does not transform in the same way as Ψ^α itself, a covariant derivative is introduced by help of a connection $A = A^a \otimes X_a \in \Omega_1(M, \mathfrak{g})$, i.e. the components A^a of the Lie-algebra valued vectorfield A are one-forms. The definition is $D\Psi = d\Psi + A \wedge \Psi$, or explicitly

$$D\Psi^\alpha := d\Psi^\alpha + (\hat{A})^\alpha{}_\beta \wedge \Psi^\beta = d\Psi^\alpha + A^a \wedge (\hat{X}_a)^\alpha{}_\beta\, \Psi^\beta. \tag{E.17}$$

Requiring $\delta_\gamma D\Psi = \hat{\gamma} D\Psi$ the connection must transform according to $\delta_\gamma A = -d\hat{\gamma} - [A, \hat{\gamma}]$, which, if split as $\delta_\gamma A = \delta_\gamma A^a \otimes X_a + A^a \otimes \delta_\gamma X_a$ is

$$\delta_\gamma A^a = -d\gamma^a \qquad \text{and} \qquad \delta_\gamma X^a = [\gamma, X^a].$$

Keeping the basis fixed, we also define for the components the variation

$$\hat{\delta}_\gamma A^a = -d\gamma^a - f^a{}_{bc} A^b \gamma^c.$$

Originally defined on elements from $\Omega(M, V_d)$, it is possible to extend the covariant derivative to forms $\alpha \in \Omega(M, \mathfrak{g})$, by defining

$$D\alpha := d\alpha + [A, \alpha] \qquad \text{or} \qquad (D\alpha)^a := d\alpha^a + f^a{}_{bc} A^b \wedge \alpha^c; \tag{E.18}$$

symbolically written as $D = d + [A, -]$. By this we have

$$\delta_\gamma A = -D\gamma. \tag{E.19}$$

E.4.3 Group Curvature

Applying the covariant derivative twice we have (at first in a matrix notation)

$$\begin{aligned} DD\Psi &= D(d\Psi + A \wedge \Psi) = d(d\Psi + A \wedge \Psi) + A \wedge (d\Psi + A \wedge \Psi) \\ &= d(A \wedge \Psi) + A \wedge d\Psi + A \wedge (A \wedge \Psi). \end{aligned}$$

The first two terms are nothing but $dA \wedge \Psi$, the last one can be reformulated as

$$\begin{aligned} A \wedge (A \wedge \Psi) &= A \wedge (A^a \wedge \hat{X}_a \Psi) = A^b \wedge A^a \wedge \hat{X}_b \hat{X}_a \Psi \\ &= \frac{1}{2} A^b \wedge A^a \wedge \left[\hat{X}_b, \hat{X}_a\right] \Psi = \frac{1}{2}(A \wedge A) \wedge \Psi. \end{aligned}$$

This gives rise to the definition of a curvature

$$F = dA + \frac{1}{2} A \wedge A \qquad \text{or} \qquad F^a := dA^a + \frac{1}{2} f^a{}_{bc} A^b \wedge A^c, \tag{E.20}$$

such that

$$DD\Psi^\alpha := (\hat{F})^\alpha{}_\beta \wedge \Psi^\beta.$$

The covariant derivative of the curvature is calculated as

$$\begin{aligned} DF &= D\big(dA + \frac{1}{2} A \wedge A\big) = d\big(dA + \frac{1}{2} A \wedge A\big) + A \wedge \big(dA + \frac{1}{2} A \wedge A\big) \\ &= \frac{1}{2} d(A \wedge A) + A \wedge dA + \frac{1}{2} A \wedge (A \wedge A). \end{aligned}$$

The first two terms cancel and the last term vanishes because of the Jacobi identity:

$$A \wedge (A \wedge A) = \frac{1}{2} A \wedge A^a \wedge A^b \otimes [X_a, X_b] = \frac{1}{4} A^c \wedge A^a \wedge A^b \otimes [X_c, [X_a, X_b]] = 0.$$

Thus there is the Bianchi-identity

$$DF \equiv 0.$$

The transformation of the curvature is calculated as

$$\delta_\gamma F = d\delta_\gamma A + A \wedge \delta_\gamma A = D\delta_\gamma A = -DD\gamma = -[F, \gamma],$$

and thus F transforms homogeneously under the group.

E.4.4 Isometries and Coset Manifolds

Killing Vector Fields and Isometries

For a space with metric $g_{\mu\nu}(x)$, a coordinate transformation $x \to x'$ is called an isometry iff $g'_{\mu\nu}(x') = g_{\mu\nu}(x')$, which can also be written $\bar{\delta} g_{\mu\nu} = 0$. Since a metric transforms like a tensor, i.e.

$$g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x),$$

the isometry condition can equivalently be expressed as

$$g_{\mu\nu}(x) \stackrel{!}{=} \frac{\partial x'^\rho}{\partial x^\mu} \frac{\partial x'^\sigma}{\partial x^\nu} g_{\rho\sigma}(x').$$

In general, this is a complicated system of differential equations for the isometric coordinate transformations $x(x')$. Since we are looking for an isometry group, it is possible and allowed to investigate infinitesimal coordinate transformations

$$x'^\mu = x^\mu + \xi^\mu \qquad \text{with} \qquad |\xi^\mu| \ll 1,$$

for which the isometry condition becomes

$$\begin{aligned} g_{\mu\nu}(x) &\stackrel{!}{\simeq} (\delta^\rho_\mu + \xi^\rho_{,\mu})(\delta^\sigma_\nu + \xi^\sigma_{,\nu}) g_{\rho\sigma}(x + \xi) \\ &\simeq (\delta^\rho_\mu \delta^\sigma_\nu + \delta^\rho_\mu \xi^\sigma_{,\nu} + \delta^\sigma_\nu \xi^\rho_{,\mu})\big(g_{\rho\sigma}(x) + \xi^\lambda g_{\rho\sigma,\lambda}\big). \end{aligned}$$

In order that this equation is fulfilled one arrives at the infinitesimal form of the isometry condition:

$$g_{\mu\nu,\lambda}\xi^{\lambda} + g_{\mu\lambda}\xi^{\lambda}_{,\nu} + g_{\lambda\nu}\xi^{\lambda}_{,\mu} = 0. \tag{E.21}$$

Any solution ξ of (E.21) is called a Killing vector (field) of the metric space. Let $\{k_A\}$ be the complete set of linear independent Killing vectors of the manifold. The coordinate change can be expressed as $\xi^{\mu} = \epsilon^A k^{\mu}_A$.
Some further remarks on Killing vectors:

- In a Riemann space V the condition (E.21) can be written in terms of covariant derivatives as

$$\xi_{\mu;\nu} + \xi_{\nu;\mu} = 0.$$

- Since isometries form a continuous group they have an associated Lie algebra with generators X_A fulfilling $[X_A, X_B] = f_{ABC}X_C$.
- The generators can be expressed by the Killing vectors as $X_A := k^{\mu}_A\partial_{\mu}$. Therefore

$$k^{\mu}_A\,\partial_{\mu}k^{\nu}_B - k^{\mu}_B\,\partial_{\mu}k^{\nu}_A = f_{AB}^{\ \ C}k^{\nu}_C. \tag{E.22}$$

- The Killing condition can also be formulated by the Lie derivative as $\pounds_{\xi}g_{\mu\nu} = 0$. The Lie bracket of two Killing vector fields is again a Killing vector. The Killing vector fields on a Riemann manifold V thus form a Lie subalgebra of vector fields on V. This is the Lie algebra of the isometry group of V.
- The maximal number of linear independent Killing vectors in D dimensions is $\frac{D(D+1)}{2}$.

These notions shall be illustrated at the example of S^2, the surface of a sphere with radius r. In polar coordinates $(x_1 = \theta, x_2 = \varphi)$ the metric is given by

$$g_{ij} = r^2\begin{pmatrix}1 & 0\\ 0 & \sin^2\theta\end{pmatrix}.$$

There are three Killing Eqs. (E.21) corresponding to

$$\xi^{\theta}_{,\theta} = 0 \qquad \sin^2\theta\,\xi^{\varphi}_{,\theta} + \xi^{\theta}_{,\varphi} = 0 \qquad \cos\theta\,\xi^{\theta} + \sin\theta\,\xi^{\varphi}_{,\varphi} = 0.$$

The solutions of these differential equations

$$\xi^{\theta} = A\sin(\varphi + \varphi_0) \qquad\qquad \xi^{\varphi} = a + A\cos(\varphi + \varphi_0)\cot\theta$$

depend on three free parameter (A, φ_0, a) giving rise to three independent Killing vectors, which may be written as

$$k_1 = \begin{pmatrix}\sin\varphi\\ \cos\varphi\cot\theta\end{pmatrix} \qquad k_2 = \begin{pmatrix}\cos\varphi\\ -\sin\varphi\cot\theta\end{pmatrix} \qquad k_3 = \begin{pmatrix}0\\ 1\end{pmatrix}. \tag{E.23}$$

The generators $X_A = k_A^\theta\, \partial_\theta + k_A^\varphi\, \partial_\varphi$ obey the algebra

$$[X_A, X_B] = \epsilon_{ABC} X_C$$

which is isomorphic to $\mathfrak{so}(3)$. This result can be extended to the D-dimensional sphere for which the isometry group is $\mathbf{SO(D+1)}$.

Manifolds which host their maximal number of independent Killing vectors are called *maximally symmetric*. For these the Riemann curvature tensors are uniquely specified by their Riemann curvature as

$$R_{\mu\nu\rho\sigma} = \frac{R}{D(D-1)}\{g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}\}. \tag{E.24}$$

For D=1 this statement is also true because any one-dimensional Riemann manifold is flat.

Group Action

In the context of symmetries we understand symmetry operations as elements of (symmetry) groups. These operate on entities, which generically are sets. A *left action* of a group $\mathbf{G}$ on a set S is a map

$$\Phi : \mathbf{G} \times S \to S$$

such that for $s \in S$ and the unit element $e \in \mathbf{G}$ the identity axiom $\Phi(e, s) = s$ and the associativity axiom $\Phi(g \circ h, s) = \Phi(g, \Phi(h, s))$ are obeyed. For a *right action* holds instead $\Phi(g \circ h, s) = \Phi(h, \Phi(g, s))$. The set S is also called a left (respectively right) G-set. In most applications the sets have also manifold properties and we talk of *G-spaces*. A group action is termed

- *transitive* if for every pair $(s, s') \in S \times S$ there is a group element g, such that $\Phi(g, s) = s'$
- *effective* if $\Phi(g, s) = s \quad \forall s$ implies $g = e$
- *free* if $\Phi(g, s) = s$ forsome s implies $g = e$.

For every $s \in S$ define the *stabilizer* subgroup[7] as

$$\mathbf{Stab_S}(s) := \{g \in \mathbf{G} |\ \Phi(g, s) = s\}.$$

Thus the stabilizer of a point s contains all symmetry transformations in $\mathbf{G}$ which leave s invariant. It is indeed a group since the product of any two leaving the point fixed also leaves it fixed, and since the identity of $\mathbf{G}$ is in this set. This amounts to identifying the set S with the coset space

$$S = \mathbf{G}/\mathbf{Stab_S}.$$

[7] In other contexts also called the isotropy group or the little group.

The stabilizer is a subgroup of **G** which, however, in general is not invariant. The action of **G** on S is free iff all stabilizers are trivial.

A closely related notion is the *orbit*: The orbit of a point $s \in S$ is the subset of S to which s can be moved by the elements of **G**:

$$\mathrm{Orb}_G\ (s) := \{\Phi(g, s)|\ g \in \mathbf{G}\} \subseteq S.$$

Two points s and s' may be defined as being equivalent $s \sim s'$ iff there exists a $g \in \mathbf{G}$ such that $\Phi(g, s) = s'$. The orbits are therefore recognized as equivalence classes under this relation: s and s' are equivalent iff $\mathrm{Orb}_G\ (s) = \mathrm{Orb}_G\ (s')$.

Homogeneous and Isotropic Spaces

Let **G** be the Lie group of isometries of a Riemann manifold V with metric g. As explained above, the Lie algebra associated to **G** is that of the Killing vector fields of g. The manifold is said to be *homogeneous* if its isometry group acts transitively on it. (The term homogeneous reflects that the geometry is the same everywhere in V). The Riemann manifold V is said to be *isotropic* around a point p if there is a nontrivial stabilizer $\mathbf{H_p}$. In textbooks on differential geometry one finds proofs of: A homogeneous space that is isotropic about one point is maximally symmetric. A space that is isotropic around every point, is homogeneous.

In the example, the isometry group of the sphere S^2 is the rotation group $\mathbf{SO(3)}$. It acts transitively on S^2. For every point p on the sphere there exists a subgroup $\mathbf{H_p} \cong \mathbf{SO(2)}$ that leaves p fixed. Forming left and right cosets of $\mathbf{SO(2)} \subset \mathbf{SO(3)}$ one finds that they are identical. Therefore the factor group $\mathbf{SO(3)}/\mathbf{SO(2)}$ is a group. Since we are dealing with Lie groups the factor group itself can be assigned a topology, giving it the properties of a *coset space*. Indeed one can show that

$$S^D = \mathbf{SO(D+1)}/\mathbf{SO(D)}.$$

Further Definitions

Let M be a left G-space: $(g, x) \mapsto gx$. Then a left-invariant vector field X_V generated by $V \in T_eG$ gives rise to an *induced vector field*

$$X_V(x) = \frac{d}{dt} \exp(tV)x\Big|_{t=0}.$$

The homomorphism

$$\mathrm{ad}_h : G \to G \qquad g \mapsto hgh^{-1}$$

is called an *adjoint representation* of G. It induces the tangent space maps ad_{h^*} : $T_gG \to T_{hgh^{-1}}G$, and specifically the adjoint map in the Lie algebra

$$\mathrm{Ad}_h : \mathfrak{g} \to \mathfrak{g} \qquad V \mapsto hVh^{-1}.$$

For a matrix group, this becomes explicitely $\mathrm{Ad}_h(V) = hVh^{-1}$.

E.5 Fibre Bundles

The notion of a fibre bundle generalizes that of a Cartesian product. For the motivation and illustration of fibre bundles, all textbooks use a cylinder and the Möbius strip as examples: Start with a circle S^1 (called the base space) and attach to each point of it a closed interval $D \subset \mathbb{R}$, the fibre. By simply taking the Cartesian product $S^1 \times D$ the geometrical interpretation is that of a cylinder. But this glueing together of the two manifolds can be done in another way: By one twist the resulting object receives the topology of a Möbius strip. Both the cylinder and the Möbius strip are *locally trivial*, that is, locally homoeomorphic to $S^1 \times D$.

We met already another example, namely the tangent bundle in which the tangent spaces T_xM, at points x of a manifold M represent the fibres. And there are further examples abound: Aristotelean space-time E is a *trivial bundle*, namely the Cartesian product $E = T \times S$ of a base space ("time" $T = E^1$) and a three-dimensional space $S = E^3$. Galileian space-time, in which one allows for Galilei-boosts mixing space and time, can be described by a projection $\pi : E \to T$ that associates to any event $p \in E$ the corresponding instant of time $t = \pi(p) \in T$. In the language of fibre bundles, E is called the *total space*, and T the *base space*. The set $\pi^{-1}(t)$ (that is all events that happen simultaneously with p) is called the *fibre* over t. Each fibre is isomorphic to $\mathbb{R}^3$, called the *typical fibre* or *standard fibre*. The Galileian bundle can be trivialized, i.e. there is a smooth map $h : E \to T \times \mathbb{R}^3$

In the context of gauge symmetries the bundle language allows for a proper understanding of gauge potentials, connections, and curvatures. In the main text various structural similarities in the description of Yang-Mills and of gravitational theories became visible, but the dissimilarities are most properly recognized in the language of fibre bundles.

E.5.1 Definition, Various Types, and Examples of Fibre Bundles

Definition of a Fibre Bundle

A fibre bundle is a structure[8] (E, π, M, F, G), where the elements are (1) a topological space E, the total space; (2) a topological space M, the base space; (3) a

[8] Strictly speaking, the whole complex $\mathbb{E} = (E, \pi, M, F, G)$ constitutes a fibre bundle, but one lazily also calls E the fibre bundle.

surjection called the projection $\pi : E \to B$; (4) a topological space F, the standard fibre. This is homeomorphic to the inverse images $\pi^{-1}(x) := F_x$ with $x \in M$; the F_x are called fibres; (5) a *structure group* G of homeomorphism of the fibre F, the fibres being left G-spaces; (6) A set of open neighborhoods U_α covering B, for which for every U_α there is a homeomorphism

$$\varphi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times F \quad \text{where} \quad \pi \cdot \varphi_\alpha^{-1}(x, f) = x \quad x \in U_\alpha \subset B, \ f \in F.$$

The latter feature expresses the intuitive notion of *local triviality*.

These terms are illustrated for the Möbius strip[9] in Fig. E.2.

The structure group G originates from the transitions from one set of local bundle coordinates $(U_\alpha, \varphi_\alpha)$ to another set (U_β, φ_β). If $U_\alpha \cap U_\beta \neq \emptyset$ it makes sense to define $g_{\alpha\beta} = \varphi_\alpha \cdot \varphi_\beta^{-1}$, which are homeomorphic mappings $F \to F$. These *transition functions* constitute the structure group G. They obey what is called a *cocycle condition*: $g_{\alpha\beta} \cdot g_{\beta\gamma} = g_{\alpha\gamma}$ on $U_\alpha \cap U_\beta \cap U_\gamma \neq \emptyset$. Thus specifically $g_{\alpha\alpha} = e$, and taking $\alpha = \gamma$: $g_{\beta\alpha} = g_{\alpha\beta}^{-1}$. By this it is guaranteed that the local pieces of a bundle are glued together in a consistent manner. Of course the topological and geometrical properties of a fibre bundle should not depend on the choice of $(U_\alpha, \varphi_\alpha)$. A change from φ to φ' induces a change in the group element from $g_{\alpha\beta}$ to $g'_{\alpha\beta} = \varphi'_\alpha \cdot \varphi'^{-1}_\beta$. In order that the $g'_{\alpha\beta}$ constitute the structure group, there must exist $\lambda \in G$ such that $g'_{\alpha\beta} = \lambda_\alpha^{-1}\, g_{\alpha\beta}\lambda_\beta$.

The structure group carries information on the global properties of the fibre bundle. For example, the cylinder and the Möbius strip differ from each other in that for the cylinder the structure group is built by the group unit alone, whereas the structure group for the Möbius strip is isomorphic to Z_2, containing also the twist mapping.

When is a fibre bundle globally trivial? It is easily shown that global triviality can be obtained if and only if all transition functions can be chosen to be identity maps. Furthermore, if the base space is a contractible manifold (that is, if it can continuously be shrunk to a point inside the manifold itself) any fibre bundle with this base space is globally trivial. The more general answer is as follows: If and only if the principal bundle associated to a fibre bundle admits a global section, both bundles are globally trivial. This needs some further terminology and definitions.

A *global section*–or section[10] for short–in the bundle (E, π, M, F, G) is defined as a continuous map s from the base space to the total space, $s : M \to E$ such that $\pi \cdot s(x) = x$ for $\forall\, x \in M$. Denote the set of sections by $\Gamma(M, E)$. If a section can only be defined for an open set $U \subset M$, this is called a local section, denoted by $\Gamma(U, E)$.

Vector, Tangent, Frame, Principal, Associated.... Bundles

By imposing additional structures or conditions on a fibre bundle one arrives at specific bundles:

---

9 Taken from [379].

10 Sometimes also called 'cross-section'.

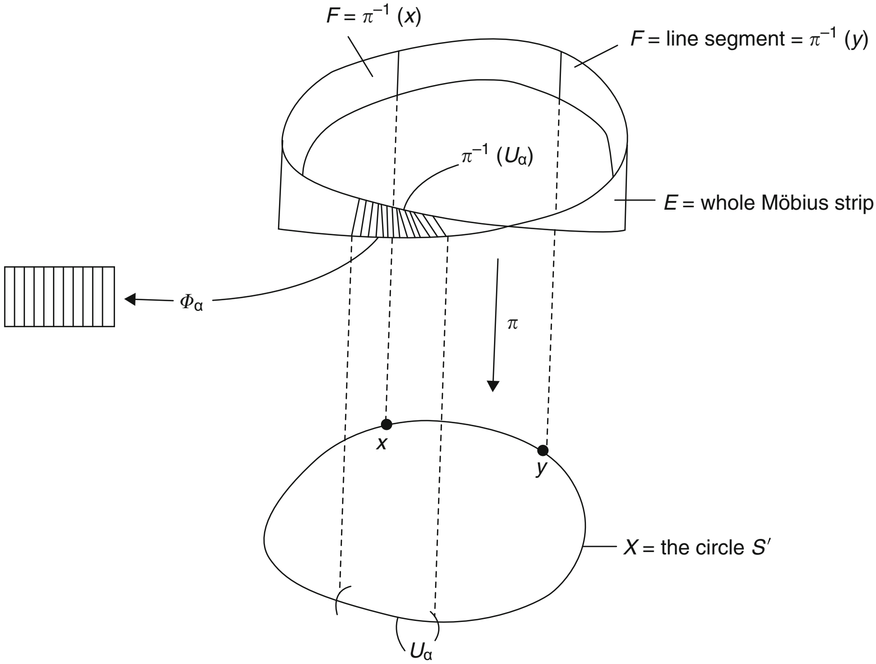


Fig. E.2 The Möbius strip as a fibre bundle

A *vector bundle* (E, π, M, V, G) of rank n is a fibre bundle where the fibre is an n-dimensional vector space V and where the structure group acts linearly on the fibres; thus $G \subseteq GL(n)$.

The *tangent bundle* to a manifold M^D is an example of a vector bundle. It is obtained by choosing the fibre F at a point x of a base manifold M^D to be the tangent space $T_x M^D$. Locally the tangent bundle has the direct product structure $T(U_\alpha) \cong \mathbb{R}^D \times \mathbb{R}^D$, and the structure group is $G = GL(D, \mathbb{R})$. Quite in analogy, one can build the *cotangent bundle*.

Related (below the term ‘associated bundle’ will be defined) to the tangent bundle TM is the *linear frame bundle* LM. A ‘linear frame’ at $x \in M$ is an ordered basis $\{X_1, \ldots, X_D\}$ of $T_x M$. Denote the collection of all linear frames in x by $L_x M$; then the set union over all points of the manifold is called the linear frame bundle LM. Any frame can be reached by acting with an element of $GL(D, \mathbb{R})$ on a fixed frame. Let this group act on LM to the right by

$$R_a : LM \to LM \;\; (x, \{X_1, \ldots, X_D\}) \mapsto (x, \{a_1^i X_i, \ldots, a_D^i X_i\}) \;\; a_i^k \in GL(D, \mathbb{R}).$$

This group action is free and thus the quotient $LM/GL(D, \mathbb{R})$ is a manifold. Define a map from this manifold to M, such that the orbit through $u \in L_x M$ is mapped to $x \in M$. By this the orbit space can be identified with M and there is a natural projection $\pi : LM \to M$. By construction, this defines a bundle in which the fibre is $GL(D, \mathbb{R})$, that is the structure group itself. A linear frame bundle is an example of a principal bundle, defined below.

For any fibre bundle $\mathbb{E} := (E, \pi, M, F, G)$ one can construct a correlated *principal bundle* $\mathbb{P}(\mathbb{E}) = (P, \pi, M, G)$, characterized by replacing the fibre with the structure group ($F \equiv G$). The principal bundle contains the information of whether a fibre bundle can be trivialized globally.

A more formal definition: A *principal G-bundle* (P, π, M, G) is a fibre bundle which is a right G-space such that the group action is free and $\pi(pg) = \pi(p)$ for $p \in P$ and $g \in G$. When restricted to a fibre $\pi^{-1}(x)$, the action is also transitive. This implies that the fibre (i.e. the orbit of any point x) is isomorphic to the structure group itself, and the base space is isomorphic to the orbit space: $M \cong P/G$.

Indeed the homogeneous space construction directly leads to a principal bundle: Due to a theorem by Cartan, any topologically closed subgroup H of a Lie group G is a Lie group, and therefore the quotient $M = G/H$, called a homogeneous space, is a manifold. The map $\pi : G \to G/H$ is a projection. And since $\pi^{-1}(x)$ of any point $x \in M$ is diffeomorphic to H, this is the fibre. The right action $(h, g) \mapsto gh$ satisfies $\pi(gh) = \pi(g)$ by definition.

Any principal G-bundle (P, π, M, G) has fibre bundles $\mathbb{E}_P$ associated to it. For every choice of a left G-space fibre F, the *associated bundle* $\mathbb{E}_P$ with fibre F has the same base space and structure group as the principal bundle. It can be constructed by considering $P \times F$ and the equivalence relation $(pg, f) \sim (p, gf)$ for $p \in P$, $g \in G$, $f \in F$. The total space of $\mathbb{E}_P$ is built by the equivalence classes of $(P \times F)/G$, and the projection in $\mathbb{E}_P$ is $\pi_E\,[(p, f)] = \pi_P(p)$. By this, for example the linear frame bundle is recognized as the principal bundle associated to the tangent bundle.
A specific vector bundle $\mathbb{V}$ associated to a principal bundle P can be constructed from a linear representation $\rho : G \to V$ of the structure group G on a finite-dimensional vector space V. The representation map ρ defines in a canonical way the associated vector bundle $\mathbb{V} = (P \times V)/G$ by the equivalence relation $(p, \rho(g) \cdot v) \simeq (p \cdot g, v)$ for all $p \in P$, $v \in V, g \in G$.

E.5.2 Connections in Fibre Bundles

Connections are needed if we want to describe how geometric objects (tensor fields, say) are transported from one point in the space to another. In Riemann-Cartan geometry, which is the arena of gravitational theories, a connection Γ is introduced in terms of parallel transport, and it entails the notion of a Γ-covariant derivative, of curvature R and torsion T. In Yang-Mills type theories a connection A is understood as a one-form field. It is introduced in order to take care of consistent local transformations of derivatives of matter fields, with an ensuing definition of an A-covariant derivative and of field strengths F associated to the gauge fields A. Indeed, there is an astounding likeness of these notions, and the language of fibre bundles reveals the similarities (and the dissimilarities as well.)

There are various equivalent ways to define connections in fibre bundles. The authors of [146] distinguish the ‘parallel transport’, ‘tangent space’, ‘cotangent space’, ‘axiomatic’, and ‘change of frame’ approaches. Here I sketch the parallel transport approach in which parallel transport is defined ‘vertically’ for the base manifold and extended ‘horizontally’ to a principal bundle, and the cotangent space approach in which a one-form connection becomes introduced that is “fed” by vertical vector fields only. The axiomatic approach is based on the definition of connections as in Appendix E.3.

Parallel Transport in a Principal Bundle

Consider a principal bundle (P, π, M, G) and assume that the base space is a manifold. Since the Lie group is a manifold too, the principal bundle is a manifold itself, and it makes sense to talk about tangent and cotangent spaces at points $p \in P$, and of tangent and cotangent bundles of the principal bundle.

To prepare the notion of parallel transport, we choose a curve C in M and extend this to a curve $\tilde{C}$ in the total space P. A natural way to do so is to define vectors tangent to $\tilde{C}$: Given a vector $X_x \in T_x M$ tangent to C at a point x along the curve, the curve $\tilde{C}$ through the fibre attached to this point is a ‘lift’ of X_x, namely an element of $T_p P$ for some $p \in P$ with $\pi(p) = x$. In order to lift a vector in $T_C M$ to TP one proceeds as follows: Decompose $T_p P$ at every point p into a subspace of vectors tangent to the fibre, called the vertical subspace $ver_p P$, and a complement $hor_p P$, called the horizontal subspace. A choice of a connection is now equivalent to a choice of the horizontal subspace, since the vertical subspace acts along the fibres: $ver_p P = \{X \in T_p P | \pi_* X = 0\}$. Since $\dim P = \dim M + \dim G$ and $\dim ver_p P = \dim G$, the horizontal space has $\dim hor_p P = \dim M$.

In short, the connection on a principal bundle is a $\mathfrak{g}$-valued one form which projects $T_p P$ to $ver_p P \cong \mathfrak{g}$. The full definition is: A *connection* on P is defined as a smooth and unique split of the tangent space $T_p P$ at each point $p \in P$

$$T_p P = ver_p P \oplus hor_p P \qquad X = ver X + hor X$$

where $X \in T_p P$, $ver X \in ver_p P$, $hor X \in hor_p P$, such that the choice of a horizontal subspace at a point p determines all horizontal subspaces at points pg:

$$hor_{pg} P = R_{g^*} hor_p P \qquad \text{for every} \qquad g \in G,$$

where R_{g^*} is push-forward of the right translation in G. This is called the *equivariance condition*.

Parallel transport is then defined as a horizontal lift: If $C : [0, 1] \to M$ is a curve in the base manifold, its horizontal lift $\tilde{C} : [0, 1] \to P$ obeys (1) $\pi(\tilde{C}) = C$, (2) all vectors X_P tangent to $\tilde{C}$ are horizontal: $X_P \in hor_{\tilde{C}} P$.

Given the notion of parallel transport in $\mathbb{P}$ one is able to define parallel transport on any fibre bundle associated to $\mathbb{P}$. Take for example the canonical vector bundle $\mathbb{V}$ with the representation space of the structure group: If $c_P(t)$ is a horizontal lift of a curve $c(t)$ in the base space, the curve $c_V(t)$ is a horizontal lift in $\mathbb{V}$ if $c_V(t) = [(c_P(t), v)]$, where v is a constant element of V.

Connection as a Lie-Algebra Valued One-Form

A connection in a principal bundle can also be characterized by a one-form ω with values in $\mathfrak{g} \cong T_e G$. Explicitly: Define the connection one-form $\omega \in \Gamma(T^*P) \otimes \mathfrak{g}$ on P by the projection of the tangent space $T_p P$ onto the vertical subspace $ver_p P$ satisfying

$$\begin{aligned} \omega(\hat{X}_A) &= A && \text{for every} && A \in \mathfrak{g} \\ R_g^* \omega &= Ad_{(g^{-1})}\omega && \text{for every} && g \in G. \end{aligned}$$

Here $\hat{X}_A$ is the vector field on P induced by the the exponential map $R_g(\exp tA)$, R_g^* is the pull-back of the right-action $R_g : p \mapsto pg$ and $\mathrm{Ad} : A \mapsto hAh^{-1}$ is the automorphism of $\mathfrak{g}$ associated to $h \in G$ by the adjoint representation of G in its algebra. The second relation serves to show that the definition of the connection is equivalent to its definition *via* parallel transport. Namely, for an arbitrary vector $X_p \in T_p P$ it reads $R_g^* \omega_p(X) = \omega_{pg}(R_{g*}X) = g^{-1}\omega(X)g$, and it can be shown that the horizontal subspace defined by the kernel of the map $\omega : T_p P \to \mathfrak{g}$,

$$hor_p P = \{X \in T_p P | \omega(X) = 0\}$$

satisfies the equivariance condition $hor_{pg} P = R_{g^*} hor_p P$.

The explicit form of the connection and its dependency on A can be verified to be:

$$\omega = g^{-1}dg + g^{-1}Ag. \tag{E.25}$$

Covariant Derivative and Curvature

In order to define the curvature on a (principal) bundle we first need a definition of the covariant derivative: Consider a Lie-algebra valued q-form $\alpha \in \Omega_q P \otimes \mathfrak{g} = \Omega(P, \mathfrak{g})$ and decompose it as $\alpha = \alpha^a \otimes X_a$ where α^a is an "ordinary" q-form and the set $\{X_a\}$ constitutes a basis of $\mathfrak{g}$. Furthermore let $X_1, \cdots, X_{q+1} \in T_p P$ be tangent vectors on P. The *exterior covariant derivative* is defined by

$$D\alpha(X_1, \cdots, X_{q+1}) = d_P \alpha \; (hor \; X_1, \cdots, hor \; X_{q+1})$$

where $d_P \alpha = (d_P \alpha^a) \otimes X_a$ is the exterior derivative on P.

The curvature 2-form Ω of a connection ω on the bundle P is defined by

$$\Omega = D\omega \in \Omega_2 P \otimes \mathfrak{g}. \tag{E.26}$$

The curvature 2-form has the property $R_g^* \Omega = g^{-1} \Omega g$ and can be shown to fulfill the Cartan structure equation $\Omega = d_P \omega + \omega \wedge \omega$.

E.5.3 Yang-Mills Gauge Field Theory in Fibre Bundle Language

How do the mathematical bundle notions relate to the physics language used to describe Yang-Mills gauge theories or gravitational theories. In this subsection this will be answered for a Yang-Mills theory.

Indeed the "stage" for gauge theories are principal bundles and their associated vector bundles. The idea is to enlarge the space-time arena M, taken to be for instance 4-dimensional Minkowski space, by attaching to each spacetime point a fibre, each fibre being homoemorphic to the gauge group.

Matter fields are represented by sections in real or complex vector bundles $\mathbb{V}$ associated with the principal bundle $\mathbb{P}$. They are defined by local sections $s : U \to P$ (where $U \subset M$). The section provides a local trivialization of V by fixing D independent basis vectors $e_\alpha : U \to V$. By this the fibre V_x obtains a linear basis $\{e_\alpha\}_x$. The fields $\Psi : U \to V$ are then uniquely represented by $\Psi(x) = \Psi^\alpha(x)\{e_\alpha\}_x$.

What in physics is called "local gauge transformation" corresponds to an automorphism in the principal bundle: Consider mappings $G(p) : P \to P$ which are equivariant, i.e. $G(gp) = gG(p)$ for $g \in G$, and which preserve each fibre, meaning that they act trivially on the base space: $\pi \circ G(p) = \pi$. These maps constitute the group $\mathfrak{G}$ of local gauge transformations. It has group elements depending on spacetime. It is therefore an infinite-dimensional group. From its definition, $\mathfrak{G}$ is a subgroup of the automorphism group $Aut(P)$. It has a subgroup $\mathfrak{G}_o$ (vertical automorphism) with group elements mediating "pure gauge transformations". In general, the group $\mathfrak{G}$ can be taken (locally) to be isomorphic to the semidirect product of the covariance group $Cov(M)$ of the base manifold and the structure group:

$$\mathfrak{G} \cong Cov(M) \ltimes G. \tag{E.27}$$

For an $SU(N)$ Yang-Mills theory in 4-dimensional Minkowski space this reads $\mathfrak{G} \cong ISO(1,3) \ltimes SU(N)$. It is important not to confuse the two groups G and $\mathfrak{G}$. The structure group G constitutes the fibre and is of relevance when it comes to define the connection. The group of local gauge transformations $\mathfrak{G}$ (with the exception of $\mathfrak{G}_0$) mediates via the gauge principle the interaction of matter fields with the gauge fields.

In the language of infinitesimal transformations the Ψ transform as (E.16): $\delta_\gamma \Psi^\alpha = (\hat{\gamma})^\alpha{}_\beta \ \Psi^\beta = \gamma^a (\hat{X}_a)^\alpha{}_\beta \ \Psi^\beta$, where $(\hat{X}_a)^\alpha{}_\beta$ are d-dimensional representation matrices for the generators X_a of the structure group. This is often termed "gauge transformations of the first kind".

The gauge potentials and gauge field strengths are related to the connection one-form and the curvature two-form, respectively, in the principal bundle. The relation becomes explicit as follows: The connection form (E.25) and the curvature form (E.26) are defined globally on the principal bundle. The gauge potential $\mathcal{A}$ and the field strength $\mathcal{F}$ are differential forms in the base manifold. If the principal bundle is non-trivial, a relation between the forms in the principal bundle and its base manifold can only be achieved locally. To understand its meaning in geometric terms let us consider at first the relation between the connection one-form A and the gauge potential $\mathcal{A}$: Take an open subset U of the base space and choose a local section s on U. The *gauge potential* $\mathcal{A}$ is defined as $\mathcal{A} := s^*\omega \in \Gamma(U, T^*M) \otimes \mathfrak{g}$, that is as the pull-back map from P to M.

This shall be made more explicit (following Sect. 7.10 in [379]): Let the local coordinates in P be given by $p = (x, g),\ x \in M,\ g \in G$. Choose

$$\partial^{\alpha\beta} := \frac{\partial}{\partial g_{\alpha\beta}} \in ver_{(x,g)}P, \qquad \tilde{\partial}_\mu := \frac{\partial}{\partial x^\mu} + C_{\mu\alpha\beta}\frac{\partial}{\partial g_{\alpha\beta}} \in hor_{(x,g)}P$$

as a basis in ver_pP and hor_pP, respectively. Then $X = X_{\alpha\beta}\partial^{\alpha\beta} + X^\mu\tilde{\partial}_\mu$ is a generic vector in the tangent space of P. The connection form (E.25) is locally

$$\omega_p = g^{-1}_{\alpha\beta}\, dg_{\alpha\beta} + g^{-1}_{\alpha\beta}\, A^a_\mu\, (T_a)^{\alpha\gamma}\, dx^\mu\, g_{\gamma\beta}$$

where (T_a) are representation matrices for the generators of G. Now

$$\langle \omega_p, X_p\rangle = g^{-1}_{\alpha\beta}\left(X_{\alpha\beta} + X^\mu C_{\mu\alpha\beta}\right) + g^{-1}_{\alpha\beta}\, A^a_\mu\, (T_a)^{\alpha\gamma}\, g_{\gamma\beta}\, X^\mu.$$

For $X_p \in hor_pP$, for which all coefficients $X_{\alpha\beta}$ vanish, we require $\langle \omega_p, hor X_p\rangle = 0$. And since this must hold for all vectors in hor_pP, we get $g^{-1}_{\alpha\beta}\left[C_{\mu\alpha\beta} + A^a_\mu\, (T_a)^{\alpha\gamma}\right] g_{\gamma\beta}$. After multiplication with g we get a relation that fixes the coefficients $C_{\mu\alpha\beta}$ in the basis vectors for hor_pP, that is the coefficients in the derivative $\tilde{\partial}_\mu$:

$$C_{\mu\alpha\beta} = -A^a_\mu\, (T_a)^{\alpha\gamma}\, g_{\gamma\beta}.$$

From this we sense that the derivative $\tilde{\partial}_\mu$ corresponds to what physicists call a covariant derivative in a gauge theory.

Quite in analogy to the gauge potential, the *Yang-Mills field strength* is defined as $\mathcal{F} := s^*\Omega \in \Gamma(U, \Omega_2 B) \otimes \mathfrak{g}$. And indeed from the Cartan structure equation we derive

$$\mathcal{F} = s^*\Omega = s^* d_P\omega + s^*\omega \wedge \omega = d(s^*\omega) + s^*\omega \wedge s^*\omega = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}.$$

(Here use was made of (E.2).) Notice that both the gauge potential and the field strength are defined with respect to a chart U in the base manifold and with respect to a section in this chart. Consider now two intersecting charts $U \cap U' \neq \emptyset$ for which $s' = sg$. One shows that the corresponding gauge potentials are related by $\mathcal{A}' = g^{-1}\mathcal{A}g + g^{-1}dg$, also called "gauge transformations of the second kind". This reveals that gauge freedom corresponds to the freedom to choose local coordinates on a principal bundle. Similarly one finds for the field strengths defined on two charts U_α and U_β with $\mathcal{F}_\alpha = s_\alpha^* \Omega$ on U_α, etc., the compatibility relation $\mathcal{F}_\beta = g_{\alpha\beta}^{-1}\mathcal{F}_\alpha g_{\alpha\beta}$ on $U_\alpha \cap U_\beta$, where $g_{\alpha\beta}$ are the transition functions.

A *pure gauge* is defined by

$$\mathcal{A} = g^{-1}dg \iff \mathcal{F} = 0.$$

which are inner automorphism in the gauge algebra.

Table (E.1), which is adapted from other similar comparisons found in the literature (originating from [568]) summarizes the previous statements[11].

The $U(1)$ principal bundle for electrodynamics and the $SU(N)$ bundles for non-Abelian gauge theories can be trivialized if the base manifold is the full Minkowski space. Thus the bundle approach seems to be nothing but an overload in formalism. But this is no longer true if the base space is non-contractible. An example is the Aharonov-Bohm effect, where in its most simple configuration the base space is a 2-dimensional plane from which a disc is cut out. Another example is electromagnetism with monopoles in which the base space is $\mathbb{R}^2 \times S^2$. Bundle techniques became important also for other topological configurations like instantons in non-Abelian gauge theories. And the non-existence of a global gauge choice in Yang-Mills theories, known as Gribov ambiguity by physicists, relates to the fact that the principal bundle on which gauge choices correspond to sections, is non-trivial; see [292].

Even for globally trivial bundles the fibre bundle approach to Yang-Mills gauge field theories and gravitational theories (if formulated in terms of metric and tetrads) is not simply an overload in formalism, but is a means to isolate the 'physically significant' parts of these theories and distinguish these from their 'merely mathematical' aspects [250].

E.5.4 Metric and Tetrad Gravity on a Bundle

Metric Manifolds

Gravitational theories as compared to Yang-Mills theories have an additional structure, namely a metric. The existence of a metric on a manifold allows to establish additional mappings between different bundles erected over a base space.

---

[11] Notice, that according to this table, we have a quite generic definition of symmetry transformations as "automorphisms in bundles".

Table E.1 Fibre bundle and gauge field entities

Fibre bundle	gauge field theory
principal bundle $\mathbb{P}$	"stage" of gauge field theory
base space M	spacetime
structure group G	gauge group
connection on $\mathbb{P}$	gauge potential
curvature on $\mathbb{P}$	field strength
vector bundle $\mathbb{V}$ associated to $\mathbb{P}$	"stage" of matter fields
section in $\mathbb{V}$	matter field
automorphisms of $\mathbb{P} \in \mathfrak{G}$	local gauge transformations
vertical automorphisms of $\mathbb{P} \in \mathfrak{G}_0$	pure gauge transformations
covariant derivative	dynamical coupling

A *Riemannian metric* g_x is a tensor on $T_xM \otimes T_xM$ which is (i) symmetric: $g_x(X, Y) = g_x(Y, X)$; (ii) positive definite: $g_x(X, X) \geqslant 0$, and $g_x(X, X) = 0$ only for $X = 0$. A metric is *pseudo-Riemannian* if instead of (ii) holds (iii) if $g_x(X, Y) = 0$ for $\forall X \in T_xM \Rightarrow Y = 0$.

We recall that on arbitrary manifolds, an inner product can be defined as a map

$$\langle -, - \rangle : T_x^*M \otimes T_xM \to \mathbb{R} \qquad \langle \omega, X \rangle = \langle \omega_\mu dx^\mu, X^\nu \partial_\nu \rangle = \omega_\mu X^\mu.$$

The metric defines an inner product between two vectors $X, Y \in T_x(M)$ by $g_x(X, Y) : T_xM \otimes T_xM \to \mathbb{R}$. This gives rise to a map

$$g_x(X, -) : T_xM \to \mathbb{R} \qquad Y \mapsto g_x(X, Y).$$

This map can be identified with a one-form: $g_x(X, -) \leftrightarrow \omega$. Since on the other hand, the one-form ω induces a unique vector X by $\langle \omega, Y \rangle = g_x(X, Y)$, the metric makes available an isomorphism

$$T_xM \cong T_x^*M.$$

Locally the metric is represented as $g_x = g_{\mu\nu}(x) dx^\mu \otimes dx^\nu$, and the isomorphism between vectors and forms becomes

$$\omega_\mu = g_{\mu\nu} X^\nu, \quad X^\mu = g^{\mu\nu} \omega_\nu \qquad \text{and} \quad g(X, Y) = \langle \omega, Y \rangle = \omega_\mu X^\mu = g_{\mu\nu} X^\nu Y^\mu.$$

Raising and lowering indices with a metric is of course familiar from Riemannian geometry. But the previous expression makes it clear that a Riemann contravariant vector is the component of a vector in a tangent space, and that the covariant vector is the component of a one-form in a cotangent space. It is only by the metric that one entity can be expressed by the other.

The metric components $g_{\mu\nu}$ are defined with respect to base vectors in the tangent spaces, e.g. in the holonomic basis by $g_{\mu\nu}(x) = g_x(\partial_\mu, \partial_\nu)$, or in an anholonomic basis by $g_{IJ} = g(e_I, e_J)$. (From now on I suppress the reference to the point x, but keep in mind that the subsequent procedure[12] is bound to work for each point separately.) If one switches to another basis, in matrix notation $e' = \Upsilon e$, the metric components change as $g' = \Upsilon^T g \Upsilon$. Now by choosing $\Upsilon = O \cdot D$, where O is an orthogonal matrix ($O^T = O^{-1}$) and D is a diagonal matrix, one can achieve that g' becomes diagonal with either +1 or -1 on the diagonal. By an appropriate choice of the orthonormal matrix one can arrange that all the negative values appear first. This shows that any metric vector space locally has a basis on which the metric tensor has the diagonal form $(\eta) = \mathrm{diag}(-1, \cdot\cdot, -1, 1, \cdot\cdot, 1)$. The basis in which this is true is said to be *orthonormal*.

In the main text (see Subsect. 7.3.) the essentials of Riemann-Cartan geometry are explained, including its description in terms of tetrads. In the language of fibre bundles the

$$e_I = e_I^{\mu} \partial_\mu \qquad \text{with} \qquad \{e_I^{\mu}\} \in GL(D, \mathbb{R})$$

constitute a frame of basis vectors. According to the Gram-Schmidt algorithm these can be made orthogonal: $g_x(e_I, e_J) = e_I^{\mu} e_J^{\nu} g_{\mu\nu}(x) = \eta_{IJ}$. By this, the D-beins e_I^{μ} are only defined up to $SO(D\text{-}1, 1)$ transformations in the case of a Lorentzian metric and $SO(D)$ in the case of a Riemannian metric. If at every point of a base space is attached an orthonormal frame as a fibre, this results in the orthonormal frame bundle $OM \subset LM$. This is a principal bundle with structure group $O(D)$. The tetrads are interpreted as sections in the associated tangent bundle.

Soldering in Gravitational Theories

The geometry of a linear frame bundle LM associated to the tangent bundle of a manifold M is closely tied to the geometry of the base manifold. This can be expressed by means of a one-form θ on LM called the solder form. Let $u \in L_x M$ be a frame at $x \in M$. It was mentioned previously that the map $u : \mathbb{R}^D \to T_x M$ is a linear isomorphism. The solder form of the linear frame bundle is the R^D-valued one-form

$$\theta_u(Y) = u^{-1} \pi_*(Y).$$

Here $Y \in T_{(x,u)} LM$ is a tangent vector to LM at the point (x, u), and π is the projection $LM \to M$. The solder form vanishes on vectors tangent to the fiber and it is equivariant

$$\theta_u(Y^V) = 0 \qquad\qquad R_g^* \theta = g^{-1} \theta$$

where R_g is right translation by $g \in GL(D, \mathbb{R})$. The form θ gives rise to a soldering of the base space and the fibres.

---

[12] Commonly known as Gram-Schmidt algorithm.

The soldering also becomes visible in the possibility to choose locally an orthonormal basis in the tangent and the cotangent spaces. The tetrads act in the same way as the soldering form: They transform the coordinate basis to an orthonormal basis. And there is still another view on soldering: The metric tensor which lives in the base space, can be expressed through the tetrads, which as explained before are sections in a tensor bundle. In Riemann geometry this soldering is even stronger, since the connections (living in the principal bundle) can uniquely be expressed by derivatives of the metric.

GR as a Gauge Theory–a Fibre Bundle Perspective

The two previous subsections revealed that despite there are similarities of Yang-Mills theories and gravitational theories, the fibre bundle language clearly uncovers dissimilarities. I couldn't state this better than A. Trautman [511]:

"For me, a gauge theory is any physical theory of a dynamical variable which, at the classical level, may be identified with a connection on a principal bundle. The structure group G of the bundle P is the group of gauge transformations of the first kind; the group $\mathfrak{G}$ of gauge transformations of the second kind may be identified with a subgroup of the group Aut(P) of all automorphisms of P. In this sense, gravitation is a gauge theory; the basic gauge field is a linear connection ω (or a connection closely related to a linear connection). $\cdots$ The most important difference between gravitation and other gauge theories is due to the soldering of the bundle of frames LM to the base manifold M. The bundle LM is constructed in a natural and unique way from M, whereas a noncontractible M may be the base of inequivalent bundles in the same structure group. $\cdots$ The soldering form θ leads to torsion which has no analog in nongravitational theories. Moreover, it affects the group $\mathfrak{G}$, which now consists of the automorphisms of LM preserving θ. This group contains no vertical automorphisms other than the identity; it is isomorphic to the group Diff(M) of all diffeomorphisms of M. In a gauge theory of the Yang-Mills type over Minkowski space-time, the group $\mathfrak{G}$ is isomorphic to the semidirect product of the Poincaré group by the group $\mathfrak{G}_0$ of vertical automorphisms of P. In other words, in the theory of gravitation, the group $\mathfrak{G}_0$ of 'pure gauge' transformations reduces to the identity; all elements of $\mathfrak{G}$ correspond to diffeomorphisms of M."

I leave this here with only few commentaries, since most terms in this quotation were defined before. The relevant principal bundle in gravitational theories is the orthonormal frame bundle. As the structure group one may for instance take the Lorentz group, but as discussed in Subsect. 7.6.3, other gauge groups might be taken as well. The role of "matter" fields in Yang-Mills theories is in gravitational theories taken on by the tetrad fields. The group $\mathfrak{G}$ of local gauge transformations is due to (E.27) $\mathfrak{G} \cong \mathrm{Diff}(M) \ltimes SO(1,3) \cong \mathrm{Diff}(M)$. These structures demonstrate the benefits of the fibre bundle language if one clearly discriminates the objects involved: There are three different groups: (i) the covariance group of the spacetime manifold $\mathrm{Diff}(M)$, (ii) the structure group G = SO(3,1), (iii) the group of local gauge transformations $\mathfrak{G}$. The diffeomorphism group carries no physical meaning.

Its appearance reflects the option of choosing coordinates in the base manifold. The peculiarity of gravitational theories is that the group $\mathfrak{G}$ is isomorphic to the covariance group $\mathrm{Diff}(M)$. This becomes still more intricate, if for the structure group one would take the Poincaré group instead of the Lorentz group. Then even $G \cong \mathrm{Diff}(M)$ (as can be understood from the fact that local translations correspond to local diffeomorphisms).

These subtleties reflect again the still persistent debate on a "principle" of general covariance and on the interrelation of covariance and invariance (see Subsect. 7.5.3). The insights which can be gained by the fibre bundle view of gravitational theories are discussed in [250].

Appendix F
*Symmetries in Terms of Differential Forms

Today most of the theoretical research articles on symmetries in fundamental physics are written in terms of differential forms and the exterior calculus. Especially in the discussion of symmetries in gravitational actions (and their extension to supergravity), the argumentation in terms of the exterior calculus is preferable, because otherwise one easily gets lost in the tensor notation. In this appendix I describe the structural relations encoded in the "world action" in a very condensed form. More (or rather, all) of the details–especially with respect to Poincaré gauge theories and metric affine gauge theories–may be found in [265].

F.1 Actions and Field Equations

F.1.1 The "World" Action

The classical action, in D dimensions, written as

$$S = \int d^D x \, \mathcal{L}_c(Q^A, Q^A_{,\mu}) \tag{F.1}$$

is a real quantity and a scalar with respect to Lorentz transformations. As described, the Lagrange density $\mathcal{L}_c$ is a function depending on a collection of fields–generically called Q^A –and their first derivatives only. In this appendix, where everything is cast in terms of differential forms, this restriction comes quite naturally. Higher-derivative theories can in principle be expressed in the language of differential forms by the introduction of Lagrange multipliers. As elucidated in the main text, the functional form of $\mathcal{L}_c$ may also be restricted by (dimensional) renormalization requirements, implying restrictions on the dependence of the Lagrangian on the fields and their derivatives.

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

In fundamental physics we are dealing with specific variants of fields, where each variant is characterized by its behavior under Lorentz transformations, gauge group transformations and diffeomorphisms.

In terms of differential forms the previous coordinate-referred action is

$$S = \int_{M^D} \mathcal{L}(Q, dQ). \tag{F.2}$$

Here $\mathcal{L} \in \Omega_D(M, R)$ is a D-form, which locally can be written with the volume form η as:

$$\mathcal{L} = \mathcal{L}_c\ \eta \equiv \mathcal{L}_c\ \vartheta^0 \wedge \vartheta^1 \wedge \ldots \wedge \vartheta^{D-1},$$

where the ϑ^I constitute a basis in T^*M. This shows the (local) relation between the coordinate action (F.1) and the form action (F.2). Observe that in (F.2) there is no explicit reference to coordinates, thus this formulation of the action is by shape and construction invariant with respect to general coordinate transformations–so long as $\mathcal{L}$ transforms as a Riemannian scalar.

In the following, we deal with the fields $Q = \{\vartheta, \omega, A, \psi, \phi\}$, that is, tetrads ϑ (1-forms), spin-connections ω (1-forms), Yang-Mills gauge fields A (1-forms), fermion fields ψ (0-forms), and Higgs scalars ϕ (0-forms). According to present knowledge, the "world action" splits into a (G)ravitational and a (M)atter part[1]:

$$S[\vartheta, \omega, A, \psi, \phi] = S_G[\vartheta, \omega] + S_M[\vartheta, \omega, A, \psi, \phi]. \tag{F.3}$$

We may assume that because of diffeomorphism invariance the gravitational part of the Lagrangian contains the dependence on the derivatives of the tetrads and on the spin-connection only via the torsion Θ and the curvature Ω (both being 2-forms)[2]

$$\Theta^I = d\vartheta^I + \omega^I{}_J \wedge \vartheta^J = \frac{1}{2} T^I_{KJ} \vartheta^{KJ} \tag{F.4a}$$

$$\Omega^I{}_J = d\omega^I{}_J + \omega^I{}_K \wedge \omega^K_J = \frac{1}{2} R^I_{JKL} \vartheta^{KL}. \tag{F.4b}$$

Therefore,

$$\mathcal{L}_G(\vartheta, d\vartheta, \omega, d\omega) = \tilde{\mathcal{L}}_G\big(\vartheta, \Theta(\vartheta, d\vartheta, \omega), \Omega(\omega, d\omega), \lambda\big). \tag{F.5}$$

---

[1] Although perhaps anchored in our minds by a scientific education, this might not be true–see for instance the equivalence of Brans-Dicke theory with f(R)-gravity (Sect. 7.6.4): Brans-Dicke theory contains a scalar field. Another example is the CDJ form of GR with its scalar field η; see Sect. 7.5.2 "Optional forms of actions for GR".

[2] Similar to the case of local phase invariance (see Sect. 5.3.4) we might prove this assumption by Noether's second theorem, or the Klein-Noether identities.

Here, the extra variable(s) λ are introduced as non-dynamical multiplier fields in the case of any constraints that might be imposed on the theory. Take for instance the goal to preclude torsion right from the beginning. Then $\tilde{\mathcal{L}}_G = \tilde{\tilde{\mathcal{L}}}_G(\vartheta, \Theta, \Omega) + \lambda_I \Theta^I$, so that the Euler-Lagrange equations for λ_I result in $\Theta^I = 0$.

We also assume minimal coupling of matter to gravity, that is a matter Lagrangian in which the dependence on the spin connection stems solely from replacing the derivative of the "matter" fields $\Phi = \{A, \psi, \phi\}$ by the covariant derivatives:

$$\mathcal{L}_M(\vartheta, \omega, \Phi, d\Phi) = \tilde{\mathcal{L}}_M(\vartheta, \Phi, D\Phi) \tag{F.6}$$

where the covariant derivative is defined as

$$D\Phi^\alpha = d\Phi^\alpha + \omega^I{}_J \wedge (\hat{X}_I{}^J)^\alpha{}_\beta \Phi^\beta,$$

with the matrices $\hat{X}$ originating from the representation of the Lorentz group to which the fields belong. In case of non-minimal coupling the matter Lagrangian can also depend on $d\vartheta$ and $d\omega$, but in order to be diffeomorphism-invariant, this dependence comes only via the torsion and the curvature[3].

General Relativity

Most of the following is universal in the sense that it makes no difference which gravitational Lagrangian is selected. The Lagrangian (F.5) covers for instance GR, the Einstein-Cartan theory, and Poincaré gauge theories in their most general form.

Taking as an example the gravitational part of general relativity with its Hilbert-Einstein action (including a cosmological term Λ) we find

$$S_{HE} = \frac{1}{2\kappa}\int d^4x\sqrt{-g}(R-2\Lambda) = \frac{1}{2\kappa}\int_M (R-2\Lambda)\,\eta = \frac{1}{2\kappa}\int *(R-2\Lambda), \tag{F.7}$$

where R is the Riemann curvature scalar with $*R = \Omega^{IJ} \wedge \eta_{IJ}$ because of

$$\begin{aligned}\Omega^{IJ} \wedge \eta_{IJ} &= \frac{1}{2} R^{IJ}{}_{KL}\vartheta^K \wedge \vartheta^L \wedge \eta_{IJ} \\ &= \frac{1}{2} R^{IJ}{}_{KL}\vartheta^K \wedge (\delta^L_J \eta_I - \delta^L_J \eta_I) = -R^{IJ}{}_{KI}\vartheta^K \wedge \eta_J = -R^{IJ}{}_{JI}\eta = R\eta.\end{aligned}$$

Maxwell and Yang-Mills

In order to see the relation between the component and the form notations, we consider Maxwell's theory in Minkowski space. The vector potential may be expressed

[3] A dependence of $\mathcal{L}_M$ on ω can be shown to be ruled out by requiring Lorentz invariance [325].

as a one-form

$$A = A_\mu dx^\mu = A_I \vartheta^I$$

and the (two-form) field strength is

$$F := dA = A_{\mu,\nu}\, dx^\nu \wedge dx^\mu = \frac{1}{2}(A_{\mu,\nu} - A_{\nu,\mu}) dx^\nu \wedge dx^\mu = \frac{1}{2} F_{\nu\mu}\, dx^\nu \wedge dx^\mu.$$

We immediately recover from $dF = d(dA) \equiv 0$ the homogeneous Maxwell Eqs. (3.1c, 3.1d). Furthermore, the gauge invariance of the field strength is more or less trivial: with the choice of a different vector potential $A^\Lambda = A + d\Lambda$, the field strength is unchanged: $F = F^\Lambda$. A gauge choice amounts to impose a condition on Λ. Take for instance the Lorenz gauge condition

$$\triangle\Lambda = -d^\dagger A \tag{F.8}$$

where $\triangle = dd^\dagger + d^\dagger d$ is the Laplace-Beltrami operator and $d^\dagger$ the coderivative (E.9). Then $d^\dagger A^\Lambda = d^\dagger A + d^\dagger d\Lambda = d^\dagger A + \triangle\Lambda = 0$, revealing that in this gauge the gauge potential has to fulfill $d^\dagger A = 0$. You may convince yourself that in four-dimensional Minkowski space $d^\dagger A = -\partial_\mu A^\mu$.

The coordinate version of the ED Lagrangian is

$$\mathcal{L}_c = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + A_\mu J^\mu.$$

In terms of a D-form $\mathcal{L} = \mathcal{L}_c \eta$, we expect a kinetic term proportional to the "square" of the field strength. Since the field strength is itself a 2-form, there is only the $(D\text{-}2)$-form $*F$ available to build the "square":

$$\begin{aligned} F \wedge *F &= (\frac{1}{2} F_{IJ} \vartheta^I \wedge \vartheta^J) \wedge *(\frac{1}{2} F_{KL} \vartheta^K \wedge \vartheta^L) = \frac{1}{4} F_{IJ} F^{KL}\, \vartheta^{IJ} \wedge \eta_{KL} \\ &= \frac{1}{4} F_{IJ} F^{KL}\, \vartheta^I \wedge 2\delta^J_{[L} \eta_{K]} = \frac{1}{4} F_{IJ} F^{KL}\, \vartheta^I \wedge (\delta^J_L \eta_K - \delta^J_K \eta_L) \\ &= \frac{1}{2} F_{IJ} F^{IJ}\, \eta. \end{aligned}$$

Thus the action for electrodynamics in terms of forms is

$$S_{ED}[\vartheta, A] = \int_M (-\frac{1}{2} F \wedge *F + A \wedge *J). \tag{F.9}$$

This is immediately extended to free Yang-Mills theory, since extra group indices on the gauge fields are simply add-ons and do not interfere with the form type: If $\mathbf{G}$ is the gauge group, the gauge fields A are $\mathfrak{g}$-valued one-forms: $A \in \mathfrak{g} \otimes \Omega_1(M)$. According to (E.20) the field strength in the non-Abelian theory is the two-form

$F = dA + \frac{1}{2}A \wedge A$. The Yang-Mills action is

$$S_{YM}[\vartheta, A] = \int_M (-\frac{1}{2}\mathrm{tr}(F \wedge *F) + A \wedge *J) \tag{F.10}$$

where tr is the trace on the representation of $\mathfrak{g}$ with the conventional normalization condition $\mathrm{tr}(T^a T^b) = \frac{1}{2}\delta^{ab}$. Furthermore, it is to be understood that also J is a Lie algebra valued one form, e.g. $J = J^a_\mu T_a dx^\mu$.

Scalars

Let us next examine the field theory for a neutral scalar field: The field ϕ is a zero-form. A generic Lagrangian is

$$\mathcal{L} = \frac{1}{2}(d\phi \wedge *d\phi) - V(\phi).$$

The potential can for instance be of the $\lambda\phi^4$ type:

$$V(\phi) = \frac{1}{2}m^2(\phi \wedge *\phi) + \frac{\lambda}{4!}(\phi \wedge *\phi) \wedge *(\phi \wedge *\phi).$$

The interaction of the scalar field with a Yang-Mills field is described by minimal coupling, that is by the replacement

$$d\phi \;\to \overset{A}{D}\phi = d\phi + A \wedge \phi,$$

(here a coupling constant g is absorbed into the A-field) where $\overset{A}{D}$ is the covariant derivative with respect to the gauge group connection A. The precise meaning of this modified exterior derivative will be defined below in F.4. Therefore

$$S_H[\vartheta, A, \phi] = \int_M \frac{1}{2}(\overset{A}{D}\phi \wedge * \overset{A}{D}\phi) - V(\phi). \tag{F.11}$$

Fermions

As exhibited in the main text, fermion fields ψ transform with the spinor representation of the Lorentz group. They additionally may have values in the representation of the gauge group. Thus a covariant derivative is

$$D_\mu\psi = \partial_\mu\psi + \omega^I{}_{J\mu}L^J{}_I\psi + A^a_\mu T^a\psi \qquad \leftrightarrow \qquad D\psi = d\psi + \omega \wedge \psi + A \wedge \psi,$$

where $L^J{}_I$ are the generators of the Lorentz group and T^a the generators of the gauge group, both taken in the representations to which ψ belongs. Defining $\not{D}\Psi = \gamma^I e^\mu_I D_\mu \psi$ the fermion sector of the theory has the action

$$S_F[\vartheta, \omega, A, \phi, \psi] = \int_{M^D} d^4x \; |e|(\bar{\psi}\not{D}\psi + Y(\phi, \bar{\psi}, \psi)). \tag{F.12}$$

Here spinors are considered to be "ordinary" zero-forms. In the mathematical literature you can find approaches to understanding spinors as Clifford-valued differential forms; see e.g. [444].

F.1.2 Field Equations

Field Equations in Terms of Forms

The field equations are found by the principle of stationary action: The generic variation of the Lagrangian in (F.2) is

$$\delta\mathcal{L} = \delta Q \wedge \frac{\partial\mathcal{L}}{\partial Q} + \delta(dQ) \wedge \frac{\partial\mathcal{L}}{\partial(dQ)}$$

(here all contingent indices on the fields Q are suppressed.) The variation δ is meant to be the "active" variation; thus it commutes with the exterior derivative: $(\delta d = d\delta)$. Reshuffling derivatives, $\delta\mathcal{L}$ can be written as

$$\delta\mathcal{L} = \delta Q \wedge \mathcal{E} + d\Theta(Q, \delta Q) \tag{F.13}$$

with the Euler derivatives

$$\mathcal{E} := \frac{\delta\mathcal{L}}{\delta Q} = \frac{\partial\mathcal{L}}{\partial Q} - (-1)^q d\frac{\partial\mathcal{L}}{\partial(dQ)} \tag{F.14}$$

(where q is the form degree of Q) and the exact form

$$\Theta = \delta Q \wedge \frac{\partial\mathcal{L}}{\partial(dQ)},$$

in other contexts also called the *symplectic potential*; see Appendix F.3.1. If we introduce the field momenta

$$P := \frac{\partial\mathcal{L}}{\partial(dQ)} \tag{F.15}$$

which are $(D\text{-}q\text{-}1)$-forms, the Euler derivatives and the symplectic potentials become

$$\mathcal{E} = \frac{\partial \mathcal{L}}{\partial Q} - (-1)^q dP \qquad \Theta = \delta Q \wedge P.$$

The symplectic potential depends (locally) on the dynamical fields and linearly on their variations. "Normally", this term is thrown away as being "nothing but" a boundary term. But we are aware that this can be somewhat premature, since even for a compact spacetime M with boundary ∂M the action $S = \int_M \mathcal{L}$ with

$$\delta S = \int_M \delta \mathcal{L} = \int_M \delta Q \wedge \mathcal{E} + \int_{\partial M} \Theta(Q, \delta Q) \tag{F.16}$$

will, in general, not be extremized by solutions of $\mathcal{E} = 0$. Only if Θ vanishes on ∂M does the requirement $\delta S \stackrel{!}{=} 0$ imply the vanishing of the Euler derivatives:

$$\mathcal{E}_{\{\vartheta, \omega, A, \phi, \psi\}} = 0,$$

and these are the field equations.[4]

We know that an action of a theory is not unique in that we can add surface terms. And it is stressed in the main part of this book that these terms are not at all void of a physical interpretation. Therefore, these will be traced in this appendix in terms of differential forms: Consider instead of the Lagrangian D-form $\mathcal{L}$ the Lagrangian $\mathcal{L}' = \mathcal{L} + d\mathcal{K}$. Then

$$\delta \mathcal{L}' = \delta \mathcal{L} + \delta d\mathcal{K} = [\delta Q \wedge \mathcal{E} + d\Theta] + d\delta\mathcal{K} = \delta Q \wedge \mathcal{E} + d[\Theta + \delta\mathcal{K})].$$

Thus, if the Lagrangian is altered by addition of an exact form, this leaves the Euler derivatives untouched, but causes a shift in the symplectic potential: $\Theta \to \Theta + \delta\mathcal{K}$. Strictly speaking, in general $\mathcal{L}'$ describes a physical system different from the one described by $\mathcal{L}$, because the boundary conditions are changed. In the case that $\mathcal{K}$ depends only on the fields Q and not on its derivatives we have $\delta\mathcal{K} = \delta Q \wedge \frac{\partial \mathcal{K}}{\partial Q}$. Then both Lagrangians lead to the same field equations $\mathcal{E} = 0$ if δQ vanishes on ∂M. In that case, we may call the Lagrangians $\mathcal{L}'$ and $\mathcal{L}$ solution and boundary equivalent. Adding a term $d\mathcal{K}$ to the Lagrangian with $\mathcal{K}$ depending on the derivative of the fields is for instance possible for $\mathcal{K} = -(Q \wedge P)$. Then

$$\delta \mathcal{L}' = \delta Q \wedge \mathcal{E} - d[Q \wedge \delta P].$$

The field equations are unchanged (the Lagrangians are solution-equivalent). However, in this case the field equations only derive from the variation of the Lagrangian

[4] Some authors prefer to call the Hodge dual of these expressions field equations ($*\mathcal{E} = 0$). This can be advantageous since the components of these forms directly compare to the field equations in local frames.

if the boundary condition is $\delta P = 0$; and therefore boundary equivalence of $\mathcal{L}$ and $\mathcal{L}'$ is no longer valid, in general. This is the field-theoretic generalization of the case of finite-dimensional systems, as derived in Sect. 2.1.4; see (2.24). Aside from the possible presence of a $d\mathcal{K}$ term, the symplectic potential Θ is not unique for another reason: One can add to it an exact $(D\text{-}1)$-form, $\Theta \to \Theta + dY(Q, \delta Q)$, without altering (F.13).

On the other hand, these considerations point out how boundary conditions on fields and surface terms in an action functional can be balanced [294]: Suppose that (F.16) does not allow one to infer the field equations $\mathcal{E} = 0$ because the surface term with Θ does not vanish for the boundary conditions on ∂M, hereafter symbolized by X. We seek a modified action S_X such that the field equations result whenever δQ satisfies the conditions X on ∂M. A sufficient condition for S_X to exist is the existence of a (D-1)-form $B(Q)$ and a (D-2)-form $\mu(Q, \delta Q)$ such that

$$\Theta(Q, \delta Q)_{|\partial M} = \delta B(Q) + d\mu(Q, \delta Q)_{|\partial M}$$

for all variations δQ that satisfy X. Then

$$S_X = \int_M \mathcal{L} - \int_{\partial M} B.$$

Gravitational and Matter Field Equations

With the split (F.3), the variation splits into

$$\begin{aligned}\delta\mathcal{L} &= \left[\delta\vartheta \wedge \frac{\delta\mathcal{L}_G}{\delta\vartheta} + \delta\omega \wedge \frac{\delta\mathcal{L}_G}{\delta\omega}\right] + \left[\delta\vartheta \wedge \frac{\delta\mathcal{L}_M}{\delta\vartheta} + \delta\omega \wedge \frac{\delta\mathcal{L}_M}{\delta\omega} + \delta\Phi \wedge \frac{\delta\mathcal{L}_M}{\delta\Phi}\right] + b.t. \\ &= \left[\delta\vartheta^I \wedge E_I + \delta\omega^I{}_J \wedge C_I{}^J\right] + \left[-\,\delta\vartheta^I \wedge T_I - \delta\omega^I{}_J \wedge S_I{}^J + \delta\Phi \wedge \frac{\delta\mathcal{L}_M}{\delta\Phi}\right] + b.t.\end{aligned}$$

with the $(D\text{-}1)$-forms

$$E_I := \frac{\delta\mathcal{L}_G}{\delta\vartheta^I} = \frac{\partial\mathcal{L}_G}{\partial\vartheta^I} + d\frac{\partial\mathcal{L}_G}{\partial(d\vartheta^I)} \tag{F.17a}$$

$$C_I{}^J := \frac{\delta\mathcal{L}_G}{\delta\omega^I{}_J} = \frac{\partial\mathcal{L}_G}{\partial\omega^I{}_J} + d\frac{\partial\mathcal{L}_G}{\partial(d\omega^I{}_J)} \tag{F.17b}$$

also called the *Einstein forms* and the *Cartan forms*, respectively, and the matter energy-momentum and the spin form[5]

[5] I use essentially the notation of [325].

$$T_I := -\frac{\delta \mathcal{L}_M}{\delta \vartheta^I} = -\frac{\partial \mathcal{L}_M}{\partial \vartheta^I} \tag{F.18a}$$

$$S_I{}^J := -\frac{\delta \mathcal{L}_M}{\delta \omega^I{}_J} = -\frac{\partial D\Phi}{\partial \omega^I{}_J} \wedge \frac{\partial \mathcal{L}_M}{\partial D\Phi} = -\hat{X}_I{}^J \Phi \wedge P_\Phi . \tag{F.18b}$$

Therefore the Euler derivatives become

$$\mathcal{E}_\alpha := \frac{\delta \mathcal{L}_M}{\delta \Phi^\alpha}, \qquad \mathcal{E}_I := E_I - T_I, \qquad \mathcal{E}_I{}^J := C_I{}^J - S_I{}^J . \tag{F.19}$$

The matter field equations $\mathcal{E}_\alpha = 0$ are $(D\text{-}q_\alpha)$-forms; the field equations $\mathcal{E}_I = 0$ and $\mathcal{E}_I{}^J = 0$ are $(D\text{-}1)$-forms.

Gravitational Vacuum Field Equations

The gravitational vacuum field equations, that is the equations $E_I = 0$ and $C_I{}^J = 0$, are not manifestly covariant, because of the explicit appearance of the spin connections $\omega^I{}_J$. Therefore, these expressions shall be rewritten in terms of torsion and curvature according to (F.5). Instead of varying the Lagrangian with respect to the tetrads and the spin-connections, one has alternatively

$$\delta \tilde{\mathcal{L}}_G = \delta \vartheta^I \wedge \frac{\partial \tilde{\mathcal{L}}_G}{\partial \vartheta^I} + \delta \Theta^I \wedge \frac{\partial \tilde{\mathcal{L}}_G}{\partial \Theta^I} + \delta \Omega^I{}_J \wedge \frac{\partial \tilde{\mathcal{L}}_G}{\partial \Omega^I{}_J} + \delta \lambda \wedge \frac{\partial \tilde{\mathcal{L}}_G}{\partial \lambda} . \tag{F.20}$$

First note that

$$\frac{\partial \tilde{\mathcal{L}}_G}{\partial \Theta^I} = \frac{\partial \mathcal{L}_G}{\partial d\vartheta^I} =: p_I \qquad \frac{\partial \tilde{\mathcal{L}}_G}{\partial \Omega^I{}_J} = \frac{\partial \mathcal{L}_G}{\partial d\omega^I{}_J} =: \pi_I{}^J , \tag{F.21}$$

where p_I and $\pi_I{}^J$ are the momenta conjugate to ϑ_I and $\omega^I{}_J$. Furthermore,

$$\delta \Theta^I = \delta d\vartheta^I + \delta \omega^I{}_J \wedge \vartheta^J + \omega^I{}_J \wedge \delta \vartheta^J = D\delta \vartheta^I + \delta \omega^I{}_J \wedge \vartheta^J \tag{F.22a}$$

$$\delta \Omega^I{}_J = \delta d\omega^I{}_J + \delta \omega^I{}_K \wedge \omega^K{}_J + \omega^I{}_K \wedge \delta \omega^K{}_J = D\delta \omega^I{}_J . \tag{F.22b}$$

(Incidentally notice that although the derivative d commutes with the variation δ, the covariant derivative D does not.) From this,

$$\begin{aligned} \delta \Theta^I \wedge p_I &= \delta \vartheta^I \wedge Dp_I + \delta \omega^I{}_J \wedge \vartheta^J \wedge p_I + d(\delta \vartheta^I \wedge p_I) \\ \delta \Omega^I{}_J \wedge \pi_I{}^J &= \delta \omega^I{}_K \wedge D\pi_I{}^J + d\big(\delta \omega^I{}_J \wedge \pi_I{}^J\big). \end{aligned}$$

Collecting all the pieces, (F.20) can be written as

$$\delta\widetilde{\mathcal{L}}_G = \delta\vartheta^I \wedge \big(\frac{\partial \tilde{L}_G}{\partial\vartheta^I} + Dp_I\big) + \delta\omega^I{}_J \wedge \big(D\pi_I{}^J + \vartheta^{[J} \wedge p_{I]}\big)$$
$$+ \delta\lambda \wedge \frac{\partial\tilde{\mathcal{L}}_G}{\partial\lambda} + d\big(\delta\vartheta^I \wedge p_I + \delta\omega^I{}_J \wedge \pi_I{}^J\big), \tag{F.23}$$

and therefore the vacuum gravitational-field equations can also be written in the manifestly covariant form

$$E_I = \frac{\partial\tilde{\mathcal{L}}_G}{\partial\vartheta^I} + Dp_I = 0, \qquad C_I{}^J = D\pi_I{}^J + \vartheta^{[J} \wedge p_{I]} = 0 \tag{F.24}$$

together with constraint equations $\frac{\partial\tilde{\mathcal{L}}_G}{\partial\lambda} = 0$. For later purposes, I introduce the abbreviations

$$e_I = \frac{\partial\tilde{\mathcal{L}}_G}{\partial\vartheta^I}, \qquad c_I{}^J = \vartheta^{[J} \wedge p_{I]}. \tag{F.25}$$

As an example, for the Hilbert-Einstein Lagrangian $\tilde{\mathcal{L}}_G \rightarrow \mathcal{L}_{HE} = \frac{1}{2\kappa}\Omega^{IJ} \wedge \eta_{IJ} + \lambda_I \Theta^I$ we have $p_I = 0$, $\pi_{IJ} = \frac{1}{2\kappa}\eta_{IJ}$. The field equations become

$$E_I = \Omega^{KJ} \wedge \frac{\partial}{\partial\vartheta^I}\eta_{KJ} = \Omega^{KJ} \wedge \eta_{KJI} = 0, \qquad C_{IJ} = D\eta_{IJ} = 0$$

together with the equation of vanishing torsion. As a matter of fact, the field equations $D\eta_{IJ} = 0$ again imply, due to $D\eta_{IJ} = \eta_{IJK} \wedge \Theta^K$, that the torsion must vanish. (This is true only in the vacuum theory without spin currents acting as sources.). Regarding the vanishing torsion in the first set of field equations, the components of (the dual of) the $(D\text{-}1)$-form equation $E_I = 0$ are indeed identical to Einstein's vacuum field equations.

Maxwell and Yang-Mills Field Equations

Electrodynamics is defined by (F.9). Varying this action for a fixed source current, we get

$$-\delta F \wedge *F + \delta A \wedge *J = -\delta(dA) \wedge *F + \delta A \wedge *J = -d(\delta A) \wedge *F + \delta A \wedge *J$$
$$= -d(\delta A \wedge *F) - \delta A \wedge \big(d * F - *J\big).$$

Dropping the first term, we get the field equations ($d * F = *J$) which in four-dimensional Minkowski space (in which $** = 1$ and $d^\dagger = *d*$) can be written

$$d^\dagger F = J.$$

Because of $d^\dagger d^\dagger = 0$, we immediately obtain current conservation: $d^\dagger J = 0$.

The equations $d^\dagger F = J$ together with $dF = 0$ comprise the full set of Maxwell's equations. The latter are fulfilled for $F = dA$ (at least locally). In the Lorenz gauge (F.8), for which $d^\dagger A = 0$, we get from $d^\dagger F = d^\dagger dA = \triangle A = J$ the other relation which the gauge potential must fulfill.

Let me also outline how to arrive from the Yang-Mills action (F.10) to the field equations. We now have

$$\begin{aligned}\delta F \wedge *F &= \delta\Big(dA + \frac{1}{2}[A, A]\Big) \wedge *F = \Big(\delta dA + [\delta A, A]\Big) \wedge *F \\ &= d(\delta A \wedge *F) + \delta A \wedge \Big(d * F + [A, *F]\Big)\end{aligned}$$

from which it is readily understood that the field equations are $D{*}F = *J$. Just as for the exterior derivative, a coderivative to $D^\dagger$ can be introduced. In 4dim-Minkowski space this amounts to $D^\dagger = *D*$, and the Yang-Mills field equations read $D^\dagger F = J$. One can show that $D^\dagger D^\dagger F = 0$, and therefore the current is covariantly conserved.

Einstein-Maxwell Field Equations

The Einstein-Maxwell theory is defined by the action

$$S[\vartheta, \omega, A] = S_{HE}[\vartheta, \omega] + S_{MW}[\vartheta, A] = \frac{1}{2\kappa}\int_M \Omega^{IJ} \wedge \eta_{IJ} - \frac{1}{2}\int_M F \wedge *F.$$

In the previous passage, we calculated

$$\delta\mathcal{L}_{HE} = \frac{1}{2\kappa}\delta\vartheta^L \wedge \Omega^{IJ} \wedge \eta_{IJL}.$$

Here the term proportional to $\delta\omega^{IJ}$ is dropped because it is proportional to the torsion. This vanishes in the vacuum theory and in this case also because the Maxwell-Lagrangian does not depend on the spin connection. Now use (E.7) to derive

$$\delta\mathcal{L}_{MW} = -\frac{1}{2}\delta(F \wedge *F) = -\delta F \wedge *F - \frac{1}{2}\delta\vartheta^L \wedge [F \wedge i_L * F - i_L F \wedge *F].$$

The matter-field equations are enclosed in the first term and these were derived previously as $d * F = 0$, which are the Maxwell equations[6] as the components of the 1-form

$$d^\dagger F = 0. \tag{F.26}$$

[6] This highly compact notation reveals again that physics became simpler with the notion of symmetries; compare this with the "original" Maxwell equations written in terms of derivatives of the components of the E- and B-fields with respect to time and to the (x, y, z)- coordinates. And with (F.26), you even have electrodynamics coupled to a gravitational field.

The terms in $\delta\mathcal{L}_{HE} + \delta\mathcal{L}_{MW}$ which are proportional to $\delta\vartheta^L$, constitute the gravitational field equations

$$\frac{1}{2\kappa}\Omega^{IJ} \wedge \eta_{IJL} - \frac{1}{2}[F \wedge i_L * F - i_L F \wedge *F] = 0, \tag{F.27}$$

from which one reads off the energy-momentum current for the electromagnetic field:

$$T_L^{(ED)} = -\frac{1}{2}[F \wedge i_L * F - i_L F \wedge *F].$$

Field Equations Derived by Palatini Variation

In some situations, it turns out to be advantageous to derive the field equations by the Palatini variations[7]. Take for instance the action for flat-space Maxwell vacuum theory

$$S[A, F] = \int_M (\frac{1}{2}F \wedge *F - dA \wedge *F) = \int_M \overline{\mathcal{L}}$$

where A and F are now understood to be independent variables. The Euler-Lagrange equations derive from

$$\begin{aligned}\delta\overline{\mathcal{L}} &= F \wedge \delta * F - \delta dA \wedge *F - dA \wedge \delta * F \\ &= (F - dA)\delta * F - \delta A \wedge d * F - d(\delta A \wedge *F)\end{aligned}$$

as $F = dA$ and $d{*}F = 0$. The latter is the flat-space variant of (F.26), and the former relates the field strength to the connection. Since $\Pi = -\frac{1}{2} * F$ is the momentum canonically conjugate to A, the Lagrangian can also be written as $\overline{\mathcal{L}}[F(A), \Pi] = F \wedge \Pi$.

In the case of gravitational theories, one can verify that the variation of the Lagrangian

$$\overline{\mathcal{L}}_G(\vartheta^K, \omega^{KL}, p_K, \pi_{KL}) = \Theta^I \wedge p_I + \Omega^{IJ} \wedge \pi_{IJ} - \Lambda(p_K, \pi_{KL}), \tag{F.28}$$

explicitly (but suppressing indices)

[7] This is sometimes also called "first-order" variation. In the context of differential forms, this terminology might lead to confusion since we already have a first-order formalism in treating the tetrads and the spin connections as independent objects.

$$\delta\overline{\mathcal{L}}_G(\vartheta, \omega, p, \pi) = \delta\Theta \wedge p \;+\; \delta\Omega \wedge \pi \;+\; \Theta \wedge \delta p \;+\; \Omega \wedge \delta\pi \\ -\; \delta p \wedge \frac{\partial\Lambda}{\partial p} \;-\; \delta\pi \wedge \frac{\partial\Lambda}{\partial\pi}$$

can be rewritten as

$$\delta\overline{\mathcal{L}}_G = \delta\vartheta^I \wedge E_I + \delta\omega^I{}_J \wedge C^I{}_J + (\Theta - \frac{\partial\Lambda}{\partial p}) \wedge \delta p + (\Omega - \frac{\partial\Lambda}{\partial\pi}) \wedge \delta\pi \quad \text{(F.29)}$$

and thus leads to the field Eqs. (F.24) together with

$$\frac{\delta\overline{\mathcal{L}}_G}{\delta p_I} = \Theta^I - \frac{\partial\Lambda}{\partial p_I} = 0, \qquad \frac{\delta\overline{\mathcal{L}}_G}{\delta\pi_{IJ}} = \Omega^{IJ} - \frac{\partial\Lambda}{\partial\pi_{IJ}} = 0.$$

Here we see one of the advantages of the Palatini formalism, in that restrictions on the geometry (like for instance vanishing torsion) can be imposed on the Lagrangian by selecting the potential Λ to be independent of the associated momenta. Specifically GR is defined by

$$\overline{\mathcal{L}}_{GR} = \Theta^I \wedge p_I + \Omega^{IJ} \wedge \pi_{IJ} - \Lambda^{IJ} \wedge \left(\pi_{KL} - \frac{1}{2\kappa}\eta_{KL}\right).$$

The p variation leads to vanishing torsion, the variation of the multipliers Λ enforce that $\pi_{KL} = \frac{1}{2\kappa}\eta_{KL}$ in accordance with previous findings. Furthermore, the variation of π results in $\Omega = V$, and the variation of the spin connection ω leads to vanishing p, again in agreement with previous findings.

The Palatini description allows a more or less straightforward transition to the Hamiltonian; more about this is given in Sect. F.3.1.

F.2 Symmetries

F.2.1 Generic Variational Symmetries

Noether Currents and Superpotentials

In the previous section, we considered arbitrary variations of the action and derived the field equations–provided boundary terms drop out. In a shorthand notation due to (F.13), the generic variation of the Lagrangian is $\delta\mathcal{L} = \delta Q \wedge \mathcal{E} + d\Theta(Q, \delta Q)$. A variational symmetry is present if the Lagrangian is quasi-invariant, that is, if under an infinitesimal transformation $\delta_\epsilon Q$

$$\delta_\epsilon\mathcal{L} = dF_\epsilon. \tag{F.30}$$

This is by virtue of (F.13) equivalent to

$$dF_\epsilon = \delta_\epsilon Q \wedge \mathcal{E} + d\Theta(Q, \delta_\epsilon Q) \qquad \text{or} \qquad \delta_\epsilon Q \wedge \mathcal{E} + dJ_\epsilon \equiv 0, \tag{F.31}$$

with the Noether current (which is a $(D$-1)-form)

$$J_\epsilon := \Theta(Q, \delta_\epsilon Q) - F_\epsilon = \delta_\epsilon Q \wedge P - F_\epsilon. \tag{F.32}$$

If a Lagrangian $\mathcal{L}$ obeys a symmetry condition (F.30), the solution-equivalent Lagrangian $\mathcal{L}' = \mathcal{L} + d\mathcal{K}$ obeys $\delta_\epsilon \mathcal{L}' = d(F_\epsilon + \delta_\epsilon \mathcal{K})$, showing that the theory described by $\mathcal{L}'$ is also quasi-invariant. The ambiguity in Θ gives rise to an ambiguity in the Noether current as

$$J_\epsilon \to J_\epsilon + \delta_\epsilon \mathcal{K} + dY(Q, \delta_\epsilon Q).$$

On solutions (denoted as $\mathcal{E} \doteq 0$), for every symmetry transformation characterized by the parameter ϵ, the Noether current is conserved:

$$dJ_\epsilon \doteq 0.$$

Therefore, on-shell, every Noether current can be derived from a *superpotential* $(D$-2)-form U_ϵ, so that

$$J_\epsilon \doteq dU_\epsilon.$$

The superpotential is not unique, since it can be replaced by $U_\epsilon + dV_\epsilon$ (and it also inherits the ambiguities from the currents).

Klein-Noether Identities

Assuming that the symmetry variation can be written in the form[8]

$$\delta_\epsilon Q = \epsilon \mathcal{A} + d\epsilon \wedge \mathcal{B}, \qquad F_\epsilon = \epsilon\sigma + d\epsilon \wedge \rho, \tag{F.33}$$

we also expand the Noether current as

$$J_\epsilon = \epsilon j + d\epsilon \wedge k = \epsilon(j - dk) + d(\epsilon k).$$

From (F.32) and (F.33), one finds the current "components"

$$j = \mathcal{A} \wedge P - \sigma \qquad\qquad k = \mathcal{B} \wedge P - \rho. \tag{F.34}$$

[8] This assumption is justified in the case of all fundamental interactions.

Now inserting (F.33) into the identity (F.31), one obtains

$$\begin{aligned}0 &\equiv (\epsilon\mathcal{A} + d\epsilon \wedge \mathcal{B}) \wedge \mathcal{E} + d(\epsilon(j - dk))\\ &= \epsilon(\mathcal{A} \wedge \mathcal{E} + dj) + d\epsilon \wedge [\mathcal{B} \wedge \mathcal{E} + j - dk]\end{aligned}$$

leading to the Klein-Noether identities

$$\mathcal{A} \wedge \mathcal{E} + dj \equiv 0 \tag{F.35a}$$

$$\mathcal{B} \wedge \mathcal{E} + j - dk \equiv 0. \tag{F.35b}$$

In the language of differential forms, we obtain solely these two sets of identities. A third one, corresponding to (3.72a) in the component notation, is "automatically" present, since dk is a closed form. Observe that the second identity is present only for local symmetries ($d\epsilon \neq 0$). In any case, the current j is on-shell conserved:

$$dj = -\mathcal{A} \wedge \mathcal{E} \doteq 0.$$

Since aside from this expression for dj we also have, according to (F.34),

$$dj = d(\mathcal{A} \wedge P) - d\sigma$$

we can equate these expressions to arrive at

$$\begin{aligned}\mathcal{A} \wedge \frac{\partial \mathcal{L}}{\partial Q} - (-1)^q \mathcal{A} \wedge dP + d\mathcal{A} \wedge P + (-1)^q \mathcal{A} \wedge dP - d\sigma&\\ = \mathcal{A} \wedge \frac{\partial \mathcal{L}}{\partial Q} + d\mathcal{A} \wedge P - d\sigma \equiv 0.&\end{aligned}$$

In the case of local symmetries we also realize that on-shell

$$j \doteq dk.$$

From (F.35), the Noether identities result immediately:

$$\mathcal{A} \wedge \mathcal{E} - d(\mathcal{B} \wedge \mathcal{E}) \equiv 0. \tag{F.36}$$

Inserting the second Klein-Noether identity into (F.34), we find from

$$J_\epsilon \equiv -\epsilon(\mathcal{B} \wedge \mathcal{E}) + d(\epsilon k), \tag{F.37}$$

that the Noether current itself can be written as a linear combination of field equations and the total derivative of a superpotential $U_\epsilon = \epsilon k$. This was first discovered for diffeomorphism symmetries in GR by P. Bergmann and coworkers around 1950; but as we have just derived, it holds for any local symmetry. Observe that according to (F.37), for the on-shell conservation of the Noether current, it suffices that the field equations are fulfilled for those fields which transform with derivatives of the symmetry parameter ϵ, that is for which $\mathcal{B} \neq 0$ in (F.33). Also keep in mind that the (D-1)-form F_ϵ is not unique since it changes with surface terms in the action. Therefore, also the current components j and k are not unique. It is true that for a given action one can calculate the F_ϵ. For Yang-Mills theory $\delta_\theta \mathcal{L} = 0$ (and thus $F_\theta = 0$), and for general relativity (and all diffeomorphism covariant theories) $\delta_\xi \mathcal{L} = \pounds_\xi \mathcal{L}$, and thus $F_\xi = i_\xi \mathcal{L}$.

Noether Charges

We identified two currents that are conserved on-shell, namely the "full" Noether currents J_ϵ and its "component" j. In taking the latter, define quantities

$$q = \int_S j = \int_{\partial S} k$$

by integrating the (D-1)-form j over a (D-1)-dimensional spatial hypersurface S. On shell, this current component can be derived from a superpotential (D-2)-form as $j = dk$. Therefore the quantity q can be expressed by an integral over the spatial boundary ∂S. But one will find that in general these q do not transform as scalars w.r.t. the symmetries unless one imposes restrictions on the allowed symmetry transformations on the boundary ∂S. In YM-theories the quantities q are gauge dependent and in GR they are coordinate dependent. Thus they do not qualify as conserved charges. Instead take the full Noether current and define for each parameter set ϵ an associated conserved charge by

$$C[\epsilon] = \int_S J_\epsilon = \int_{\partial S} U_\epsilon. \tag{F.38}$$

This seems to open a treasure chest of charges. Which ones can be given a meaning depends on the theory in question. Specifically for generally covariant theories one is interested to identify those charges which correspond to conserved energy and momentum.

In fundamental physics we have two kinds of invariances, either "geometric" ones, associated to coordinate symmetries, or "internal" symmetries, associated to gauge symmetries. The coordinate symmetries we are interested in are the local Lorentz transformations (related to the use of tetrads) and general coordinate transformations. These will be dealt with in the next two subsections.

F.2.2 Lorentz Transformations

Since the action is a Lorentz scalar, the theory is quasi-invariant with respect to Lorentz transformations

$$\delta_L(\lambda)\mathcal{L} = dF_\lambda.$$

In the following, the case of an invariant Lagrangian, that is $dF_\lambda = 0$, is assumed. A generalization to the case of quasi-invariance is possible without difficulty. It is needed for instance in teleparallel theories, see [397]. An infinitesimal Lorentz transformation is parameterized as $\Lambda^I{}_J = \delta^I_J + \lambda^I{}_J$ with $\lambda^{IJ} + \lambda^{JI} = 0$. The λ^{IJ} are coefficients of vector fields λ in the Lie-algebra associated to the Lorentz group: $\lambda = \lambda^{IJ} \otimes X^{IJ}$, where the generators X^{IJ} constitute a basis in the algebra; for more details about these mathematical structures see Appendix F.4. below. Under the Lorentz transformations, the fields transform as

$$\begin{aligned} \delta_L(\lambda)\Phi^\alpha &= \lambda^\alpha{}_\beta \\ \Phi^\beta, \delta_L(\lambda)\vartheta^I &= \lambda^I{}_J\, \vartheta^J, \\ \delta_L(\lambda)\omega^I{}_J &= -D\lambda^I{}_J = -d\lambda^I{}_J - \omega^I{}_K\lambda^K{}_J + \lambda^I{}_K\omega^K{}_J. \end{aligned} \tag{F.39}$$

Here $\lambda^\alpha{}_\beta = (\lambda^I{}_J \hat{X}_I{}^J)^\alpha{}_\beta$ is the Lorentz generator for the matter fields (i.e. matrices of the representation to which Φ^α belongs). The group composition (near the identity) is characterized by the commutator of two infinitesimal Lorentz transformations:

$$[\delta_L(\lambda), \delta_L(\lambda')] = \delta_L([\lambda, \lambda']), \tag{F.40}$$

where the commutator of the two parameters is taken in the algebra $\mathfrak{so}(3, 1)$.

Currents for Local Lorentz Symmetry

The current corresponding to (F.32) is

$$\begin{aligned} J_\lambda = \delta_\lambda Q \wedge P &= \lambda^\alpha{}_\beta\, \Phi^\beta \wedge P_\alpha + \lambda^I{}_J\, \vartheta^J \wedge p_I - D\lambda^I{}_J \wedge \pi_I{}^J \\ &= -\lambda^I{}_J S_I{}^J + \lambda^I{}_J c_I{}^J - d(\lambda^I{}_J \wedge \pi_I{}^J) + \lambda^I{}_J \wedge D\pi_I{}^J \end{aligned}$$

with the definition of the spin momentum S from (F.18b) and the abbreviation for c from (F.25). Then, because of (F.24) and (F.19), this can finally be cast in the form

$$J_\lambda = \lambda^I{}_J \mathcal{E}_I{}^J - d(\lambda^I{}_J \wedge \pi_I{}^J). \tag{F.41}$$

This is of course compatible with the general finding (F.37) and with the variation of the fields corresponding to (F.33). The superpotential is $U_\lambda = -\lambda^I{}_J \wedge \pi_I{}^J$. The Noether currents are conserved on the solutions of the field equations for the spin connections.

Klein-Noether Identities for Local Lorentz Symmetry

From (F.13), we find directly $\delta_L(\lambda)\mathcal{L} = \delta_L(\lambda)Q \wedge \mathcal{E} + d\big(\delta_L(\lambda)Q \wedge P\big) = 0$, and thus the identity

$$\begin{aligned} &\lambda^\alpha{}_\beta\, \Phi^\beta \wedge \mathcal{E}_\alpha + \lambda^I{}_J \vartheta^J \wedge \mathcal{E}_I - D\lambda^I{}_J \wedge \mathcal{E}_I{}^J \\ &\qquad + d\big(\lambda^\alpha{}_\beta\, \Phi^\beta \wedge P_\alpha + \lambda^I{}_J \vartheta^J \wedge p_I - D\lambda^I{}_J \wedge \pi_I{}^J\big) \equiv 0. \end{aligned} \tag{F.42}$$

In order to go from this expression to the Klein-Noether identities, we shift the covariant derivative away from the parameter λ by making use of

$$D\lambda^I{}_J \wedge Z_I{}^J = d(\lambda^I{}_J \wedge Z_I{}^J) - \lambda^I{}_J \wedge DZ_I{}^J.$$

This is valid because $\lambda^I{}_J \wedge Z_I{}^J$ is a Lorentz scalar for arbitrary $Z_I{}^J$. Then (F.42) can be written as

$$\lambda^I{}_J\, A_I{}^J + d(\lambda^I{}_J\, B_I{}^J) \equiv 0, \tag{F.43}$$

with

$$A_I{}^J = (\hat{X}_I{}^J)^\alpha{}_\beta\, \Phi^\beta \wedge \mathcal{E}_\alpha + \vartheta^{[J} \wedge \mathcal{E}_{I]} + D\mathcal{E}_I{}^J \equiv 0, \tag{F.44a}$$

$$B_I{}^J = (\hat{X}_I{}^J)^\alpha{}_\beta\, \Phi^\beta \wedge P_\alpha + \vartheta^{[J} \wedge p_{I]} + D\pi_I{}^J - \mathcal{E}_I{}^J \equiv 0. \tag{F.44b}$$

The first line is the expected result: These are the Noether identities originating from local Lorentz invariance. And since by the definition of the spin form (F.18b) and the Cartan form from (F.24), $B_I{}^J = -S_I{}^J + C_I{}^J - \mathcal{E}_I{}^J$, the second set of identities (F.44b) yields the field Eqs. (F.19) with respect to the spin connections. There is additional information hidden in the consistency condition following from taking the covariant derivative of (F.44b) and inserting this into the Noether identities (F.44a):

$$D(S_I{}^J - C_I{}^J) \equiv -\vartheta^{[J} \wedge (T - E)_{I]} - (\hat{X}_I{}^J)^\alpha{}_\beta\, \Phi^\beta \wedge \mathcal{E}_\alpha.$$

In case of vacuum gravity, this amounts to a relation between the Cartan and the Einstein form,

$$DC_I{}^J + \vartheta^{[J} \wedge E_{I]} \equiv 0 \tag{F.45}$$

and for a matter-field in a gravitational background field to

$$DS_I{}^J + \vartheta^{[J} \wedge T_{I]} \equiv -(\hat{X}_I{}^J)^\alpha{}_\beta \, \Phi^\beta \wedge \mathcal{E}_\alpha. \tag{F.46}$$

The superpotential becomes $U_\lambda = -\lambda^I{}_J \, \pi_I{}^J$, so that charges

$$C\left[\lambda^I{}_J\right] = -\int_{\partial S} \lambda^I{}_J \pi_I{}^J = -\frac{1}{2\kappa}\int_{\partial S} \lambda^I{}_J \, \eta_I{}^J \tag{F.47}$$

result, where the last expression on the right-hand side are the charges from the Einstein-Cartan theory.

F.2.3 Diffeomorphisms

General coordinate transformations are described infinitesimally

$$\delta_D(\xi) = \pounds_\xi, \qquad \text{where} \qquad \xi = \xi^\mu \partial_\mu = \xi^I e_I$$

is a vector field[9]. They form a group whose algebra near the identity transformation is encoded in the commutator

$$\begin{aligned}[][\delta_D(\xi), \delta_D(\xi')] &= [\pounds_\xi, \pounds_{\xi'}] = [di_\xi, \pounds_{\xi'}] + [i_\xi d, \pounds_{\xi'}] \\ &= d[i_\xi, \pounds_{\xi'}] + [i_\xi, \pounds_{\xi'}]d = di_{[\xi,\xi']} + i_{[\xi,\xi']}d = \pounds_{[\xi,\xi']},\end{aligned}$$

that is

$$[\delta_D(\xi), \delta_D(\xi')] = \delta_D([\xi, \xi']). \tag{F.48}$$

For the commutator of a general coordinate transformation and a Lorentz transformation, one finds

$$[\delta_D(\xi), \delta_L(\lambda)] = \delta_L(\pounds_\xi \, d\lambda) = \delta_L(i_\xi \, d\lambda). \tag{F.49}$$

If the Lagrangian is a weight-one scalar, it transforms as

$$\delta_\xi \mathcal{L} = \pounds_\xi \mathcal{L} = i_\xi d\mathcal{L} + di_\xi \mathcal{L} = di_\xi \mathcal{L},$$

since $d\mathcal{L} = 0$, $\mathcal{L}$ being a D-form. Thus we identify $F_\xi = i_\xi \mathcal{L}$ in (F.30).

---

9 This corresponds to a local transformation $x^\mu \to x^\mu - \xi^\mu$. That is, a minus sign is chosen in this appendix, in order to avoid an abundance of minus signs in subsequent expressions. This differs from the convention used in the main text.

Diffeomorphism Currents

The current corresponding to (F.32) can be rewritten as

$$\begin{aligned} J_\xi &= \delta_\xi Q \wedge P - i_\xi \mathcal{L} = \pounds_\xi Q \wedge P - i_\xi Q \wedge \frac{\partial \mathcal{L}}{\partial Q} - i_\xi dQ \wedge \frac{\partial \mathcal{L}}{\partial dQ} \\ &= d i_\xi Q \wedge P - i_\xi Q \wedge \frac{\partial \mathcal{L}}{\partial Q} = -i_\xi Q \wedge \mathcal{E} + d(i_\xi Q \wedge P), \end{aligned}$$

or explicitly as

$$J_\xi = -\xi^I \mathcal{E}_I - i_\xi \omega^I{}_J \wedge \mathcal{E}_I{}^J - i_\xi \Phi^\alpha \wedge \mathcal{E}_\alpha + dU_\xi. \tag{F.50}$$

Again, the Noether current can be represented as a sum of a terms that vanish on-shell and a superpotential $U_\xi = i_\xi Q \wedge P$, such that $dJ_\xi \doteq 0$.

From (F.31), we immediately see that in this case of diffeomorphisms

$$dJ_\xi = -\delta_\xi Q \wedge \mathcal{E} = -\pounds_\xi Q \wedge \mathcal{E}.$$

As usual, this vanishes on-shell. But it also vanishes for vacuum gravity if $\pounds_\xi\{\vartheta, \omega\} = 0$, that is in the case of Killing symmetries.

The expression for the current (F.50) in the case of vacuum gravity and in the case of matter in a fixed gravitational background becomes:

$$\begin{aligned} J_\xi^v &= -\xi^I E_I - i_\xi \omega^I{}_J \wedge C_I{}^J + dU_\xi^v \\ J_\xi^M &= \xi^I T_I + i_\xi \omega^I{}_J \wedge S_I{}^J - i_\xi \Phi^\alpha \wedge \mathcal{E}_\alpha + dU_\xi^M. \end{aligned}$$

If the spin current vanishes, we have in particular

$$J_I^M = T_I - i_I \Phi^\alpha \wedge \mathcal{E}_\alpha + d(i_I \Phi \wedge P_\Phi). \tag{F.51}$$

Thus, on-shell ($\mathcal{E}_\alpha \doteq 0$), the Noether current J_I^M is identical to the Hilbert matter energy-momentum current T_I plus an exact form.

Gravitational Klein-Noether Identities for Diffeomorphism Symmetry

Covariant Lie Derivatives

Let us first consider a pure gravitational theory in order to motivate the introduction of a covariant Lie derivative. Starting from (F.20) and using the definitions of the momenta according to (F.21) and the abbreviations from (F.25), the condition for diffeomorphism invariance specializes to

$$\pounds_\xi \vartheta^I \wedge e_I + \pounds_\xi \Theta^I \wedge p_I + \pounds_\xi \Omega^I{}_J \wedge \pi_I{}^J - \pounds_\xi \tilde{\mathcal{L}}_G \equiv 0. \tag{F.52}$$

This does not look very revealing; it even is not manifestly Lorentz invariant. In view of an improvement, introduce a covariant derivative with respect to an optional Lorentz connection $\Gamma = \Gamma^I{}_J \Lambda^I{}_J$

$$\overset{\Gamma}{D} = d + \Gamma$$

together with a covariant Lie derivative referring to this connection:

$$\overset{\Gamma}{\pounds}_X \mathcal{F} = (i_X \overset{\Gamma}{D} + \overset{\Gamma}{D} i_X)\mathcal{F} = (i_X(d+\Gamma) + (d+\Gamma)i_X)\mathcal{F} = \pounds_X \mathcal{F} + i_X(\Gamma\mathcal{F}).$$

General properties of these covariant derivatives for arbitrary Γ are investigated in [396]. A quite natural choice is $\Gamma = \omega (= \omega^I{}_J \Lambda_I{}^J)$. Then $\overset{\omega}{D} = D$ and

$$\overset{\omega}{\pounds}_X = \pounds_X + i_X \omega^I{}_J \Lambda^I{}_J. \tag{F.53}$$

The Lorentz-covariant Lie derivative obeys (in this compact notation $\Omega = \Omega^I{}_J \Lambda_I{}^J$)

$$\begin{aligned} [i_X, \overset{\omega}{\pounds}_Y]\mathcal{F} &= i_{[X,Y]}\mathcal{F} \\ \left[\overset{\omega}{\pounds}_Y, D\right]\mathcal{F} &= i_X \Omega \wedge \mathcal{F} \\ \left[\overset{\omega}{\pounds}_X, \overset{\omega}{\pounds}_Y\right]\mathcal{F} &= \overset{\omega}{\pounds}_{[X,Y]} \mathcal{F} + i_Y i_X \Omega \wedge \mathcal{F}. \end{aligned}$$

Now it is obvious to define a modified diffeomorphism:

$$\overset{\omega}{\delta}_D (\xi) := \overset{\omega}{\pounds}_\xi = \delta_D(\xi) + \delta_L(i_\xi \omega) \tag{F.54}$$

which is a mixture of a general coordinate transformation and a specific Lorentz transformation for which

$$[\overset{\omega}{\delta}_D (\xi), \delta_L(\lambda)] = \delta_L(\overset{\omega}{\pounds}_\xi \lambda) \qquad \left[\overset{\omega}{\delta}_D (\xi), \overset{\omega}{\delta}_D (\xi')\right] = \overset{\omega}{\delta}_D ([\xi, \xi']) + \delta_L(i_{\xi'} i_\xi \Omega).$$

The Lorentz-covariant Lie derivative acts on the fields as

$$\begin{aligned} \overset{\omega}{\pounds}_\xi \vartheta^I &= i_\xi \Theta^I + D i_\xi \vartheta^I \\ \overset{\omega}{\pounds}_\xi \Theta^I &= i_\xi \Omega^I{}_J \wedge \vartheta^J + \xi^J \Omega^I{}_J + D i_\xi \Theta^I \\ \overset{\omega}{\pounds}_\xi \Omega^I{}_J &= D i_\xi \Omega^I{}_J \wedge P_I{}^J. \end{aligned}$$

Substituting the Lorentz-covariant Lie derivative into (F.52), one finds after some algebra

$$\overset{\omega}{\pounds}_\xi\, \vartheta^I \wedge e_I + \overset{\omega}{\pounds}_\xi\, \Theta^I \wedge p_I + \overset{\omega}{\pounds}_\xi\, \Omega^I{}_J \wedge \pi_I{}^J - \overset{\omega}{\pounds}_\xi\, \tilde{\mathcal{L}}_G \equiv i_\xi \omega^I{}_J A_I{}^J \equiv 0, \tag{F.55}$$

where $A_I{}^J \equiv 0$ is the (Lorentz) Noether identity (F.44a). Indeed, it is shown in [396] that (F.52) has the same appearance for any covariant Lie derivative.

Noether Identities

As for the Lorentz symmetry case, we aim at writing the previous identity in the form $A^G + dB^G = \xi^I A_I^G + d(\xi^I B_I^G) \equiv 0$, and then inferring from the arbitrariness of the ξ^I that the expressions A_I^G and B_I^G must themselves vanish. (The upper index G, standing for "gravitation", is superfluous here, but it is introduced because later the corresponding expressions for the matter fields are investigated.) The different terms in (F.55) are exploited as

$$\begin{aligned}
\overset{\omega}{\pounds}_\xi\, \vartheta^I \wedge e_I &= i_\xi \Theta^I \wedge e_I + d(i_\xi \vartheta^I \wedge e_I) - \xi^I \wedge De_I \\
\overset{\omega}{\pounds}_\xi\, \Theta^I \wedge p_I &= i_\xi \Omega^I{}_J \wedge c_I{}^J + \xi^J \Omega^I{}_J \wedge p_I + d(i_\xi \Theta^I \wedge p_I) + i_\epsilon \Theta^I \wedge Dp_I \\
\overset{\omega}{\pounds}_\xi\, \Omega^I{}_J \wedge \pi_I{}^J &= d(i_\xi \Omega^I{}_J \wedge \pi_I{}^J) + i_\xi \Omega^I{}_J \wedge D\pi_I{}^J \\
\overset{\omega}{\pounds}_\xi\, \tilde{\mathcal{L}}_G &= (i_\xi D + D i_\xi)\tilde{\mathcal{L}}_G = D i_\xi \tilde{\mathcal{L}}_G = d i_\xi \tilde{\mathcal{L}}_G .
\end{aligned}$$

Collecting all the terms, one eventually obtains

$$A^G = i_\xi \Theta^I \wedge (e_I + Dp_I) + i_\xi \Omega^I{}_J \wedge (c^J{}_I + D\pi_I{}^J) - \xi^I \wedge De_I + \xi^J \Omega^I{}_J \wedge p_I .$$

This can be expressed as per (F.24, F.25) in terms of (the vacuum gravity) Euler derivatives as

$$A^G = i_\xi \Theta^I \wedge E_I + i_\xi \Omega^I{}_J \wedge C_I^J - \xi^I \wedge DE_I$$

since $\Omega^J{}_I \wedge p_J = -DDp_I$. Furthermore,

$$B^G = \xi^I \wedge (E_I - Dp_I) + i_\xi \Theta^I \wedge p_I + i_\xi \Omega^I{}_J \wedge \pi^J{}_I - i_\xi \tilde{\mathcal{L}}_G .$$

By writing

$$i_\xi F = i_{(\xi^I e_I)} F = \xi^I i_{(e_I)} F =: \xi^I i_I F,$$

both A^G and B^G can be expanded as

$$A_I^G = i_I \Theta^J \wedge E_J + i_I \Omega^K{}_J \wedge C_K^J - DE_I \equiv 0 \tag{F.56a}$$

$$B_I^G = (E_I - Dp_I) + i_I \Theta^J \wedge p_J + i_I \Omega^K{}_J \wedge \pi^J{}_K - i_I \tilde{\mathcal{L}}_G \equiv 0. \tag{F.56b}$$

The first set of these relations represents the Noether identities of any generally covariant vacuum gravity theory with respect to the diffeomorphism symmetry. This can also be read as another relation between the Einstein and the Cartan forms:

$$DE_I = i_I \Theta^J \wedge E_J + i_I \Omega^K{}_J \wedge C^J_K. \tag{F.57}$$

The second set of identities in (F.56) yields an alternative definition of the Einstein form:

$$E_I = D\big(\frac{\partial \tilde{\mathcal{L}}_G}{\partial \Theta^I}\big) - i_I \Theta^J \wedge \big(\frac{\partial \tilde{\mathcal{L}}_G}{\partial \Theta^J}\big) - i_I \Omega^K{}_J \wedge \big(\frac{\partial \tilde{\mathcal{L}}_G}{\partial \Omega^K{}_J}\big) + i_I \tilde{\mathcal{L}}_G.$$

Noether Identities for Diffeomorphism Symmetries in Full

If matter is coupled to gravity[10], we need to explore the full theory defined by a Lagrangian $\mathcal{L} = \mathcal{L}_G + \mathcal{L}_M$. Then (F.55) is augmented by "matter terms":

$$\begin{aligned}
&\overset{\omega}{\pounds}_\xi\, \vartheta^I \wedge e_I + \overset{\omega}{\pounds}_\xi\, \Theta^I \wedge p_I + \overset{\omega}{\pounds}_\xi\, \Omega^I{}_J \wedge \pi_I{}^J - \overset{\omega}{\pounds}_\xi\, \tilde{\mathcal{L}}_G \\
&\quad - \overset{\omega}{\pounds}_\xi\, \vartheta^I \wedge T_I + \overset{\omega}{\pounds}_\xi\, \Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{\partial \Phi^\alpha} + \overset{\omega}{\pounds}_\xi\, D\Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{D\Phi^\alpha} - \overset{\omega}{\pounds}_\xi\, \tilde{\mathcal{L}}_M \\
&\qquad = A + dB = (A^G + dB^G) + (A^M + dB^M) \equiv 0.
\end{aligned}$$

Here the first line is identical to $(A^G + dB^G)$, that is the expression derived in the previous paragraph. The second line is evaluated as

$$\begin{aligned}
A^M + dB^M = {}& -i_\xi \Theta^I \wedge T_I - D i_\xi \vartheta^I \wedge T_I + i_\xi D\Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{\partial \Phi^\alpha} + D i_\xi \Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{\partial \Phi^\alpha} \\
& + i_\xi DD\Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{D\Phi^\alpha} + D i_\xi D\Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{D\Phi^\alpha} - d i_\xi \tilde{\mathcal{L}}_M.
\end{aligned}$$

After some algebra (shifting the D-derivatives and relating the curvature term to the covariant derivative of the matter field equation), one arrives at

$$\begin{aligned}
A^M + dB^M = {}& - i_\xi \Theta^I \wedge T_I + \xi^I DT_I - i_\xi \Omega^I{}_J \wedge S_I{}^J + (-1)^{|\Phi^\alpha|} i_\xi \Phi^\alpha \wedge D\mathcal{E}_\alpha \\
& + i_\xi D\Phi^\alpha \wedge \mathcal{E}_\alpha + d\Big(i_\xi \Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{\partial \Phi^\alpha} + i_\xi D\Phi^\alpha \wedge \frac{\partial \tilde{\mathcal{L}}_M}{\partial D\Phi^\alpha} - \xi^I T_I - i_\xi \tilde{\mathcal{L}}_M\Big).
\end{aligned}$$

Taking together the contributions for the gravitational part (according to (F.56)) and the "matter part" results in:

[10] Observe how this formulation is self-evident.

$$A_I = i_I\Theta^J \wedge \mathcal{E}_J + i_I\Omega^K{}_J \wedge \mathcal{E}_K{}^J - D\mathcal{E}_I + (-1)^{|\Phi^\alpha|} i_I\Phi^\alpha \wedge D\mathcal{E}_\alpha \\ + i_I D\Phi^\alpha \wedge \mathcal{E}_\alpha \equiv 0 \tag{F.58a}$$

$$B_I = (\mathcal{E}_I - Dp_I) + i_I\Theta^J \wedge p_J + i_I\Omega^K{}_J \wedge \pi^J{}_K + i_I\Phi^\alpha \wedge \frac{\partial\tilde{\mathcal{L}}_M}{\partial\Phi^\alpha} \\ + i_I D\Phi^\alpha \wedge \frac{\partial\tilde{\mathcal{L}}_M}{\partial D\Phi^\alpha} - i_I\tilde{\mathcal{L}} \equiv 0. \tag{F.58b}$$

The first expression $A_I \equiv 0$ is the complete diffeomorphism Noether identity. Observe that this holds for any gravitational theory built solely from curvature and torsion and which couples minimally to arbitrary matter fields, as long as they transform according to a representation of the Lorentz group.

Field Theory in a Fixed Gravitational Background

Various conclusions can be drawn from the identities (F.58) if the gravitational field is not dynamical; they become

$$A_I^M = -i_I\Theta^J \wedge T_J - i_I\Omega^K{}_J \wedge S_K{}^J - DT_I + (-1)^{|\Phi^\alpha|} i_I\Phi^\alpha \wedge D\mathcal{E}_\alpha \\ + i_I D\Phi^\alpha \wedge \mathcal{E}_\alpha \equiv 0 \tag{F.59a}$$

$$B_I^M = -T_I + i_I\Phi^\alpha \wedge \frac{\partial\tilde{\mathcal{L}}_M}{\partial\Phi^\alpha} + i_I D\Phi^\alpha \wedge \frac{\partial\tilde{\mathcal{L}}_M}{\partial D\Phi^\alpha} - i_I\tilde{\mathcal{L}}_M \equiv 0. \tag{F.59b}$$

Therefore, on the solutions of the matter field equations, we find from (F.59a) the energy-momentum "conservation law"

$$DT_I \doteq i_I\Theta^J \wedge T_J + i_I\Omega^K{}_J \wedge S_K{}^J. \tag{F.60}$$

This can be considered together with the angular-momentum "conservation law" (F.46), namely $DS_I{}^J \doteq -\vartheta^{[J} \wedge T_{I]}$. Both of these relations can also be derived from the identities (F.45, F.57) and the field equations.

In Minkowski space, (F.60) becomes $dT_\mu = 0$, and therefore (F.46) can be written as

$$dS^{\mu\nu} + \frac{1}{2}dx^\mu \wedge T^\nu - \frac{1}{2}dx^\nu \wedge T^\mu = d(S^{\mu\nu} + \frac{1}{2}x^\mu \wedge T^\nu - \frac{1}{2}x^\nu \wedge T^\mu) \\ =: dJ^{\mu\nu} = 0.$$

From this, we recover the angular momentum part and the spin part in the total angular momentum density $J^{\mu\nu}$.

The identity (F.58b) on a fixed gravitational background becomes

$$T_I = i_I\Phi^\alpha \wedge \frac{\partial\tilde{\mathcal{L}}_M}{\partial\Phi^\alpha} + i_I D\Phi^\alpha \wedge \frac{\partial\tilde{\mathcal{L}}_M}{\partial D\Phi^\alpha} - i_I\tilde{\mathcal{L}}, \tag{F.61}$$

an expression for the energy-momentum current defined in terms of the matter fields. Notice that originally it is defined via the derivative of the matter Lagrangian with respect to the tetrads; compare (F.18a)[11]

Example: Klein-Gordon Field Coupled to Einstein-Cartan Gravity

The previous expressions are quite generic and hold for all theories with Lagrangians of the form $\mathcal{L} = \mathcal{L}_G(\vartheta, \Theta, \Omega) + \mathcal{L}_M(\vartheta, \Phi, D\Phi)$. In order to comprehend their essence, let us consider the simplest case:

$$\mathcal{L}_G = \frac{1}{2\kappa}\Omega^{IJ} \wedge \eta_{IJ}, \qquad \mathcal{L}_M = \frac{1}{2}(d\phi \wedge *d\phi) - V, \qquad V = \frac{1}{2}m^2(\phi \wedge *\phi).$$

This describes a massive scalar field in a Einstein-Cartan theory (without a cosmological constant). The variation of the gravitational part was already treated previously with the result

$$\delta\mathcal{L}_G = \frac{1}{2\kappa}\delta\vartheta^I \wedge \Omega^{KJ} \wedge \eta_{KJI} + \frac{1}{2\kappa}\delta\omega^I_J \wedge D\eta_I{}^J + dB_G,$$

with an exact form B_G which has no influence on the field equations. From this we read off the Einstein and the Cartan forms according to (F.17):

$$E_I = \frac{1}{2\kappa}\Omega^{JK} \wedge \eta_{JKI}, \qquad C_K{}^J = \frac{1}{2\kappa}D\eta_K{}^J.$$

The variation of the matter Lagrangian is

$$\begin{aligned}\delta\mathcal{L}_M &= \frac{1}{2}\delta(d\phi) \wedge *d\phi \\ &= \delta d\phi \wedge *d\phi + \frac{1}{2}\delta\vartheta^I \wedge [(-1)d\phi \wedge i_I * d\phi - i_I d\phi \wedge *d\phi] - \delta V.\end{aligned}$$

Furthermore,

$$\delta V = \frac{1}{2}m^2\ \delta(\phi \wedge *\phi) = m^2\ \delta\phi \wedge *\phi + \frac{1}{2}m^2\ \delta\vartheta^I \wedge \phi \wedge i_I * \phi$$

so that

$$\delta\mathcal{L}_M = -\delta\varphi \wedge (d * d\phi + m^2 * \phi) - \delta\vartheta^I \wedge T_I + dB_M$$

[11] This reflects the astounding finding in Sect. 7.5.2 that the Belinfante energy-momentum tensor for a matter theory is derived by coupling the matter fields to gravity.

with the matter energy-momentum current according to (F.18a)

$$T_I = \frac{1}{2}\left[d\phi \wedge i_I * d\phi + i_I d\phi \wedge *d\phi\right] + \frac{1}{2}m^2\ \phi \wedge i_I * \phi. \tag{F.62}$$

The spin current $S_I{}^J$ vanishes identically. Collecting all terms, the Euler-Lagrange equations are found:

$$\begin{aligned}
\mathcal{E}_\phi &= -d * d\phi - m^2 * \phi = 0\\
\mathcal{E}_I &= \frac{1}{2\kappa}\Omega^{JK} \wedge \eta_{JKI} - T_I = 0\\
\mathcal{E}_K{}^J &= \frac{1}{2\kappa} D\eta_K{}^J = 0.
\end{aligned}$$

We verify that the equation $\mathcal{E}_\phi = 0$ is equivalent to the Klein-Gordon equation: Write

$$d * d\phi = d * \phi_{,\mu} dx^\mu = d(\phi_{,\mu} * dx^\mu) = \phi_{,\mu\nu} dx^\nu \wedge *dx^\mu = \phi_{,\mu\nu} g^{\nu\mu} \eta.$$

Then $\mathcal{E}_\phi = -(\partial_\mu \partial^\mu + m^2)\phi\ \eta$, and $*\mathcal{E}_\phi = 0$ is indeed the Klein-Gordon equation. Furthermore, the energy-momentum tensor (F.62) can be rewritten as

$$T_I = \frac{1}{2}[2i_I d\phi \wedge *d\phi - i_I(d\phi \wedge *d\phi)] + i_I V = i_I d\phi \wedge *d\phi - i_I \mathcal{L}_M,$$

and is then identical to the canonical energy-momentum tensor for the scalar field. This complies with (F.51) in that $i_I \phi \equiv 0$.

F.3 Gravitational Theories

F.3.1 Energy-Momentum Conservation

Since in Minkowski space energy-momentum conservation has its roots in the invariance of the field theory actions with respect to spacetime transformations, it makes sense to try to identify energy-momentum as conserved charges adjoined to diffeomorphism currents. And since in all other theories energy is always the numerical value of the Hamiltonian, it makes sense to look for an energy-momentum derived from a Hamiltonian.

From (F.50) (and the preceding equation) we identify the diffeomorphism current

$$J_\xi = \pounds_\xi Q \wedge P - i_\xi \mathcal{L} = -i_\xi Q \wedge \mathcal{E} + d(i_\xi Q \wedge P). \tag{F.63}$$

leading according to (F.38) to diffeomorphism charges

$$C[\xi] = \int_{\partial S} U_\xi = \int_{\partial S} (i_\xi Q \wedge P).$$

Symplectic Approach à la Nester

The following is largely adopted from [383]; J. M. Nester and his research group opened up another approach towards a comprehensive understanding of the long-term problem of energy-momentum conservation in general relativity with its relations to the notions of quasilocal energy, boundary conditions in the GR action, and pseudo energy-momentum tensors; see also [89], [382]. They specifically place the Hamiltonian of a theory at the center of their considerations.

In the main text, it is explained that a Hamiltonian in a field theory is defined only with respect to the choice of a time variable; see Subsect. 3.4.4. In geometrical terms this means that spacetime is foliated into spacelike surfaces Σ_t (parametrized by t) together with a transversal timelike vector field N, such that $i_N\, dt = 1$. Then any form α can be decomposed into its "time" component $\hat{\alpha} := i_N \alpha$ and its "spatial" projection $\underline{\alpha} = \alpha - dt \wedge \hat{\alpha}$. Specifically,

$$\mathcal{L} = dt \wedge i_N \mathcal{L}.$$

Take for instance the first-order (or Palatini) version of the Maxwell Lagrangian: $\overline{\mathcal{L}}_{ED} = F \wedge \Pi = dA \wedge \Pi$, where Π is the momentum canonically conjugate to A. From this Lagrangian one can construct the Hamiltonian three-form

$$\mathcal{H}_{ED}(N) = \pounds_N A \wedge \Pi - i_N \overline{\mathcal{L}}_{ED}. \tag{F.64}$$

Explicitly,

$$\begin{aligned} \mathcal{H}_{ED}(N) &= \pounds_N A \wedge \Pi - i_N dA \wedge \Pi - F \wedge i_N \Pi \\ &= -d i_N A \wedge \Pi - F \wedge i_N \Pi \\ &= -i_N A \wedge d\Pi - F \wedge i_N \Pi + d(i_N A \wedge \Pi). \end{aligned} \tag{F.65}$$

Now, also in the case of gravitational fields it is possible to rewrite the Lagrangian (F.28) as

$$i_N \overline{\mathcal{L}}_G = \pounds_N \vartheta \wedge p + \pounds_N \omega \wedge \pi - \mathcal{H}_G(N) = \pounds_N Q \wedge P - \mathcal{H}_G(N), \tag{F.66}$$

where Q stands for the fields in the set $\{\vartheta^K, \omega^K{}_L\}$ and P for their conjugate momenta $\{p_K, \pi_K{}^L\}$. The explicit expression for the Hamiltonian $\mathcal{H}_G(N)$ is calculated to be

$$\begin{aligned} \mathcal{H}_G(N) = -\Theta \wedge i_N p - \Omega \wedge i_N \pi - i_N \vartheta \wedge Dp - i_N \omega \wedge \big(D\pi + [\vartheta \wedge p]\big) \\ + i_N \Lambda \ + d\mathcal{B}(N), \end{aligned}$$

with

$$\mathcal{B}(N) = i_N\vartheta \wedge p + i_N\omega \wedge \pi = i_N Q \wedge P. \tag{F.67}$$

The gravitational Hamiltonian can also be written as

$$\mathcal{H}_G(N) = -i_N Q \wedge \frac{\delta\overline{\mathcal{L}}_G}{\delta P} - \frac{\delta\overline{\mathcal{L}}_G}{\delta Q} \wedge i_N P + d\mathcal{B}(N).$$

The expression (F.66) together with previous findings allows us to derive conditions which must be met in order that the Hamiltonian field equations can be derived. First, vary (F.66):

$$\delta(i_N\overline{\mathcal{L}}_G) = (\delta\pounds_N Q) \wedge P + \pounds_N Q \wedge \delta P - \delta\mathcal{H}_G(N).$$

Now compare this with $i_N(\delta\overline{\mathcal{L}}_G)$ using (F.29):

$$\begin{aligned} i_N\delta\overline{\mathcal{L}}_G &= i_N d(\delta Q \wedge P) + [\text{on} - \text{shell terms}] \\ &\doteq \pounds_N(\delta Q \wedge P) - d i_N(\delta Q \wedge P) \\ &= (\pounds_N\delta Q) \wedge P + \delta Q \wedge \pounds_N P - d i_N(\delta Q \wedge P). \end{aligned}$$

Since N is not varied, $\delta i_N\overline{\mathcal{L}}_G = i_N\delta\overline{\mathcal{L}}_G$, and thus

$$\delta\mathcal{H}_G(N) = \pounds_N Q \wedge \delta P - \delta Q \wedge \pounds_N P + d\mathcal{C}(N) \qquad \mathcal{C}(N) = i_N(\delta Q \wedge P). \tag{F.68}$$

Therefore, if the surface term $d\mathcal{C}(N)$ vanishes, we are allowed to derive from the previous expression the Hamiltonian field equations

$$\pounds_N Q^\alpha = \frac{\delta\mathcal{H}_G(N)}{\delta P_\alpha} \qquad \pounds_N P_\alpha = -\frac{\delta\mathcal{H}_G(N)}{\delta Q^\alpha}.$$

This is not only reminiscent of the findings by T. Regge and C. Teitelboim [437], details in Appendix C.3, but it also offers the possibility of introducing appropriate closed forms in the gravitational Lagrangians in order to cope with different asymptotic spacetime geometries.

Observe that the Hamiltonian $\mathcal{H}_G(N)$ defined by (F.66) is identical with the diffeomorphism Noether current J_N (F.63) with the superpotential $\mathcal{B}(N) = U_{(\xi=N)}$. Thus,

$$d\mathcal{H}_G(N) = \big[\text{linear combination of field equations}\big], \tag{F.69}$$

and by Stokes' theorem the integrated Hamiltonian is

$$H(N) = \int_\Sigma \big[\text{linear combination of field equations}\big] + \int_{\partial\Sigma} \mathcal{B}(N).$$

Therefore $H(N)$, which can be identified with the energy in the finite region Σ (a three-dimensional spatial hypersurface) is numerically determined solely by the boundary term, and is thus a quasilocal energy-momentum.

However, the Hamiltonian is not unique, since one can add to it a total derivative without affecting the conservation law (F.69). This amounts to a change of $\mathcal{B}(N)$ and $\mathcal{C}(N)$. Instead of interpreting this as a weakness (indeed this indeterminacy is common to any Noether current), it opens an arena for defining different gravitational energies in a controlled way, where the control is due to boundary terms. As a matter of fact, it is found that with the further requirement of covariant boundaries there are only two possibilities for the $\mathcal{B}(N)$. I refrain from giving these expressions here (they are defined in terms of a background metric) and refer to [89], [382], were you may also find details for the case of general relativity. With the different boundary terms, they reproduce the various pseudotensors, superpotentials, and the Komar and the ADM mass. This "covariant Hamiltonian formulation", as it is called by Nester *et al.*, also allows us to understand anew the role of the Gibbons-Hawking-York boundary term in the gravitational Lagrangian and the Brown-York approach towards quasi-local energies.

On this level of consideration, the phase-space constraints which result from local invariances, in the sense discussed in Appendix C, are not yet visible. The constraints are found in the differential form notation by varying

$$S = \int dt \int_{\Sigma_t} \pounds_N Q^\alpha \wedge P_\alpha - \mathcal{H}_G(N).$$

with respect to the temporal components $\hat{Q}^\alpha$, $\hat{P}_\alpha$, and the time-evolution equation results from the variations with respect to the spatial components $\underline{Q}^\alpha$, $\underline{P}_\alpha$.

Symplectic Approach à la Wald

The proper technical basis for the following Noether charge approach was developed by J. Lee and R. Wald [338], originally not in the language of differential forms, in an attempt to understand the relationship between local symmetries occurring in a Lagrangian formulation of a field theory, and the corresponding constraints present in a phase-space formulation. Later, these techniques gave an impetus for the derivation of the black-hole thermodynamics laws directly from diffeomorphism invariance of a theory [527]. The techniques were refined in [528] in order to become applicable to defining counterparts to conserved quantities at null infinity.

In the approach of Wald *et al.*, considerable weight is given to the symplectic potential Θ as it appears in the variation of a Lagrangian

$$\delta\mathcal{L} = \delta\Phi \wedge \mathcal{E} + d\Theta(Q, \delta Q). \tag{F.70}$$

It serves to define the 'symplectic current' ω:

$$\omega(Q; \delta_1 Q, \delta_2 Q) := \delta_1\Theta(Q, \delta_2 Q) - \delta_2\Theta(Q, \delta_1 Q). \tag{F.71}$$

For its exterior derivative one finds, by using (F.70)

$$d\omega = -\delta_2 Q \wedge \delta_1 \mathcal{E} + \delta_1 Q \wedge \delta_2 \mathcal{E}.$$

This vanishes on-shell and/or for variations $\delta_1 Q$, $\delta_2 Q$ that solve the linearized field equations. Furthermore, define the symplectic form relative to a Cauchy surface Σ as

$$\Omega(Q; \delta_1 Q, \delta_2 Q) := \int_\Sigma \omega(Q; \delta_1 Q, \delta_2 Q).$$

This form is independent of Σ if $Q, \delta_1 Q$, and $\delta_2 Q$ are solutions–provided that either Σ is compact or the fields satisfy appropriate asymptotic conditions. The form Ω is degenerate (presymplectic) on the space $\mathcal{F}$ of all field configurations, but after factoring $\mathcal{F}$ by the degeneracy subspace of Ω, one can define a phase space Γ; for details see [338]. A Hamiltonian H_ξ conjugate to a vector field ξ^μ is a function on Γ whose pull-back to $\mathcal{F}$ satisfies

$$\delta H_\xi = \Omega(\Phi; \delta Q, \pounds_\xi Q). \tag{F.72}$$

The previous considerations apply to any field theory with local symmetries. Next, consider specifically diffeomorphism invariant theories: Diffeomorphisms are generated by smooth vector fields ξ^μ. Every vector field generates a local symmetry and gives rise to a Noether current $(D\text{-}1)$-form

$$J_\xi := \Theta(Q, \pounds_\xi Q) - i_\xi \mathcal{L}.$$

We verify that from

$$dJ_\xi = d\Theta(Q, \pounds_\xi Q) - di_\xi \mathcal{L} = \pounds_\xi \mathcal{L} - \pounds_\xi Q \wedge \mathcal{E} - di_\xi \mathcal{L} = -\pounds_\xi Q \wedge \mathcal{E},$$

the Noether currents are closed forms for all ξ whenever the field equations are fulfilled (i.e. $\mathcal{E} \doteq 0$) or whenever ξ is a symmetry (i.e $\pounds_\xi Q = 0$). Therefore, on solutions, the Noether current is exact: $J_\xi = dU_\xi$ with $(D\text{-}2)$-forms U_ξ. It is shown in [294]) that on-shell

$$J_\xi = dU_\xi + \xi^\mu C_\mu.$$

where $C_\mu = 0$ are the constraint equations of the theory. The variation of the Noether current is (on-shell)

$$\begin{aligned}
\delta J_\xi &= \delta\Theta(Q, \pounds_\xi Q) - i_\xi \delta\mathcal{L} = \delta\Theta(Q, \pounds_\xi Q) - i_\xi d\Theta(Q, \delta Q) \\
&= \delta\Theta(Q, \pounds_\xi Q) - \pounds_\xi \Theta(Q, \delta Q) + d(i_\xi \Theta(Q, \delta Q)) \\
&= \omega(Q; \delta Q, \pounds_\xi Q) + d(i_\xi \Theta(Q, \delta Q))
\end{aligned}$$

where in the last step the symplectic current according to (F.71) is introduced. Therefore

$$\omega(Q;\delta Q,\pounds_\xi Q)=\delta J_\xi-d(i_\xi\Theta)=\xi^\mu\delta C_\mu+\delta dQ_\xi-d(i_\xi\Theta)$$

which integrates to

$$\Omega(Q;\delta Q,\pounds_\xi Q)=\int_\Sigma\xi^\mu\delta C_\mu+\int_{\partial\Sigma}(\delta Q_\xi-i_\xi\Theta).$$

If a D-1 form B with

$$\delta\int_{\partial\Sigma}i_\xi B=\int_{\partial\Sigma}i_\xi\Theta$$

can be found, then a Hamiltonian exists according to (F.72), and is given by

$$H_\xi=\int_\Sigma\xi^\mu C_\mu+\int_{\partial\Sigma}(Q_\xi-i_\xi\mathrm{B}).$$

Examples:

- The energy of a system can be defined if a Hamiltonian conjugate to "time translations" $\xi^\mu=t^\mu$ exists
- If a Hamiltonian exists that is conjugate to a rotation the angular momentum can be defined as

$$J=-H_\varphi=-\int_{\partial\Sigma}Q_\varphi$$

  where φ is assumed to be tangent to Σ, so that the B-contribution vanishes.

The previous definitions and the ensuing relations hold for all diffeomorphism invariant theories. Take now as an example general relativity with the Hilbert Lagrangian four-form components $\mathcal{L}=\frac{1}{16\pi}\epsilon_{\mu\nu\rho\sigma}R$. The components of the symplectic 3-form are

$$\Theta_{\mu\nu\rho}=\frac{1}{16\pi}\epsilon_{\lambda\mu\nu\rho}g^{\lambda\sigma}g^{\kappa\tau}(D_\kappa\delta g_{\sigma\tau}-D_\sigma\delta g_{\kappa\tau}).$$

From this, one derives the Noether current and the Noether charge (components)

$$(J_\xi)_{\mu\nu\rho}=\frac{1}{8\pi}\epsilon_{\lambda\mu\nu\rho}D_\sigma(D^{[\sigma}\xi^{\lambda]})\qquad\qquad(Q_\xi)_{\mu\nu}=-\frac{1}{16\pi}\epsilon_{\mu\nu\rho\sigma}D^\rho\xi^\sigma.$$

If spacetime is asymptotically flat, the angular momentum defined by this charge becomes identical to the ADM expression, and when (ξ) is an asymptotic Killing vector, it agrees with Komar's expression. Furthermore, for asymptotic time translations, expressions for the B can be found, such that a Hamiltonian exists and that the energy becomes the ADM energy.

F.3.2 Gravitational Theories Beyond Einstein

The previous considerations are generic and apply not only to general relativity. In principle you can "build" your own action functionals by building wedge products out of the forms ϑ, Θ, Ω, such that the product is a D-form. But by far not all of these mathematically admissible gravitational theories describe our world. They must be at least as good as GR in the sense of reproducing the experimental data. From the point of geometry, a "natural" candidate is the Einstein-Cartan theory, which as explained in Subsect. 7.3.4 is a theory in a Riemann-Cartan space. This geometry specializes to a Riemann space for vanishing torsion, and it reduces to a Weizenböck's teleparallel space in the case of vanishing curvature. From the motivation to understand gravitational theories as a gauge theory of the Poincaré group a Lagrangian emerged that has terms linear and quadratic in the curvature, and terms quadratic in the torsion; in a shorthand notation $R + T^2 + R^2$. Explicitly this is a linear combination of ten terms with, in principle, ten independent coefficients. This 10-parameter Lagrangian can of course also be expressed in terms of differential forms, and–to stress this point again–all the previous results on the Noether currents with respect to local Lorentz and diffeomorphism invariance hold true. The theories with vanishing curvature and those with vanishing torsion were–in terms of the exterior calculus–further investigated in [325]. The Noether identities for the teleparallel theory, which is a specific linear combination of the three torsion T^2 terms (more about this below), are treated in [397].

The component form of the full PGT theory was formulated explicitly by K. Hayashi and T. Shirafuji [259]. In order to better understand its structure, they split the curvature and the torsion into its irreducible components with regard to the Lorentz group. This was reformulated by R. Wallner in the language of the exterior calculus [529]: In 4D, the torsion can be written as

$$\Theta^I = \Gamma^I + \frac{1}{3}\,\vartheta^I \wedge \Lambda + \frac{1}{3}\ *(\Xi \wedge \vartheta^I).$$

Here Λ is the "trace" part (a 1-form)

$$\Lambda := i_I \Theta^I i_I, \qquad \left(\Theta^I - \frac{1}{3}\vartheta^I \wedge \Lambda\right) = 0$$

and Ξ is the 1-form "axial" part

$$\Xi := *(\Theta^I \wedge \vartheta_I), \qquad \Big(\Theta^I - \frac{1}{3} * (\Xi \wedge \vartheta^I)\Big) \wedge \vartheta_I = 0.$$

Then Γ^I can be proven to be traceless with a vanishing axial part:

$$i_I \Gamma^I = 0 \qquad\qquad \Gamma^I \wedge \vartheta_I = 0.$$

Thus, torsion Θ^I with its 24 components is decomposed in terms of the Λ (4 components), the Ξ (4 components) and the Γ^I (16 components).

Also, the curvature with its 36 components can be decomposed into $1+6+9+9+10+1$ irreducible components.

The U_4 objects can be split into V_4 objects plus further contributions by splitting the Riemann-Cartan connection $\tilde{\omega}^I{}_J$ into the Riemann connection $\omega^I{}_J$ and the contortion $K^I{}_J$ as

$$\tilde{\omega}^I{}_J = \omega^I{}_J + K^I{}_J \qquad \text{where} \qquad d\vartheta^I + \omega^I{}_J \wedge \vartheta^J = 0.$$

With this, the U_4 curvature and torsion become

$$\Theta^I = K^{IJ} \wedge \vartheta_J \qquad\qquad \tilde{\Omega}^I{}_J = \Omega^I{}_J + DK^I{}_J + K^I{}_L \wedge K^L{}_J \tag{F.73}$$

where D is the Riemann covariant derivative.

If we specifically consider the $R + T^2$ Lagrangian (without a cosmological constant), it has four terms, namely

$$\mathcal{L} = \sum_i^4 a_i \, \mathcal{L}_i = a_1(\Gamma^I \wedge *\Gamma_I) + a_2(\Lambda \wedge *\Lambda) + a_3(\Xi \wedge *\Xi) + a_4 \, \tilde{\Omega}_{IJ} \wedge \eta^{IJ}.$$

The tilde notation in $\mathcal{L}_4$ indicates that the curvature form lives in Riemann-Cartan space; indeed this Lagrangian alone describes the Einstein-Cartan theory. The (Einstein-Cartan) R term in the Lagrangian can be split into an (Einstein) R term plus further contributions by the previous splitting of the Riemann-Cartan connection: The Lagrangian term $\mathcal{L}_4$ becomes

$$\tilde{\Omega}_{IJ} \wedge \eta^{IJ} = \Omega_{IJ} \wedge \eta^{IJ} + K_{IL} \wedge K^L{}_J \wedge \eta^{IJ} + d(K_{IJ} \wedge \eta^{IJ}).$$

The second expression on the right-hand can be written in terms of torsion. This is a three-page exercise in the exterior calculus with the resulting identity

$$\tilde{\Omega}_{IJ} \wedge \eta^{IJ} = \Omega_{IJ} \wedge \eta^{IJ} + (\Gamma^I \wedge *\Gamma_I) - \frac{2}{3}(\Lambda \wedge *\Lambda) + \frac{1}{6}(\Xi \wedge *\Xi) + \text{exact form}.$$

Teleparallel GR

This leads to the remarkable observation that in a spacetime with vanishing curvature $\tilde{\Omega}$, the Hilbert-Einstein action is–up to a boundary term–equivalent to a specific T^2 theory, called the teleparallel equivalent of general relativity ($GR_{\|}$). In any teleparallel theory, one can find (under the condition that the manifold is parallelizable) an orthonormal frame in which the connection one-forms vanish; see e.g. Chapter 3 in [49]. This frame is unique only up to constant rotations. The advantage of this frame is that the torsion looks like a gauge field strength: $F^I := \Theta^I = d\vartheta^I$. It can indeed be interpreted as a field strength of a gauge theory of the translation group giving rise to a covariant derivative $d+\vartheta$. (Strictly speaking, this description of teleparallel theories differs in some aspects from the "proper" approach in which in the PGT Lagrangian a term with Lagrangian multipliers forces the curvature to become zero; see [397].)

Write the Lagrangian of the teleparallel equivalent of GR as

$$\mathcal{L}_{GR_{\|}} = -\frac{1}{2}F^I \wedge \Pi_I + \mathcal{L}_M(\vartheta, \Phi, D\Phi), \quad \Pi_I = \frac{1}{\kappa} * \left({}^{(1)}F_I - 2\,{}^{(2)}F_I - \frac{1}{2}\,{}^{(3)}F_I\right)$$

with ${}^{(2)}F^I = \frac{1}{3}\vartheta^I \wedge i_I F^I$, ${}^{(3)}F^I = \frac{1}{3}i^I(\vartheta^J \wedge F_J)$, ${}^{(1)}F^I = F_I - {}^{(2)}F_I - {}^{(3)}F_I$. The variations of the Lagrangian with respect to the coframe ϑ and the field strength F results in the field equations

$$\Sigma_I + T_I - d\Pi_I = 0$$

with the matter energy-momentum $T_I := \frac{\delta\mathcal{L}_M}{\delta\vartheta}$, and the gravitational energy-momentum

$$\Sigma_I := \frac{\partial\mathcal{L}_G}{\partial\vartheta} = i_I\mathcal{L}_G + i_I F^J \wedge \Pi_J = \frac{1}{2}\left[i_I F^J \wedge \Pi_J - F^J \wedge i_I \Pi_J\right].$$

Notice that this term has a structure similar to the energy-momentum expressions for the matter fields. One thus could be tempted to use the exact form $J_I := \Sigma_I + T_I$ to define the total four-momentum

$$P_I = \int_S (\Sigma_I + T_I) \doteq \int_{\partial S} \Pi_I.$$

However, to qualify as representing a genuine conserved quantity, this must be covariant under both general coordinate and local Lorentz transformations. Covariance under general coordinate transformations is guaranteed, since P_I is expressed in terms of differential forms. But as shown in [397], under a Lorentz transformation

$$P'_I = \int_{\partial S} (\Lambda^{-1})^J{}_I \Pi_J - \frac{1}{2\kappa}\int_{\partial S} (\Lambda^{-1})^J{}_I (\Lambda^{-1})^K{}_L\, d\Lambda^L{}_N \wedge \eta_J{}^N{}_K.$$

Thus the "energy-momentum" transforms as a vector only under those Lorentz transformations for which $d\Lambda$ vanishes on the boundary ∂S.

And of course we meet again the arbitrariness of allowing a closed form to be added to the current J_I: The changes $\Sigma_I \to \Sigma_I + d\Psi_I$ and $\Pi_I \to \Pi_I + \Psi_I$ leave the field equations untouched, but give rise to a change in the object P_I. This change can be compensated by a surface term $d\Psi$ in the action with $\frac{\partial \Psi}{\partial \vartheta^I} = \Psi_I$. Again, we see the interplay of boundary terms and conserved quantities.

F.3.3 Topological Terms

A "topological term" or "topological invariant" or "topological density" is the name for geometry-based surface terms in an action functional. The differential-form language facilitates a proper bookkeeping of topological terms.

Lanczos-Lovelock(-Cartan) Gravity

Differential forms permit an elegant formulation of Lanczos-Lovelock gravity in D dimensions as

$$\mathcal{L}_L = \sum_p \gamma_p \, \epsilon_{I_1 \ldots I_D} \, \Omega^{I_1 I_2} \wedge \ldots \wedge \Omega^{I_{2p-1} I_{2p}} \wedge \vartheta^{I_{2p+1} \ldots I_D}, \tag{F.74}$$

with (coupling) constants γ_p. Assuming the absence of torsion, the field equations are

$$\sum_p \gamma_p (D - 2p) \epsilon_{I_1 \ldots I_D} \Omega^{I_1 I_2} \wedge \ldots \wedge \Omega^{I_{2p-1} I_{2p}} \wedge \vartheta^{I_{2p+1} \ldots I_{D-1}} = 0.$$

Following [582], denote by $\mathcal{L}_{(C,V)}$ a summand in the Lanczos-Lovelock Lagrangian with the number of C curvature forms and $V = D$-$2C$ vielbein forms. Thus for $D = 2$ there are two terms

$$\mathcal{L}_{(0,2)} = \epsilon_{IJ} \vartheta^{IJ} = \eta \qquad \mathcal{L}_{(1,0)} = \epsilon_{IJ} \Omega^{IJ}.$$

The first one relates to a cosmological constant, while the second is proportional to the Euler density (to be explained below) in two dimensions. The corresponding Lagrangian term is a topological invariant. For $D = 3$, we have the terms $\mathcal{L}_{(0,3)} = \eta$, $\mathcal{L}_{(1,1)} = \epsilon_{IJK} \Omega^{IJ} \wedge \vartheta^K$, which are the cosmological constant term and the Hilbert Lagrangian. For $D = 4$ we find

$$\mathcal{L}_{(0,4)} = \eta, \qquad \mathcal{L}_{(1,2)} = \epsilon_{IJKL} \Omega^{IJ} \wedge \vartheta^{KL}, \qquad \mathcal{L}_{(2,0)} = \epsilon_{IJKL} \Omega^{IJ} \wedge \Omega^{KL}$$

corresponding to the cosmological term, the Hilbert term and the Euler density in four dimensions. This pattern is repeated for higher dimensions: $\mathcal{L}_{(0,D)}$ is always the cosmological term in a Lagrangian, $\mathcal{L}_{(1,D-2)}$ the Hilbert Lagrangian, and for D even $\mathcal{L}_{(D/2,0)}$ is the D-dimensional Euler density.

In Lanczos-Lovelock gravity, it was originally assumed that the underlying geometry is Riemannian. A Lanczos-Lovelock Lagrangian within Riemann-Cartan geometry was derived in [361]. This means that in a sum extending (F.74) additional terms are allowed, namely all those which are Lorentz invariant in this extended context. These are not only terms involving torsion explicitly, but also terms which vanish for vanishing torsion due to Bianchi identities. What does this give in four dimensions? The authors of [361] find three additional terms in the Lanczos-Lovelock Lagrangian:

$$\Omega^{IJ} \wedge \vartheta_{IJ}, \qquad \Theta^I \wedge \Theta_I, \qquad \Omega^{IJ} \wedge \Omega_{IJ}.$$

The latter is the 4D-Pontryagin density which is locally exact:

$$P_4 := \Omega^{IJ} \wedge \Omega_{IJ} = d\Big(\omega^I{}_J \wedge d\omega^J{}_I + \frac{2}{3}\omega^I{}_J \wedge \omega^J{}_K \wedge \omega^K{}_I\Big).$$

A linear combination of the other two terms is again a topological term

$$NY_4 := \Theta^I \wedge \Theta_I + \Omega_{IJ} \wedge \vartheta^{IJ} = d\Big(\vartheta^I \wedge \Theta_I\Big)$$

known as the Nieh-Yan density. Together with the previously derived Euler density

$$E_4 = \epsilon_{IJKL}\, \Omega^{IJ} \wedge \Omega^{KL} = d\Big(\omega^{IJ} \wedge (\Omega_{KL} - \frac{1}{3}\omega^K{}_M \wedge \omega^{LM})\epsilon_{IJKL}\Big)$$

we have three topological terms in 4D. The three-forms from which they are derived by applying an external derivative are called Chern-Simons forms. Written in components,

$$E_4 = (R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma})\eta$$

which shows that the Euler 4-form density has the Gauß-Bonnet tensor as components.

Playing with Topological Terms

The three topological terms in 4D are unsuited as action forms, because they would constitute surface terms (they are locally exact). However, as we are aware of the discussion of boundary terms in the main text, they may be suitable to serve different (partly interdependent) purposes. A boundary term

1. may be necessary in order to allow the definition of a functional integration (Regge/Teitelboim);
2. must be included into the action in order to cancel terms due to non-trivial boundary conditions (Gibbons, Hawking, York);
3. can serve as the generator of a canonical transformation;
4. signals a non-trivial topology–this is the case if the corresponding densities are not globally exact.

As for item 4.) above, there is the mathematical notion of cohomology classes which relates integrals of the topological terms to topologies of the manifold. And with regard to item 2.), a simple line of reasoning reveals how in terms of differential forms boundary terms are to be handled: Start from the Hilbert-Einstein Lagrangian (assuming the curvature being defined with respect to the Levi-Civita connection) and rewrite it as

$$\begin{aligned}\Omega^{IJ} \wedge \eta_{IJ} &= (d\omega^{IJ} + \omega^{I}{}_{K} \wedge \omega^{KJ}) \wedge \eta_{IJ} \\ &= d(\omega^{IJ} \wedge \eta_{IJ}) + \omega^{IJ} \wedge d\eta_{IJ} + \omega^{I}{}_{K} \wedge \omega^{KJ} \wedge \eta_{IJ}.\end{aligned}$$

Since the combination $\Omega^{IJ} \wedge \eta_{IJ} - d(\omega^{IJ} \wedge \eta_{IJ})$ involves only first derivatives of the tetrads, it is reasonable to add to the GR action a term

$$-\frac{1}{2}\int_{M} d(\omega^{IJ} \wedge \eta_{IJ}) = -\frac{1}{2}\int_{\partial M} \omega^{IJ} \wedge \eta_{IJ}.$$

It was demonstrated by B.P. Dolan [131] that this term is identical to the surface term considered by Gibbons and Hawking [216] and by York [575], as described in Subsect. 7.5.2. Dolan also showed that the same term gives rise to the canonical transformation from the tetrad variables to Ashtekar's "new" variables (see C.3.3). The previous findings, which are valid only for the case of vanishing torsion were extended to the case with torsion. Here, it turned out that one needs to add (1/2 of) the Nieh-Yan term in order to recover the canonical transformations leading to Ashtekar's variables.

The most general action describing vacuum GR in first-order formalism and being compatible with diffeomorphism invariance and Lorentz invariance, can be written (following [441]):

$$S[\vartheta, \omega] = \int \alpha_1\, \eta_{IJ} \wedge \Omega^{IJ} + \alpha_2\, \vartheta^{IJ} \wedge \Omega_{IJ} + \alpha_3\, P_4 + \alpha_3\, E_4 + \alpha_5\, NY_4 + \alpha_6\, \eta.$$

The first term is the Hilbert-Palatini action, where α_1 is related to the gravitational constant. The first two terms together are proportional to the Holst action (C.111), and α_2 is proportional to the Barbero-Immirzi parameter. The terms with $\alpha_3, \alpha_4, \alpha_5$ are the Pontryagin, the Euler, and the Nieh-Yan topological densities in 4D, respectively. Finally, the α_6 term represents the cosmological constant.

Even and Odd Spacetime Dimensions

In the section on Poincaré gauge theories, I point out the fact that the Einstein-Cartan theory in 4D is not invariant with respect to translations. This can easily be exhibited in differential forms: For translations with an infinitesimal parameter ϵ^I the vierbeins and the curvature transform as

$$\delta\vartheta^I = D\epsilon^I = d\epsilon^I + \omega^I{}_K\epsilon^K \qquad \delta\Omega^{IJ} = 0 \quad \text{since} \quad \delta\omega^I{}_K = 0.$$

For the 4D Einstein-Cartan theory $\mathcal{L}_{(1,2)} = \epsilon_{IJKL}\,\vartheta^I \wedge \vartheta^J \wedge \Omega^{KL}$, we find

$$\begin{aligned}\delta\mathcal{L}_{(1,2)} &= 2\epsilon_{IJKL}(D\epsilon^I) \wedge \vartheta^J \wedge \Omega^{KL} \\ &= d(2\epsilon_{IJKL}\,\epsilon^I\vartheta^J \wedge \Omega^{KL}) + 2\epsilon_{IJKL}\,\epsilon^I D(\vartheta^J \wedge \Omega^{KL}) \\ &= d\,\left(2\epsilon_{IJKL}\epsilon^I\vartheta^J \wedge \Omega^{KL}\right) + 2\epsilon_{IJKL}\,\epsilon^I\Theta^J \wedge \Omega^{KL}\end{aligned}$$

where the Bianchi identity $D\Omega^{KL} = 0$ was used. This is an exact form only if the torsion vanishes. On the other hand, for the 3D Einstein-Cartan theory:

$$\delta\mathcal{L}_{(1,1)} = 2\,\epsilon_{IJK}(D\epsilon^I) \wedge \Omega^{KL} = d(2\epsilon_{IJK}\epsilon^I\Omega^{KL}).$$

These results can be applied to any dimension with the insight that the Einstein Lagrangian terms $\mathcal{L}_{(2n+1,1)}$ in Lanczos-Lovelock-Cartan gravity are translation invariant. For gravity and supergravity in odd dimensions, see the lecture notes [577].

F.4 Gauge Theories

In this section[12] we will restrict ourselves to the matter Lagrangian $\mathcal{L}_M$ in (F.3), that is we assume a fixed background gravitational field. The collection of matter fields, previously called Φ^A is now split in two sets, namely "particle fields" $\{\Psi^\alpha\}$ and "force fields" $\{A^a\}$. There might be different fields Ψ, each with their own index range, dictated by the gauge symmetry group $\mathbf{G}$ as explained below. In the following, we will deal only with one variant of Ψ fields, because the others may simply be added.

F.4.1 Global Symmetry

Assume that

$$S_P = \int \mathcal{L}_P(\Psi^\alpha, d\Psi^\alpha, \vartheta, \omega) \tag{F.75}$$

[12] The mathematics needed is outlined in Appendix E.5.

is invariant under infinitesimal Lie-group transformations. This means the following: Call the infinitesimal group parameter $\gamma = \gamma^a \otimes X_a = \gamma^a X_a$ where the X_a are generators of $\mathfrak{g}$, the Lie algebra associated to the symmetry group $\mathbf{G}$. The index a runs from 1 to the dimension of the group manifold. Thus γ is a Lie-algebra valued vector field. The p-form Ψ^α is considered as an element of $\Omega_p(M, V)$, where V is a representation space of $\mathbf{G}$ (and $\alpha = 1, \ldots, \dim V$). Thus $\Psi^\alpha \in \Omega(M, V) = \oplus\Omega_p(M, V)$. Observe that this has consequences for the Lagrangian form $\mathcal{L}$, which must be an element of $\Omega_D(M, R)$. Let for instance

$$\mathcal{L} = \Psi_p \wedge \tilde{\Psi}_{D-p} = \Psi_p^a V_a \wedge \tilde{\Psi}_{D-p}^b V_b := \Psi_p^a \wedge \tilde{\Psi}_{D-p}^b (V_a, V_b).$$

To obtain an element of $\Omega_D(M, R)$, it is necessary to have an inner product for the basis vectors in V:

$$(\ ,\ ) : V \times V \to R, \qquad \text{for example} \qquad (V_a, V_b) = \delta_{ab}.$$

A $\mathbf{G}$ gauge transformation does not affect the tetrad and the spin connection in (F.75). On the Ψ-fields they are acting as

$$\delta_G(\gamma)\Psi^\alpha = \hat{\gamma}^\alpha{}_\beta \Psi^\beta = (\gamma^a \hat{X}_a)^\alpha{}_\beta \Psi^\beta, \tag{F.76}$$

where $\hat{X}_a$ are the generators of $\mathbf{G}$ in the Ψ representation. Let us subsequently write $\delta_G(\gamma) = \delta_\gamma$. (Quasi)-invariance of (F.75) means that $\delta_\gamma \mathcal{L}_P = dF_\gamma$, which by virtue of (F.32) is equivalent to the existence of an on-shell exact form J_γ

$$J_\gamma := \delta_\gamma \Psi^\alpha \wedge P_\alpha - F_\gamma \qquad \text{with} \qquad dJ_\gamma \doteq 0. \tag{F.77}$$

In the following, just for simplification at this point, we assume invariance of the action (F.75), that is $F_\gamma \equiv 0$. We might distinguish global and local gauge invariance, the former characterized by constant parameters γ^a. In this case:

$$J_\gamma = \gamma^a \wedge j_a \qquad \text{with} \qquad j_a := (\hat{X}_a)^\alpha{}_\beta \Psi^\beta \wedge P_\alpha.$$

F.4.2 Local Gauge Transformations

The globally-invariant action (F.75) is turned into a locally-invariant action by replacing the "ordinary" derivative by the covariant derivative through the introduction of a gauge field A^a and by introducing a kinetic term for the gauge field. This new action is then

$$S_M = \int \mathcal{L}_M = \int \mathcal{L}_P(\Psi^\alpha, D\Psi^\alpha, \vartheta, \omega) + \mathcal{L}_F(A, dA, \vartheta). \tag{F.78}$$

The local symmetry with a spacetime dependent γ^a, together with

$$\delta_\gamma \Psi^\alpha = (\hat{\gamma})^\alpha{}_\beta \Psi^\beta \qquad \delta_\gamma A^a = -D\gamma^a$$

(according to (E.16, E.19)), converts (F.31) into

$$\begin{aligned} 0 \equiv \delta_\gamma \mathcal{L}_M =& \hat{\gamma}^\alpha{}_\beta \Psi^\beta \wedge \mathcal{E}_\alpha - D\gamma^a \wedge \mathcal{E}_a + dJ_\gamma, \\ J_\gamma =& \hat{\gamma}^\alpha{}_\beta \Psi^\beta \wedge P_\alpha - D\gamma^a \wedge P_a \end{aligned} \tag{F.79}$$

where

$$\mathcal{E}_\alpha = \frac{\delta \mathcal{L}_M}{\delta \Psi^\alpha} = \frac{\delta \mathcal{L}_P}{\delta \Psi^\alpha} \qquad\qquad \mathcal{E}_a = \frac{\delta \mathcal{L}_M}{\delta A^a} = \frac{\partial \mathcal{L}_P}{\partial A^a} + \frac{\delta \mathcal{L}_F}{\delta A^a} \tag{F.80}$$

$$P_\alpha = \frac{\partial \mathcal{L}_M}{\partial (d\Psi^\alpha)} = \frac{\partial \mathcal{L}_P}{\partial (D\Psi^\alpha)} \qquad P_a = \frac{\partial \mathcal{L}_M}{\partial (dA^a)} = \frac{\partial \mathcal{L}_F}{\partial (dA^a)}. \tag{F.81}$$

By reshuffling the D-derivatives, (F.79) becomes

$$\hat{\gamma}^\alpha{}_\beta \Psi^\beta \wedge \mathcal{E}_\alpha + \gamma^a D\mathcal{E}_a + d(\hat{\gamma}^\alpha{}_\beta \Psi^\beta \wedge P_\alpha - \gamma^a \mathcal{E}_a + \gamma^a DP_a) = 0.$$

As exerted previously, writing this in the form $\gamma^a A_a + d(\gamma^a B_a) = 0$, gives rise to the identities

$$A_a = (\hat{X}_a)^\alpha{}_\beta \Psi^\beta \wedge \mathcal{E}_\alpha + D\mathcal{E}_a = 0 \tag{F.82a}$$

$$B_a = (\hat{X}_a)^\alpha{}_\beta \Psi^\beta \wedge P_\alpha - \mathcal{E}_a + DP_a = 0. \tag{F.82b}$$

We guess from the known result of the Noether current resulting from the Lorentz symmetry that the Noether gauge current can be expressed in terms of the gauge field equations and a superpotential as

$$J_\gamma = \gamma^a \mathcal{E}_a - d(\gamma^a P_a)$$

and confirm this by the use of (F.82b).

Clearly (F.82a) is the Noether identity, constituting a relation among the Euler derivatives of $\mathcal{L}_M$. The other identity (F.82b) will be further analyzed, since it restricts $\mathcal{L}_F$: From the second expression in (F.80) one derives

$$\mathcal{E}_a = \frac{D\Psi^\alpha}{\partial A^a} \wedge \frac{\partial \mathcal{L}_P}{D\Psi^\alpha} + \frac{\delta \mathcal{L}_F}{\delta A^a} = (\hat{X}_a)^\alpha{}_\beta \Psi^\beta \wedge P_\alpha + \frac{\delta \mathcal{L}_F}{\delta A^a}.$$

Therefore (F.82b) boils down to a condition on $\mathcal{L}_F$,

$$-\frac{\partial \mathcal{L}_F}{\partial A^a} - d\frac{\partial \mathcal{L}_F}{\partial (dA^a)} + D\frac{\partial \mathcal{L}_F}{\partial (dA^a)} \stackrel{!}{=} 0,$$

or, making use of (E.18),

$$-\frac{\partial \mathcal{L}_F}{\partial A^a} + f_{bc}{}^a A^b \wedge \frac{\partial \mathcal{L}_F}{\partial A^c} \stackrel{!}{=} 0.$$

The solution of this differential equation reveals that the dependence of $\mathcal{L}_F$ on the gauge fields A^a and their derivatives is only through the field strength $F^a := dA^a + \frac{1}{2} f_{bc}{}^a A^b \wedge A^c$.

A further restriction on $\mathcal{L}_F$ stems from the requirement that it should be a scalar with respect to the gauge group **G**. A natural choice is

$$\mathcal{L}_F = g_{ab} F^a \wedge * F^b$$

where g_{ab} is an invariant metric in **G**.

Group Covariant Diffeomorphisms

The group properties of the gauge transformations are characterized by the commutators of infinitesimal group transformations. We have

$$[\delta_G(\gamma), \delta_G(\gamma')] = \delta_G([\gamma, \gamma']), \tag{F.83}$$

where the commutator is taken in the algebra of **G**. In analogy to the case of local Lorentz transformations (F.49) the commutator of group transformations and diffeomorphisms is

$$[\delta_D(\xi), \delta_G(\gamma)] = \delta_G(i_\xi d\gamma). \tag{F.84}$$

As with the "mixture" of diffeomorphisms with the Lorentz transformations we may define a modified diffeomorphism. Its form can be derived by observing that

$$£_\xi A = i_\xi dA + d i_\xi A = i_\xi \big(F - \frac{1}{2}(A \wedge A)\big) + \big(D i_\xi A - A \wedge i_\xi A\big) = i_\xi F + D(i_\xi A),$$

where in this case the covariant derivative is defined with respect to the group connection A. Notice that the last term is nothing but a gauge transformation. The modified diffeomorphism is thus reasonably defined as

$$\overset{G}{£}_\xi := \delta_D(\xi) + \delta_G(i_\xi A) \tag{F.85}$$

in terms of a modified Lie derivative

$$\overset{G}{£}_\xi := i_\xi \overset{G}{D} + \overset{G}{D} i_\xi.$$

And again, as in the mixture of diffeomorphisms with Lorentz transformations, one derives

$$
\begin{aligned}
[i_X, \overset{G}{\pounds}_Y]\Phi &= i_{[X,Y]}\Phi \\
\left[\overset{G}{\pounds}_Y, \overset{G}{D}\right]\Phi &= i_X F \wedge \Phi \\
\left[\overset{G}{\pounds}_X, \overset{G}{\pounds}_Y\right]\Phi &= \overset{G}{\pounds}_{[X,Y]}\Phi + i_Y i_X F \wedge \Phi.
\end{aligned}
$$

For the modified diffeomorphism $\overset{G}{\delta}_D(\xi) = \overset{G}{\pounds}_\xi$, it holds that

$$
[\overset{G}{\pounds}_\xi, \delta_G(\lambda)] = \delta_G(\overset{G}{\pounds}_\xi \lambda) \qquad \left[\overset{G}{\pounds}_\xi, \overset{G}{\pounds}_\xi\right] = \overset{G}{\pounds}_{[\xi,\xi']} + \delta_G(i_{\xi'} i_\xi F).
$$

References

1. Abbott, L.F., Deser, S.: Stability of gravity with a cosmological constant. Nucl. Phys. **B195**, 76–96 (1982)
2. Abbott, L.F., Deser, S.: Charge Definition in Non-Abelian Gauge Theories. Phys. Lett. **116B**, 259–263 (1982)
3. Aldrovandi, R.R., Pereira, J.G.: Teleparallel Gravity—An Introduction. Springer, Dordrecht (2012)
4. Aldrovandi, R., Beltrán Almeida, J.P., Pereira, J.G. Class: de Sitter special relativity. Quantum. Grav. **24**, 1385–1404 (2007)
5. Alvarez-Gaumé, Baulieu L.: The two quantum symmetries associated with a classical symmetry. Nucl. Phys. **B212**, 255–267 (1983)
6. Ambjorn, J., Jurkiewicz, J., Loll, R.: The Self-organized de Sitter universe. Int. J. Mod. Phys. **D17**, 2515–2520 (2009)
7. Anderson, J.L.: Covariance, Invariance, and Equivalence—A Viewpoint. Gen. Rel. Grav. **2**, 161–172 (1971)
8. Anderson, P.W.: More Is Different—Broken symmetry and the nature of the hierarchical structure of science. Science. **177**(4047), 393–396 (1972)
9. Anderson, J.L., Bergmann, P.G.: Constraints In Covariant Field Theories. Phys. Rev. **83**, 1018–1025 (1951)
10. de Andrade, V.C., Pereira, J.G.: Gravitational Lorentz force and the description of the gravitational interaction. Phys. Rev. **D56**, 4689–4659 (1997)
11. Appelquist, Th., Chodos, A., Freund, P.: Modern Kaluza Klein Theories. Addison-Wesley, Reading (1987)
12. Aragone, C., Deser, S.: Consistency Problems of Hypergravity. Phys. Lett. **B86**, 161–163 (1979)
13. Arkani-Hamed, N., Dimopoulos, S., Dvali, G.: Phenomenology, Astrophysics and Cosmology of Theories with Sub-Millimeter Dimensions and TeV Scale Quantum Gravity. Phys. Rev. **D59**, 086004 (1999)
14. Arnol'd, V.I.: Mathematical Methods of Classical Mechanics, 2nd edn. Springer-Verlag, Heidelberg (1989)
15. Arnowitt, R., Deser, S., Misner, C.W.: The dynamics of general relativity. In: Witten, L. (ed.) Gravitation: An Introduction to Current Research, p. 227. Wiley, New York (1958), slightly "retouched" by the authors in arXiv:gr-qc/0405109; Canonical Variables for General Relativity. Phys. Rev. **117**, 1595–1602 (1960)

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

16. Ashtekar, A.: New variables for classical and quantum gravity. Phys. Rev. Lett. **57**, 2244–2247 (1986). New Hamiltonian formulation of general relativity. Phys. Rev. **D36**, 1587–1602 (1987)
17. Atkinson, D., Johnson, P.W.: Current and constituent quark masses: Beyond chiral-symmetry breaking. Phys. Rev. **D**, 1661–1666 (1990)
18. Bacry, H., Lévy-Leblond, J.-M.: Possible Kinematics. J. Math. Phys. **9**, 1605–1614 (1968)
19. Bacry, H., Nuyts, J.: Classification of ten-dimensional kinematical groups with space isotropy. J. Math. Phys. **27**, 2455–2457 (1986)
20. Bailin, D., Love, A.: Kaluza-Klein theories. Rep. Prog. Phys. **50**, 1087–1170 (1987)
21. Bak, D., Cangemi, D., Jackiw, R.: Energy-momentum conservation in gravity theories. Phys. Rev. **D49**, 5173–5181 (1994). Erratum: Phys. Rev. **D52**, 3753 (1995)
22. Bakker, B.L.G.: Forms of relativistic dynamics. In: Latal, H., Schweiger, W. (eds.) Methods of Quantization. Lecture Notes in Physics, vol. 572. Springer-Verlag, Berlin (2001)
23. Banerjee, N., Banerjee, R., Ghosh, S.: Quantisation of Second Class Systems in the Batalin-Tyutin Formalism. Ann. Phys. **241**, 237–257 (1995)
24. Barbashov, B.M., Nesterenko, V.V.: Continuous Symmetries in Field Theory. Fortschr. Phys. **31**, 535–567 (1983)
25. Barbour, J.: The End of Time. Oxford University Press (1999)
26. Barbour, J.B., Bertotti, B.: Mach's principle and the structure of dynamical theories. Proc. R. Soc. Lond. **A382**, 295–306 (1982)
27. Bargmann, V.: On Unitary Ray Representations of Continuous Groups. Ann. Math. **59**, 1–46 (1954)
28. Bargmann, V.: Note on Wigner's Theorem on Symmetry Operations. J. Math. Phys. **5**, 862–868 (1964)
29. Bargmann, V., Wigner, E.: Group Theoretical Discussion of Relativistic Wave Equations. Proc. Natl. Acad. Sci. U.S.A. **34**, 211 (1948)
30. Barrow, J.D.: Dimensionality. Phil. Trans. R. Soc. Lond. **A310**, 337–346 (1983)
31. Barrow, J.D.: New Theories of Everything—The Quest for Ultimate Explanation. Oxford University Press, Oxford (2007)
32. Barut, A.O., Böhm, A.: Dynamical groups and mass formulae. Phys. Rev. **B139**, 1107–1112 (1965)
33. Batlle, C., Gomis, J., Pons, J.M., Román-Roy, N.: Equivalence between the Lagrangian and Hamiltonian formalisms for constrained systems. J. Math. Phys. **27**, 2953–2962 (1986)
34. Batlle, C., Gomis, J., Gràcia, X., Pons, J.M.: Noether's theorem and gauge transformations: Application to the bosonic string and CP_2^{n-1}model. J. Math. Phys. **30**, 1345–1350 (1989)
35. Batterman, R.W.: Falling cats, parallel parking, and polarized light. Stud. Hist. Philos. Sci. B: Stud. Hist. Philos. Mod. Phys. **34**, 527–557 (2003)
36. Baulieu, L.: Perturbative Gauge Theories. Phys. Rep. **129**, 1–74 (1985)
37. Bekaert, X., Boulanger, N.: The unitary representations of the Poincaré group in any spacetime dimension. arXiv:hep-th/0611263
38. Bern, Z., Blau, S.K., Mottola, E.: General covariance of the path integral for quantum gravity. Phys. Rev. **D43**, 1212–1222 (1991)
39. Bertlmann, R.A.: Anomalies in Quantum Field Theory. Clarendon Press, Oxford (1996)
40. Bessel-Hagen, E.: Über die Erhaltungssätze der Elektrodynamik. Math. Ann. **34**, 258–276 (1921)
41. Beig, R., Murchadha, Ó.N.: The Poincaré group as the symmetry group of canonical general relativity. Ann. Phys. (N.Y.) **174**, 463–498 (1987)
42. Belinfante, F.J.: On the spin angular momentum of mesons. Physica. **6**, 887–898 (1939)
43. Bergmann, P.: Introduction to the Theory of Relativity. Prentice-Hall, New York (1942), several editions, corrected and enlarged 1976
44. Bergmann, P.G., Komar, A.: The Coordinate Group Symmetries of General Relativity. Int. J. Theor. Phys. **5**, 15–28 (1972)
45. Beringer, J., et al., (Particle Data Group): Review of Particle Physics. Phys. Rev. **D86**, 010001 (2012). http://pdg.lbl.gov/

46. Bern, Z., Carrasco, J.J., Dixon, L.J., Johansson, H., Roiban, R.: Ultraviolet Behavior of $N = 8$ Supergravity at Four Loops. Phys. Rev. Lett. **103**, 081301 (2009)
47. Binétruy, P.: Supersymmetry—Theory. Experiment and Cosmology. Oxford University Press, Oxford (2006)
48. Blagojević, M.: Hamiltonian structure and gauge symmetries of Poincaré gauge theory. Ann. Phys. (Leipzig) **10**, 367–391 (2001)
49. Blagojević, M.: Gravitation and Gauge Symmetries. IOP Publishing, Bristol (2002); Three lectures on Poincaré gauge theory. arXiv:gr-qc/0302040
50. Blagojević, M., Hehl, F.W.: Gauge Theories of Gravitation—A Reader with Commentaries. Imperial College Press, London (2013)
51. Blagojević, M., Vasilić, M.: Hamiltonian analysis of extra gauge symmetries in an $R + T^2$ theory of gravity. Phys. Rev. **D34**, 357–366 (1986)
52. Bluhm, R.: Overview of the SME: Implications and Phenomenology of Lorentz Violation, in [150], pp. 199–236
53. Böhm, M., Denner, A., Joos, H.: Gauge Theories of the Strong and Electroweak Interaction. Teubner, Stuttgart (2001)
54. Bouatta, N., Compère, G., Sagnotti, A.: An Introduction to Free Higher-Spin Fields. arXiv:hep-th/0409068
55. Brading, K.: Which symmetry? Noether, Weyl and conservation of electric charge. Stud. Hist. Philos. Mod. Phys. **33**, 3–22 (2002)
56. Brading, K., Brown, H.R.: Noether's Theorems and Gauge Symmetry. arXiv:hep-th/0009058
57. Brading, K., Castellani, E. (eds.): Symmetries in Physics—Philosophical Reflections. Cambrige University Press (2003)
58. Brown, H.R., Holl, P.R.: The Galilean covariance of quantum mechanics in the case of external fields. Am. J. Phys. **67**, 204–214 (1999)
59. Brown, J.D.: Black hole entropy and the Hamiltonian formulation of diffeomorphism invariant theories. Phys. Rev. **D52**, 7011–7026 (1995)
60. Brown, J.D., Lau, S.R., York, J.W.: Action and Energy of the Gravitational Field. Ann. Phys. **297**, 175–218 (2002)
61. Brown, J.D., Kuchař, K.V.: Dust as a standard of space and time in canonical quantum gravity. Phys. Rev. **D51**, 5600–5629 (1995)
62. Buchbinder, I.L., Kuzenko, S.M.: Ideas and Methods of Supersymmetry and Supergravity -or- A Walk Through Superspace. Institute of Physics Publishing, Bristol (1995)
63. Burgess, C. F.: Quantum Gravity in Everyday Life: General Relativity as an Effective Field Theory, Living Reviews in Relativity, lrr-2004-5. Max Planck Society (2004)
64. Cacciatori, S., Gorini, V., Kamenshchik, A.: Special relativity in the 21st century. Ann. Phys. **17**, 728–768 (2008)
65. Cahn, R.N.: The eighteen arbitrary parameters of the standard model in your everyday life. Rev. Mod. Phys. **68**, 951–959 (1996)
66. Callan, C.G., Coleman, S., Jackiw, R.: A New Improved Energy-Momentum Tensor. Ann. Phys. **59**, 42–73 (1970)
67. Candotti, E., Palmieri, C., Vitale, B.: On the Inversion of Noether's Theorem in Classical Dynamical Systems. Am. J. Phys. **40**, 424–429 (1972)
68. Cao, T.Y.: New Philosophy of Renormalization: From the Renormalization Group Equations to Effective Field Theories. In: Brown, L.M. (ed.) Renormalization—From Lorentz to Landau (and Beyond). Springer, New York (1993)
69. Cao, T.Y.: Conceptual developments of 20th century field theories. Cambridge University Press (1997)
70. Cao, T. Y. (ed.): Conceptual Foundations of Quantum Field Theory. Proc. Confererence (1966), Boston, Cambridge University Press (1999)
71. Capovilla, R., Dell, J., Jacobson, T.: A pure spin-connection formulation of gravity. Class. Quantum Grav. **8**, 59–73 (1991)
72. Capozziello, S., Faraon, V.: Beyond Einstein Gravity, Fundamental Theories of Physics, vol. 170. Springer (2011)

73. Cariñena, J.F., del Olmo, M.A., Santander, M.: Kinematics groups and dimensional analysis. J. Phys. A: Math. Gen. **14**, 1–14 (1981)
74. Cariñena, J.F.: Theory of Singular Lagrangians. Fortschr. Phys. **38**, 641–679 (1990)
75. Carmeli, M.: Classical Fields—General Relativity and Gauge Theory. Wiley, New York (1982)
76. Carr, B. (ed.): Universe or Multiverse? Cambridge University Press (2007)
77. Carroll, L.: Through the looking-glass and what Alice found there. Macmillan, London (1871)
78. Carroll, S.M.: Spacetime and Geometry—An Introduction to General Relativity. Addison-Wesley, San Francisco (2004)
79. Cartan, E.: Sur une généralisation de la notion de courbure de Riemann et les espaces a torsion. Comptes rendus de l'Académie de sciences Paris **174**, 593 (1922)
80. Casalbuoni, R.: On the quantization of systems with anticommuting variables. Nuovo Cim. **A33**, 115–24 (1976); The classical mechanics for Bose-Fermi systems. Nuovo Cim. **A33**, 389–431 (1976)
81. Castellani, L.: Symmetries of constrained Hamiltonian systems. Ann. Phys. **143**, 357–371 (1982)
82. Chadha, S., Nielsen, H.B.: Lorentz Invariance as a Low Energy Phenomenon. Nucl. Phys. **B217**, 125–144 (1983)
83. Chaichian, M., Hagedorn, R.: Symmetries in Quantum Mechanics. Institute of Physics Publishing, Bristol (1989)
84. Chaichian, M., Martinez, D.L.: On the Noether identities for a class of systems with singular Lagrangians. J. Math. Phys. **35**, 6536–6545 (1994)
85. Chamseddine, A.H.: Applications of the Gauge Principle to Gravitational Interactions. Int. J. Geom. Methods Mod. Phys. **3**, 149–176 (2006)
86. Chamseddine, A.H., West, P.C.: Supergravity as a gauge theory of supersymmetry. Nucl. Phys. **B129**, 39–44 (1977); reprinted in [50]
87. Charap, J.M., Henneaux, M., Nelson, J.E.: Explicit form of constraint algebra in tetrad gravity. Class. Quantum Grav. **5**, 1405–1414 (1988)
88. Chang, C.C., Nester, J.M., Chen, C.M.: Pseudotensors and Quasilocal Energy-Momentum. Phys. Rev. Lett. **83**, 1879–1901 (1999)
89. Chen, C.M., Nester, J.M.: Quasilocal quantities for general relativity and other gravity theories. Class. Quantum Grav. **16**, 1279–1304 (1999)
90. Chen, C.M., Nester, J.M., Yo, H.W.: Acausal PGT modes and the nonlinear constraint effect. Acta Phys. Polonica **B29**, 961–970 (1998); reprinted in [50]
91. Chkareuli, J.L., Froggatt, C.D., Nielsen, H. B.: Lorentz Invariance and Origin of Symmetries. Phys. Rev. Lett. **87**, 091601 (2001); Deriving gauge symmetry and spontaneous Lorentz violation. Nucl. Phys. **B821**, 65–73 (2009); Spontaneously generated tensor field gravity. Nucl. Phys. **B848**, 498–522 (2011)
92. Cho, Y.M., Freund, P.G.O.: Non-Abelian gauge fields as Nambu-Goldstone fields. Phys. Rev. **D12**, 1711–1720 (1975)
93. Clifton, T., Ferreira, P.G., Padilla, A., Skordis, C.: Modified gravity and cosmology. Phys. Rep. **513**, 1–189 (2012)
94. Cohen, A.G., Glashow, S.L.: Very Special Relativity. Phys. Rev. Lett. **97**, 021601 (2006)
95. Coleman, S.: The Invariance of the Vacuum is the Invariance of the World. J. Math. Phys. **7**, 787 (1966)
96. Coleman, S.: Aspects of Symmetry—Selected Erice Lectures. Cambridge University Press (1985)
97. Coleman, S., Mandula, J.: All possible symmetries of the S matrix. Phys. Rev. **159**, 1251–1256 (1967)
98. Collins, P.D.M., Martin, A.D., Squires, E.J.: Particle Physics and Cosmology. Wiley, New York (1989)
99. Corry, L.: David Hilbert and the Axiomatization of Physics (1898–1918): From 'Grundlagen Der Geometrie' to 'Grundlagen der Physik'. Kluwer, Dordrecht (2004)
100. Corry, L., Renn, J., Stachel, J.: Belated Decision in the Hilbert-Einstein Priority Dispute. Science **278**, 1270–1273 (1997)

101. Corvin, L., Ne'eman, Y., Sternberg, S.: Graded Lie algebras in mathematics and physics (Bose-Fermi symmetry). Rev. Mod. Phys. **47**, 573–603 (1975)
102. Costa, G., Fogli, G.: Symmetries and Group Theory in Particle Physics—An Introduction to Space-Time and Internal Symmetries, Lecture Notes in Physics, vol. 823, Berlin (2012)
103. Costa, M.E.V., Girotti, H.O., Simões, T.J.M.: Dynamics of gauge systems and Dirac's conjecture. Phys. Rev. **D32**, 405–410 (1984)
104. Cremmer, E., Julia, B.: The $N = 8$ Supergravity Theory. I. The Lagrangian. Phys. Lett. **80B**, 48–51 (1978); The SO(8) Supergravity. Nucl. Phys. **B153**, 141–212 (1979)
105. Cremmer, E., Julia, B., Scherk, J.: Supergravity Theory in 11 Dimensions. Phys. Lett. **76B**, 409–412 (1978)
106. Cvetič, M., Gibbons, G.W., Lü, H., Pope, C.N.: Consistent group and coset reductions of the bosonic string. Class. Quantum Grav. **20**, 5161–5193 (2003)
107. Dadhich, N., Pons, J.M.: Consistent Levi-Civita truncation uniquely characterizes the Lovelock Lagrangians. Phys. Lett. **B705**, 139–142 (2011)
108. Dadhich, N., Pons, J.M.: On the equivalence of the Einstein-Hilbert and the Einstein-Palatini formulations of general relativity for an arbitrary connection. Gen. Rel. Grav. **44**, 2337–2352 (2012)
109. Damour, T.: Experimental Tests of Gravitational Theory in [45]
110. Damour, T.: What is missing from Minkowski's 'Raum und Zeit' lecture. Ann. Phys. **17**, 619–630 (2008)
111. Damour, T., Deser, S., McCarthy, J.: Nonsymmetric gravity theories: Inconsistencies and a cure. Phys. Rev. **D47**, 1541–1556 (1993)
112. Davies, P.C.W.: Particles do not exist. In: Christensen, S.M. (ed.) Quantum Theory of Gravity: Essays in Honor of the 60th Birthday of Brice DeWitt, pp. 66–77. Adam Hilger, Bristol (1984)
113. Deser, S.: Self-Interaction and Gauge Invariance. Gen. Rel. Grav. **1**, 9 (1970), reprinted (and updated) as arXiv:gr-qc/0411023
114. Deser, S.: Gravity from self-interaction redux. Gen. Rel. Grav. **42**, 641–646 (2010)
115. Deser, S., Zumino, B.: Consistent Supergravity. Phys. Lett. **B62**, 335–337 (1976), reprinted in [176]
116. D'Souza, I.A., Kalman, C.S.: Preons—Models of Leptons. World Scientific, Quarks and Gauge Bosons as Composite Objects (1992)
117. Deutsch, D.: On Wheeler's notion of 'Law without Law' in physics. Found. Phys. **16**, 565–572 (1986)
118. de Wit. B., Nicolai, H.: $N = 8$ supergravity. Nucl. Phys. **B208**, 323–364 (1982)
119. de Wit, B., Nicolai, H.: The consistency of the S^7 truncation in $d = 11$ supergravity. Nucl. Phys. **B281**, 211–240 (1987)
120. DeWitt, B.S.: Dynamical Theory of Groups and Fields. Blackie and Son Ltd., London (1965)
121. DeWitt, B.S.: Quantum Theory of Gravity. I. The Canonical Theory. Phys. Rev. **160**, 1113–1148 (1967); Quantum Theory of Gravity. II. The Manifestly Covariant Theory. Phys. Rev. **162**, 1195–1239; Quantum Theory of Gravity. III. Applications of the Covariant Theory, Phys. Rev. **162**, 1239–1256
122. DeWitt, B.S.: Supermanifolds, 2nd edn. Cambridge University Press, Cambridge (1992)
123. Dine, M.: Supersymmetry and String Theory. Cambridge University Press, Cambridge (2007)
124. Dirac, P.A.M.: The Principles of Quantum Mechanics, 4th edn. Oxford University Press (1st edn. 1930, since 3rd edn. 1947 bra-ket notation) (1988)
125. Dirac, P.A.M.: Relativistic Wave Equations. Proc. Roy. Soc. **A155**, 447–459 (1936)
126. Dirac, P.A.M.: Forms of Relativistic Dynamics. Rev. Mod. Phys. **21**, 392–399 (1949)
127. Dirac, P.A.M.: Generalized Hamiltonian dynamics. Canad. J. Math. **2**, 129–148 (1950)
128. Dirac, P.A.M.: The theory of gravitation in Hamiltonian form. Proc. Royal Soc. **A246**, 333 (1958); Fixation of Coordinates in the Hamiltonian Theory of Gravitation. Phys. Rev. **114**, 924–930 (1959)
129. Dirac, P.A.M.: Lectures on Quantum Mechanics. Dover, Mineola (2001); originally published by the Belfer Graduate School of Science, Yeshiva University, New York (1964)

130. Dobado, A., Gómez-Nicola, A., Maroto, A.L., Peláez, J.R.: Effective Lagrangians for the Standard Model. Springer, Berlin (1997)
131. Dolan, B.P.: On the generating function for Ashtekar's canonical transformation. Phys. Lett. **B233**, 89–91 (1989)
132. Dolan, L., Duff, M.J.: Kac-Moody-Symmetries of Kaluza-Klein Theories. Phys. Rev. Lett. **52**, 14–17 (1984)
133. Doncel, M.G., Hermann, A., Michel, L., Pais, A. (eds.): Symmetries in Physics (1600–1980). Seminar d'història de les Ciències. Universitat Autònoma de Barcelona, Bellaterra (1987)
134. Dosch, H.G.: The Standard Model of Particle, Physics, in [471], pp. 21–50
135. Dosch, H.G., Müller, V.F.: The facets of relativistic quantum field theory. Eur. Phys. J. **H35**, 331–375 (2010)
136. Doughty, N.A.: Lagrangian Interactions. Addison-Wesley (1990)
137. Dreiner, H.B., Haber, H.E., Martin, S.P.: Two-component spinor techniques and Feynman rules for quantum field theory and supersymmetry. Phys. Rep. **494**, 1–196 (2010)
138. Duff, M.J.: Kaluza-Klein Theory in Perspective, in Lindström, Ulf (ed.): Proceedings of the Symposium 'The Oskar Klein Centenary'. World Scientific, Singapore, pp. 22–35, also as arXiv:hep-th/9410046
139. Duff, M.J.: Supermembranes, arXiv:hep-th/9611203, based on lectures given at the T. A. S. I. Summer School. University of Colorado, Boulder, June 1996; the Topical Meeting, Imperial College, London, July 1996 and the 26th British Universities Summer School in Theoretical Elementary Particle Physics, University of Swansea, September 1996
140. Duff, M.J. (ed.): The World in Eleven Dimensions—Supergravity. Supermembranes and M-theory. Institute of Physics Publishing, Bristol (1999)
141. Duff, M.J., Nilsson, B.E.W., Pope, C.N.: Kaluza-Klein Supergravity. Phys. Rep. **130**, 1–142 (1986)
142. Duff, M.J., Nilsson, B.E.W., Pope, C.N., Warner, N.P.: On the Consistency of the Kaluza-Klein Ansatz. Phys. Lett. **149B**, 90–94
143. Dyson, F.J.: Missed Opportunities. Bull. Am. Math. Soc. **78**, 635–652 (1972)
144. Earman, J.: Laws, Symmetry, and Symmetry Breaking—Invariance, Conservation Principles and Objectivity. Address to the 2002 meeting of the Philosophy of Science Association.
145. Eder, G.: Quantenmechanik I. Bibliographisches Institut, Mannheim (1986)
146. Eguchi, T., Gilkey, P.B., Hanson, A.J.: Gravitation, Gauge Theories and Differential Geometry. Phys. Rep. **66**, 213–393 (1980)
147. Ehlers, J.: The nature and structure of spacetime. In: Mehra, J., et al. (eds.) The Physicist's Conception of Nature, pp. 71–91 (1973)
148. Ehlers, J.: On limit relations between and approximative explanations of, physical theories. In: Barcan Marcus, R., Dorn, G.J.W., Weingartner, P. (eds.): Logic, methodology and philosophy of science, pp. 387–403. North Holl (1986)
149. Ehlers, J.: General Relativity, in [471], pp. 91–104
150. Ehlers, J., Lämmerzahl, C. (eds.): Special Relativity: Will it survive the next 101 years? Lecture Notes in Physics 702. Springer, Berlin (2006)
151. Einstein, A.: Zur Elektrodynamik bewegter Körper. Ann. Phys. **17**, 891–921 (1905); reprinted in [438], translated in [348]
152. Einstein, A.: Die Feldgleichungen der Gravitation, Sitzungsber. d. Kgl. Preuss. Akad. der Wissensch., 844–847, 25. November 1915; translation: The Field Equations of Gravitation. In: CPAE (Collected Papers of Albert Einstein), vol. 6, 117–120, translator: A. Engel
153. Einstein, A.: HAMILTONsches Prinzip und allgemeine Relativitätstheorie, Sitzungsber. d. Kgl. Preuss. Akad. der Wissensch., 1111–1116, 26. Oktober 1916; translated as: Hamilton's Principle and the General Theory of Relativity. In CPAE (Collected Papers of Albert Einstein), vol. 6, 240–246, translator: A. Engel
154. Einstein, A., Die Grundlage der allgemeinen Relativitätstheorie. Ann. Phys. **49**, 769–822 (1916), reprinted in [438], translated in [348]
155. Einstein, A.: Relativity—The Special and General Theory. Henry Holt, New York (1920), translation of: Relativitätstheorie—Nach dem heutigen Stande gemeinverständlich dargestellt. Friedr. Vieweg und Sohn, Braunschweig (1916)

156. Einstein, A.: Der Energiesatz in der allgemeinen Relativitätstheorie, Sitzungsber. Kgl. Preuss. Akad. Wissensch. 448–459, 16. Mai 1918; translation: The Law of Energy Conservation. In CPAE (Collected Papers of Albert Einstein), vol. 7, 47–61, translator: A. Engel
157. Einstein, A.: Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle?, Sitzungsber. Kgl. Preuss. Akad. Wissensch. 349–356 (1919), translated as: Do gravitational fields play an essential part in the structure of elementary particles of matter?, in [348]
158. Einstein, A.: Einheitliche Feldtheorie von Gravitation und Elektrizität, Sitzungsber. Kgl. Preuss. Akad. Wissensch., 414–419, 1925; Unified Field Theory of Gravitation and Electricity. In: Unzicker. A., Case, T.: Translation of Einstein's Attempt of a Unified Field Theory with Teleparallelism, arXiv:physics/0503046
159. Einstein, A.: Zu Kaluza's Theorie des Zusammenhanges von Gravitation und Elektrizität, Sitzungsber. Kgl. Preuss. Akad. Wissensch., 23–30, (1927)
160. Einstein, A.: Über den gegenwärtigen Stand der Feld-Theorie. In: Honegger, E. (ed.), Festschrift zum 70. Geburtstag von Prof. Dr. A. Stodola, pp. 126–132. Orell Füssli, Leipzig (1929)
161. Einstein, A.: Considerations concerning the fundaments of theoretical physics. Science **91**, 487–492 (1940); in German: Das Fundament der Physik. In: Aus meinen späteren Jahren, Ullstein, Frankfurt (1984)
162. Einstein, A.: The Meaning of Relativity, 5th edn. Princeton University Press (1955)
163. Elliot, J.P., Dawber, P.G.: Symmetry in Physics. Mcmillan, London (1979)
164. Ellis, G.F.R.: Physics ain't what it used to be. Nature **438**, 739–740 (2005)
165. Ellis, G.F.R.: Physics and the real world. Found. Phy. **36**, 227–262 (2006)
166. Emmerson, J.L.: Symmetry Principles in Particle Physics. Clarendon Press, Oxford (1972)
167. Esposito, G., Marmo, G., Sudarshan, G.: From Classical to Quantum Mechanics. Cambridge University Press (2004)
168. Exirifard, Q., Sheikh-Jabbari, M.M.: Lovelock gravity at the Crossroads of Palatini and Metric formulations. Phys. Lett. **B661**, 158–161 (2008)
169. Faddeev, L.D.: The Feynman integral for singular Lagrangians. Theor. Math. Phys. **1**, 1–13 (1970)
170. Faddeev, L.D., Jackiw, R.: Hamiltonian Reduction of Unconstrained and Constrained Systems. Phys. Rev. Lett. **60**, 1692–1694 (1988)
171. Faddeev, L.D., Popov, V.N.: Feynman diagrams for the Yang-Mills field. Phys. Lett. **25B**, 29–30 (1967)
172. Falkenburg, B.: Particle Metaphysics: A Critical Account of Subatomic Reality. Springer, Heidelberg (2007)
173. Fang, J., Fronsdal, C.: Massless fields with half-integral spin. Phys. Rev. **D18**, 3630–3633 (1978)
174. Fatibene, L., Ferraris, M., Francaviglia, M., Lusanna, L.: ADM pseudotensors, conserved quantities and covariant conservation laws in general relativity. Ann. Phys. **327**, 1593–1616 (2012)
175. Fayet, P.: Supersymmetry and weak, electromagnetic and strong interactions. Phys. Lett. **64B**, 159–163 (1976); Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions. Phys. Lett. **69B**, 489–494 (1977)
176. Ferrara, S. (ed.): Supersymmetry. North Holland/World Scientific, Amsterdam (1987)
177. Ferrara, S., Zumino, B.: Supergauge invariant Yang-Mills theories. Nucl. Phys. **B79**, 413–421 (1974), reprinted in [176]
178. Ferrara, S., Freedman, D.S., van Nieuwenhuizen, P., Breitenlohner, P., Gliozzi, F., Scherk, J.: Scalar multiplet coupled to supergravity. Phys. Rev. **D15**, 1013–1018 (1977)
179. Ferrara S., van Nieuwenhuizen, P.: Supergravity: An Odyssey through Space-Time and Superspace, in [267], 557–585
180. Feynman, R.P.: Space-Time approach to Non-Relativistic Quantum Mechanics. Rev. Mod. Phys. **20**, 367–387 (1948); Feynman, R.P., Hibbs, A. R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965)

181. Feynman, R.P., Leighton, R.B., Sands, M.: The Feynman Lectures on Physics, Mainly Mechanics, Radiation and Heat, vol. 1. Addison-Wesley, Reading (1963)
182. Feynman, R.P.: Quantum Theory of Gravitation. Acta Phys. Polonica **XXIV**, 697–722 (1963)
183. Feynman, R.P., Morinigo, F.B., Wagner, W.G.: Feynman Lectures on Gravitation. Addison-Wesley (1995)
184. Feynman, R.P.: The Character of Physical Law. The M.I.T. Press, Cambridge (1965)
185. Fierz, M.: Über die relativistische Theorie kräftefreier Teilchen mit beliebigem Spin. Helv. Phys. Acta **12**, 3–37 (1939)
186. Fierz, M., Pauli, W.: On relativistic wave equations for particles of arbitrary spin in an electromagnetic field. Proc. Roy. Soc. London **A173**, 211–232 (1939)
187. Fisher, M.E.: The renormalization group in the theory of critical behavior. Rev. Mod. Phys. **46**, 597–616 (1974)
188. Fölsing, A.: Albert Einstein—A Biography. Penguin (1998)
189. Foerster, D., Nielsen, H.B., Ninomiya, M.: Dynamical stability of local gauge symmetry (creation of light from chaos). Phys. Lett. **94B**, 135–140 (1980)
190. Fonda, L., Ghirardi, G.C.: Symmetry Principles in Quantum Physics. Dekker, New York (1970)
191. Forger, M., Rõmer, H.: Currents and the energy-momentum tensor in classical field theory: A fresh look at an old problem. Ann. Phys. **309**, 306–389 (2004)
192. Frankel, T.H.: The Geometry of Physics. Cambridge University Press (1997, 2004)
193. Fredenhagen, K., Rehren K.-H., Seiler, E.: Quantum Field Theory—Where We Are, in [471], pp. 61–87
194. Freedman, D.Z., van Nieuwenhuizen, P., Ferrara, S.: Progress towards a theory of supergravity. Phys. Rev. **D13**, 3214–3218 (1976), reprinted in [176]
195. French, A.P. (ed.): Einstein: A Centenary Volume. Harvard University Press, Cambridge (1979)
196. Freund, P.G.O., Kaplansky, I.: Simple supersymmetries. J. Math. Phys. **17**, 228–230 (1976)
197. Freund, P.G.O., Rubin, M.: Dynamics of dimensional reduction. Phys. Lett. **B97**, 233–235 (1980)
198. Friedman, M.: Foundations of Space-Time Theories. Princeton University Press (1983)
199. Fritzsch, H., Gell-Mann, M., Leutwyler, H.: Advantages of the Color Octet Gluon Picture. Phys. Lett. B. **47**, 365–368 (1973)
200. Fritzsch, H., Minkowski, P.: Unified interactions of leptons and hadrons. Ann. Phys. **93**, 193–266 (1975)
201. Frogatt, C.D., Nielsen, H.B.: Origin of Symmetries. World Scientific, Singapore (1991)
202. Fuchs, J., Schweigert, Ch.: Symmetry, Lie Algebras and Representation. Cambridge University Press (1997)
203. Fujii, Y., Maeda, K.-I.: The Scalar-Tensor Theory of Gravitation. Cambridge University Press, Cambridge (2003)
204. Fujikawa, K.: Path-integral measure for gauge-invariant fermion theories. Phys. Rev. Lett. **42**, 1195–1198 (1979); Fujikawa K., Suzuki, H.: Path Integrals and Quantum Anomalies, Clarendon, Oxford (2004)
205. Fulton, T., Rohrlich, F., Witten, L.: Conformal invariance in physics. Rev. Mod. Phys. **34**, 442–457 (1962)
206. Fushchich, W.I., Nikitin, A.G.: Symmetries of Maxwell's Equations. D. Reidel Publishing Company, Dordrecht (1987); The complete sets of conservation laws for the electromagnetic field. J. Phys. A: Math. Gen. **25**, L231–L233 (1992)
207. Fushchich, W.I., Shtelen, W.M., Serov, N.I.: Symmetry Analysis and Exact Solutions of Equations of Nonlinear Mathematical Physics. Kluwer, Dordrecht (1993); Fushchich, W.I., Nikitin, A.G.: Symmetries of Equations of Quantum Mechanics. Allerton, New York (1994)
208. Galilei, G.: Dialogo sopra i due massimi sistemi, Florence, 1632; translated by S. Drake: Dialogue Concerning the Two Chief World Systems. University of California Press, Berkeley (1953, revised 1967)

209. García, J.A., Pons J.M.: Equivalence of Faddeev-Jackiw and Dirac approaches for gauge theories. Int. J. Mod. Phys. **A12**, 451–464 (1997); Faddeev-Jackiw approach to gauge theories and ineffective constraints. Int. J. Mod. Phys. **A13**, 3691–3710 (1998)
210. Georgi, H.: Effective Field Theory. Ann. Rev. Nuc. Part. Sci. **43**, 209–252 (1993)
211. Georgi, H.: Lie Algebras in Particle Physics—From Isospin to Unified Theories, 2nd edn. Westview (1999)
212. Georgi, H.: Weak Interactions and Modern Particle Theory. Dover Publications (2009), revised version of the original 1984 edition
213. Georgi, H., Glashow, S.L.: Gauge theories without anomalies. Phys. Rev. **D6**, 429–431 (1972)
214. Georgi, H., Glashow, S.L.: Unity of All Elementary-Particle Forces. Phys. Rev. Lett. **32**, 438–441 (1974)
215. Georgi, H., Quinn, H.R., Weinberg, S.: Hierarchy of interactions in unified gauge theories. Phys. Rev. Lett. **33**, 451–454 (1974)
216. Gibbons, G.W., Hawking, S.W.: Action integrals and partition functions in quantum gravity. Phys. Rev. **D15**, 2752–2756 (1977)
217. Gibbons, G.W., Patricot, C.E.: Newton-Hooke spacetimes, Hpp-waves and the cosmological constant. Class. Quantum Grav. **20**, 5225–5239 (2003)
218. Gitman, D.M., Tyutin, I.V.: Quantization of Fields with Constraints. Springer, Berlin (1990)
219. Giulini, D.: On Galilei Invariance in Quantum Mechanics and the Bargmann Superselection Rule. Ann. Phys. **249**, 222–235 (1996)
220. Giulini, D.: Algebraic and Geometric Structures of Special Relativity, in [150], pp. 45–111
221. Giulini, D.: Remarks on the Notions of General Covariance and Background Independence, in [471], pp. 105–120
222. Giulini, D.: Concepts of Symmetry in the Work of Wolfgang Pauli. arXiv:0802.4341 [physics.hist-ph]
223. Giulini, D.: Superselection Rules. In: Greenberger, D., et al. (eds.) Compendium of Quantum Physics: Concepts, Experiments, History, and Philosophy. Springer, Berlin (2009)
224. Giulini, D.: Equivalence principle, quantum mechanics, and atom-interferometer tests. arXiv:1105.0749v2 [gr-qc]
225. Giulini, D., Kiefer, C., Zeh, H.D.: Symmetries, superselection rule, and decoherence. Phys. Lett. **A199**, 291–298 (1995)
226. Glashow, S.L., Iliopoulos, J., Maiani, L.: Weak Interactions with Lepton-Hadron Symmetry. Phys. Rev. **D2**, 1285–1292 (1970)
227. Goenner, H.: Unified Field Theories: From Eddington and Einstein up to Now. In: de Sabbata, V., Karade, T.M. (eds.) Proceedings of the Sir Arthur Eddington Centenary Symposium, Relativistic Astrophysics and Cosmology, vol. 1, pp. 176–196. World Scientific, Singapore (1984)
228. Goenner, H., Wünsch, D.: Kaluza's and Klein's contributions to Kaluza-Klein-theory. Max-Planck-Institute for the History of Science, **235** (2003)
229. Goldberg, J.N.: Conservation laws in general relativity. Phys. Rev. **111**, 315–320 (1958)
230. Goldstein, H., Poole, C.P., Safko, J.L.: Classical Mechanics, 3rd edn. Addison-Wesley (2001)
231. Gomes H., Koslowski T.: Frequently Asked Questions About Shape Dynamics. Found. Phys. **43**, 1428–1458 (2013)
232. Gomis, J., Paris, J., Samuel, S.: Antibracket, antifields and gauge-theory quantization. Phys. Rep. **259**, 1–145 (1995)
233. Gomis, J., Henneaux, M., Pons, J.M.: Existence theorem for gauge symmetries in Hamiltonian constrained systems. Class. Quantum Grav. **7**, 1089–1096 (1990)
234. Gotay, M.J., Nester, J.M., Hinds, G.: Presymplectic manifolds and the Dirac Bergmann theory of constraints. J. Math. Phys. **19**, 2388–2399 (1978)
235. Gourdin, M.: Lagrangian Formalism and Symmetry Laws. Gordon and Breach, New York (1969)
236. Gourgoulhon, É.: 3+1 Formalism in General Relativity—Bases of Numerical Relativity. Lecture Notes in Physics 846, Springer, Heidelberg (2012)

237. Gràcia, X., Pons, J.M.: Gauge generators, Dirac's conjecture, and degrees of freedom for constrained systems. Ann. Phys. **187**, 355–368 (1988)
238. Gràcia, X., Pons, J.M.: A Hamiltonian approach to Lagrangian Noether transformations. J. Phys. A: Math. Gen. **25**, 6357–6369 (1992)
239. Greenberg, O.W., Greenberger, D.M., Greenbergest, T.V.: (Para) Bosons, (Para) Fermions, Quons and Other Beasts in the Menagerie of Particle Statistics, arXiv:hep-ph/9306225; Greenberg, O.W.: Spin-Statistics Connection and Commutation Relations. AIP Proc. **545**, 113–127 (2004)
240. Green, M.B., Schwarz, J.H., Witten, E.: Superstring Theory. Cambridge University Press, Cambridge (1987)
241. Greene, B.: The Elegant Universe—Superstring, Hidden Dimensions, and the Quest for the Ultimate Theory. W.W. Norton & Company (1999); The Fabric of the Cosmos—Time, Space, and the Texture of Reality. Alfred A. Knopf, New York (2004)
242. Grössing, G. (ed.): Emergent Quantum Mechanics 2011 (EmerQuM 11; Heinz von Foerster Congress, Vienna, Austria). J. Phys. Conf. Series **361** (2012)
243. Gronwald, F., Hehl, F.W.: On the gauge aspects of gravity. In: Bergmann, P.G., de Sabbata, V., Treder, H.J. (eds.) Quantum Gravity, Proceedings International School of Cosmology and Gravitation: 14th Course, Erice, 11–19 May 1995, arXiv:gr-qc/9602013
244. Gross, D.J.: The Triumph and the Limitations of Quantum Field Theory, in [70]. arXiv:hep-th/9704139
245. Gross, D.J.: Gauge Theory—Past, Present, Future. In: Liu C.S., Yau, S.-T. (eds.) Chen Ning Yang: A Great Physicist of the Twentieth Century. International Press of Boston (1995)
246. Gross, D.J.: Symmetry in Physics: Wigner's Legacy. Phys. Today **48**, 46–50 (1995)
247. Gross, D.J.: The role of symmetry in fundamental physics. Proc. Natl. Acad. Sci. U.S.A. **93**, 14256–14259 (1996)
248. Gruber, B., Biedenharn, L.C., Doebner, H.D.: Symmetries in Science V. Plenum Press, New York (1991)
249. Gürsey, F.: Introduction to the De Sitter Group. In: Gürsey, F. (ed.) Group Theoretical Concepts and Methods in Elementary Particle Physics. Gordon and Breach, New York (1965)
250. Guttmann, Y., Lyre, H.: Fiber Bundle Gauge Theories and 'Fields Dilemma'. arXiv:physics/0005051 [physics.hist-ph]
251. Haag, R., Łopuszanski, J., Sohnius, M.: All possible generators of supersymmetries of the S-matrix. Nucl. Phys. **B88**, 257 (1975), reprinted in [176]
252. Haber, H.E., Kane, G.L.: The search for supersymmetry—Probing physics beyond the standard model. Phys. Rep. **117**, 75–263 (1985)
253. Hagen, C.R.: Scale and conformal transformations in Galilean-covariant field theory. Phys. Rev. **D5**, 377–388 (1972)
254. Hamermesh, M.: Group Theory and its Application to Physical Problems. Dover, New York (1989)
255. Havas, P.: The connection between conservation laws and invariance groups—Folklore and fiction and fact. Acta Phys. Austriaca **38**, 145–167 (1973)
256. Hawking, S.W.: Is the end in sight for theoretical physics? Inaugural address as Lucasian Professor, Cambridge University 1979. In: Black Holes and Baby Universes and Other Essays. Bantam Dell (1994)
257. Hawking, S.W.: Gödel and The End of Physics. http://www.damtp.cam.ac.uk/events/strings02/dirac/hawking/dirac/hawking/
258. Hawking, S.W., Horowitz, G.T.: The gravitational Hamiltonian, action, entropy, and surface terms. Class. Quantum Grav. **13**, 1487–1498 (1996)
259. Hayashi, K., Shirafuji, T.: Gravity from the Poincaré gauge theory of fundamental interactions, I. General formulation, II. Equations of motion for test bodies and various limits, III. Weak field approximation, IV. Mass and energy of particle spectrum. Prog. Theor. Phys. **64**, 866, 883, 1435, 2222 (1980)
260. Hebecker, A., Wetterich, C.: Spinor gravity. Phys. Lett. **B574**, 269–275 (2003)

261. Hecht, R.D., Lemke, J., Wallner, R.P.: Can Poincaré gauge theory be saved? Phys. Rev. **D44**, 2442–2451 (1991); Hecht, R.D., Nester, J.M., Zhytnikov, V.V.: Some Poincaré gauge theory Lagrangians with well-posed initial value problems. Phys. Lett. **A222**, 37–42 (1996)
262. Hehl, F.W.: Spin and Torsion in General Relativity: I. Foundations. Gen. Rel. Grav. **4**, 333–349 (1973): II. Geometry and field equations. Gen. Rel. Grav. **5**, 491–516 (1974)
263. Hehl, F.W., von der Heyde, P., Kerlick, G.D., Nester, J.M.: General relativity with spin and torsion: Foundations and prospects. Rev. Mod. Phys. **48**, 393–416 (1976)
264. Hehl, F.W., Nitsch, J., von der Heyde, P.: Gravitation and Poincaré gauge field theory with quadratic Lagrangian, and in [267], vol. 1, 329–355
265. Hehl, F.W., McCrea, J.D., Mielke, E.W., Ne'eman, Y.: Metric affine gauge theory of gravity: Field equations, Noether identities, world spinors, and breaking dilation invariance. Phys. Rep. **258**, 1–171 (1995)
266. Hehl, F.W., Obukhov, Y.N.: Foundations of Classical Electrodynamics—Charge, Flux, and Metric. Birkhäuser, Boston (2003); Spacetime metric from local and linear electrodynamics: a new axiomatic scheme. In [150]; also as arXiv:gr-qc/0508024
267. Held, A. (ed.): General Relativity and Gravitation—One Hundred Years After the Birth of Albert Einstein. Plenum, New York (1980)
268. Henneaux, M.: Hamiltonian form of the path integral for theories with gauge freedom. Phys. Rep. **126**, 1–66 (1985)
269. Henneaux, M., Nelson, J.E., Schomblond, C.: Derivation of Ashtekar variables from tetrad gravity. Phys. Rev. **D39**, 434–437 (1989)
270. Henneaux, M., Teitelboim, C.: Relativistic quantum mechanics of supersymmetric particles. Ann. Phys. **143**, 127–159 (1982)
271. Henneaux, M., Teitelboim, C.: Hamiltonian treatment of asymptotically anti-de Sitter spaces. Phys. Lett. **142B**, 355–358 (1984); Asymptotically anti-de Sitter spaces. Commun. Math. Phys. **98**, 391–424 (1985)
272. Henneaux, M., Teitelboim, C.: The cosmological constant and general covariance. Phys. Lett. **B222**, 195–199 (1989)
273. Henneaux, M., Teitelboim, C.: Quantization of Gauge Systems. Princeton University Press, Princeton (1992)
274. Henneaux, M., Teitelboim, C., Zanelli, J.: Gauge invariance and degree of freedom count. Nucl. Phys. **B332**, 169–188 (1990)
275. Hilbert, D.: Die Grundlagen der Physik (Erste Mitteilung). Göttinger Nachrichten. Math.-Phys. Klasse, pp. 395–407 (1915); translation as: The foundations of physics (proofs of first communication). In: Renn, J., Schemmel M. (eds.) The Genesis of General Relativity: vol. 4. Gravitation in the Twilight of Classical Physics: The promise of mathematics, pp. 989–1001, Springer, Dordrecht (2007)
276. Hinterbichler, K.: Theoretical aspects of massive gravity. Rev. Mod. Phys. **84**, 671–710 (2012)
277. Hoddeson, L., Brown, L., Riordan, M., Dresden, M. (eds.): The Rise of the Standard Model—A History of Particle Physics from 1964 to 1979. Cambridge University Press (1997)
278. Hojman, S.A., Kuchař, K., Teitelboim, C.: Geometrodynamics regained. Ann. Phys. **96**, 88–135 (1976)
279. Hojman, S.A., Shepley, L.C.: No Lagrangian? No Quantization!. J. Math. Phys. **32**, 142–146 (1991)
280. Holst, S.: Barbero's Hamiltonian derived from a generalized Hilbert-Palatini action. Phys. Rev. **D53**, 5966–5969 (1996)
281. Horowitz, G.T., Strominger, A.: Witten's expression for gravitational energy. Phys. Rev. **D27**, 2793–2804 (1983)
282. Howard, D., Stachel, J. (eds.): Einstein and the History of General Relativity. Birkhäuser (1989)
283. Huang, K.: Quarks. Leptons and Gauge Fields. World Scientific, Singapore (1992)
284. Ibragimov, N.H.: Elementary Lie Group Analysis and Ordinary Differential Equations. Wiley, Chichester (1999)

285. Ibragimov, N.H., Kara, A.H., Mahomed, F.M.: Lie-Bäcklund and Noether symmetries with applications. Nonlinear Dyn. **15**, 115–136 (1998)
286. Iliopoulos, J., Nanopoulos, D.V., Tomaras, T.N.: Infrared stability or anti-unification. Phys. Lett. **94B**, 141–144 (1980)
287. Imiłkowska, K., Kowalski-Glikman, J.: Doubly Special Relativity as a Limit of Gravity, in [150], pp. 279–298
288. Inönü, E.: Contraction of Lie groups and their representations. In: Gürsey, F. (ed.) Group Theoretical Concepts and Methods in Elementary Particle Physics, pp. 391–402. Gordon, New York (1965)
289. Inönü, E., Wigner, E.: On the contraction of groups and their representations. Proc. Natl. Acad. Sci. U.S.A. **39**, 6510–6524 (1953)
290. Isenberg, J., Nester, J.: Canonical Gravity, in [267], vol. 1, pp. 23–97
291. Isham, C.: Lectures on Quantum Theory—Mathematical and Structural Foundations. Imperial College Press, London (1995)
292. Isham, C.: Modern Differential Geometry for Physicists, 2nd edn. World Scientific, Singapore (1999)
293. Itzykson, C., Zuber, B.: Quantum Field Theory. McGraw-Hill, New York (1980)
294. Iyer, V., Wald, R.M.: Some properties of the Noether charge and a proposal for dynamic black hole entropy. Phys. Rev. **D50**, 846–864 (1994); Comparison of the Noether charge and Euclidean methods for computing the entropy of stationary black holes. Phys. Rev. **D52**, 4430–4439 (1995)
295. Jackiw, R., Manton, N.S.: Symmetries and Conservation Laws in Gauge Theories. Ann. Phys. **127**, 257–273 (1980)
296. Jackiw, R., Pi, S.-Y.: Chern-Simons modification of general relativity. Phys. Rev. **D68**, 104012 (2003)
297. Jackiw, R., Pi, S.-Y.: Tutorial on scale and conformal symmetries in diverse dimensions. J. Phys. A: Math. Theor. **44**, 223001 (2011)
298. Jackson, J.D.: Classical Electrodynamics. Wiley, New York (1962, 1988)
299. Jacobson, T.: Thermodynamics of spacetime—The Einstein equation of state. Phys. Rev. Lett. **75**, 1260–1263 (1995)
300. Jahn, O., Sreedhar, V.V.: The maximal invariance group of Newton's equations for a free point particle. Am. J. Phys. **69**, 1039–1043 (2001)
301. Janssen, T., Prokopec, T.: Problems and hopes in nonsymmetric gravity. J. Phys. A: Math. Theor. **40**, 7067–7074 (2007)
302. Jauch, J.M.: Foundations of Quantum Mechanics. Addison-Wesley (1986)
303. Jizba, P., Pons, J.M.: Revisiting the gauge principle: Enforcing constants of motion as constraints. J. Phys. A: Math. Theor. **43**, 205202 (2010)
304. Jones, H.F.: Groups. Representations and Physics. Hilger, Bristol (1990)
305. José, J.V., Saletan, E.J.: Classical Dynamics—A Contemporary Approach. Cambridge University Press, Cambridge (1998)
306. Julia, B., Silva, S.: Currents and superpotentials in classical gauge-invariant theories: I. Local results with applications to perfect fluids and general relativity. Class. Quantum Grav. **15**, 2173–2215 (1998); Currents and superpotentials in classical gauge theories: II. Global aspects and the example of affine gravity. Class. Quantum Grav. **17**, 4733–4743 (2000)
307. Kaku, M.: Quantum Field Theory. Oxford University Press (1994)
308. Kaku, M.: Introduction to Superstrings and M-Theory, 2nd edn. Springer, New York (1999)
309. Kaluza, T.: Zum Unitätsproblem der Physik. Sitzungsber. Preussische Akademie der Wissenschaften 966–972 (1921). Translated as: On the unification problem of physics, in [11] and in [429]
310. Kastrup, H.: On the advancements of conformal transformations and their associated symmetries in geometry and theoretical physics. Annalen der Physik **17**, 631–690 (2008)
311. Kastrup, H.: The Contribution of Emmy Noether, Felix Klein and Sophus Lie to the Modern Concept of Symmetries in Physical Systems, in [133], pp. 113–163
312. Katz, J.: A note on Komar's anomalous factor. Class. Quantum Grav. **2**, 423–425 (1985)

313. Katz, J., et al.: Relativistic conservation laws and integral constraints for large cosmological perturbations. Phys. Rev. **D55**, 59–57 (1997)
314. Kayser, B., et al.: The Physics of Massive Neutrinos. World Scientific, Singapore (1989)
315. Kibble, T.W.B.: Lorentz invariance and the gravitational field. J. Math. Phys. **2**, 212–221 (1961)
316. Kiefer, C.: Quantum Gravity, 3rd edn. Clarendon, Oxford (2012)
317. Killing, W.: Die Zusammensetzung der stetigen endlichen Transformationsgruppen. Math. Ann. **31**(2), 252–290 (1888); **33**(1), 1–48 (1888); **34**(1) (1889); **36**(2), 161–189 (1890)
318. Klauder, J. R.: Quantization of Constrained Systems. In: Latal, H., Schweiger, W. (eds.) Methods of Quantization. Lecture Notes in Physics 572. Springer-Verlag, Berlin (2001); Coherent state quantization of constrained systems. Ann. Phys. 254, 419–453 (1997)
319. Klein, F.: Vergleichende Betrachtungen über neuere geometrische Forschungen. Programm zum Eintritt in die philosophische Fakultät und den Senat der Friedrich-Alexander-Universität zu Erlangen. Deichert, Erlangen (1872); reprinted and provided with notes from Klein. In: Gesammelte Mathematische Abhandlungen, vol. I, 460–497, Springer, Berlin (1921–1923)
320. Klein, F.: Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie. Nachrichten d. Königl. Gesellsch. d. Wissensch. zu Göttingen, Math.-phys. Klasse, Issue II, pp. 171–189 (1918); reprinted. In: Gesammelte Mathematische Abhandlungen. vol. 1, pp. 568–585. Springer, Berlin (1921)
321. Klein, O.: Quantentheorie und fünfdimensionale Relativitätstheorie. Z. Phys. **37**, 895–906 (1926)
322. Klein, O.: The atomicity of electricity as a quantum theory law. Nature **118**, 5–16 (1926)
323. Kobayashi, M., Maskawa, T.: CP-violation in the renormalizable theory of weak interaction. Prog. Theor. Phys. **49**(2), 652–657 (1973)
324. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry. Wiley Interscience (1996)
325. Kopczynski, W.: Variational principles for gravity and fluids. Annals Phys. **203**, 308–338 (1990)
326. Komar, A.: Covariant conservation laws in general relativity. Phys. Rev. **113**, 934–936 (1959)
327. Kosmann-Schwarzbach, Y.: The Noether Theorems—Invariance and Conservation Laws in the Twentieth Century. Springer-Verlag (2011)
328. Kraus, P., Griffiths, D.J.: Renormalization of a model quantum field theory. Am. J. Phys. **60**, 1013–1022 (1992)
329. Kunstatter, G.: The path integral for gauge theories—A geometrical approach. Class. Quantum. Grav. **9**, 157–168 (1992)
330. Ladyman, B.: Structural Realism. Stanford Encyclopedia of Philosophy (2009)
331. Lanczos, C.: The Variational Principles of Mechanics. University of Toronto Press 1949 (4th ed. 1974, Dover Paperback)
332. Landau, L.D., Lifschitz, E.M.: Mechanics, Course of Theoretical Physics, vol. 1, 3rd edn. Butterworth-Heinemann
333. Landau, L.D., Lifschitz, E.M.: The Classical Theory of Fields, Course of Theoretical Physics, vol. 2, 4th edn. Butterworth-Heinemann
334. Laughlin, R.B.: A Different Universe—Reinventing Physics from the Bottom Down. Basic Books, New York (2005)
335. Leader E., Lorcé, C.: The angular momentum controversy: What's it all about and does it matter?. Phys. Reports, in press (March 2014); available as arXiv: 1309.4235 [hep-ph]
336. Lederman, L.M., Hill, ChT: Symmetry and the Beautiful Universe. Prometheus Books, Amherst (2004)
337. Lee, A.R., Kalotas, T.M.: Lorentz transformations from the first postulate. Am. J. Phys. **43**, 434 (1975)
338. Lee, J., Wald, R.W.: Local symmetries and constraints. J. Math. Phys. **31**, 725–743 (1990)
339. Leutwyler, H., Stern, J.: Relativistic dynamics on a null plane. Ann. Phys. **112**, 94–164 (1978)
340. Lévy-Leblond, J.-M.: Galilei group and nonrelativistic quantum mechanics. J. Math. Phys. **4**, 776–788 (1963)

341. Lévy-Leblond, J.-M.: Une nouvelle limite non-relativiste du groupe de Poincaré. Ann. Insti. Henri Poincaré **III**, 1–12 (1965)
342. Lévy-Leblond, J.-M.: One more derivation of the Lorentz transformation. Am. J. Phys. **44**, 271 (1976)
343. Lie, S., Engel, F.: Theorie der Transformationsgruppen, Teubner, Leipzig, Three Volumes, 1888–1893, reprinted by Chelsea Publ. Co, New York (1970)
344. Lloyd, S.: Programming the Universe. Alfred A. Knopf, New York (2006)
345. Loebbert, F.: The Weinberg-Witten theorem on massless particles: An essay. Ann. Phys. **17**, 803–829 (2008)
346. Loll, R.: The emergence of spacetime or quantum gravity on your desktop. Class. Quantum Grav. **25**, 114006 (2008)
347. Łopuszanski, J.: An Introduction to Symmetry and Supersymmetry in Quantum Field Theory. World Scientific, Singapore (1991)
348. Lorentz, H.A., Einstein, A., Minkowski, H., Weyl, H.: The Principle of Relativity. Dover (1952). (reprint of a 1923 edition)
349. Louis, J., Mohaupt, T., Theisen, S.: String Theory—An Overview, [in 471], pp. 289–323
350. Lovelock, D.: The Einstein tensor and its generalization. J. Math. Phys. **12**, 498–501 (1971)
351. Luciani, J.F.: Spacetime geometry and symmetry breaking. Nucl. Phys. **B135**, 111–130 (1978)
352. Lüst, D., Theisen, S.: Lectures on string theory. Lect. Notes in Phys. **346**, Springer-Verlag, Berlin (1989)
353. Lusanna, L.: An enlarged phase space for finite-dimensional constrained systems, unifying their Lagrangian, phase- and velocity-space descriptions. Phys. Rep. **185**, 1–54 (1989); The second Noether theorem as the basis of the theory of singular Lagrangians and Hamiltonians constraints. Riv. Nuovo Cim. **14**, 1–75 (1991)
354. Lyre, H.: Lokale Symmetrien und Wirklichkeit. Mentis, Paderborn (2004)
355. Lyre, H.: Does the Higgs Mechanism Exist? arXiv:0806.1359 [physics.gen-ph]
356. MacDowell, S.W., Mansouri, F.: Unified geometric theory of gravity and supergravity. Phys. Rev. Lett. **38**, 739–742 (1977) (reprinted in [50])
357. Mahomed, F.M., Kara, A.H., Leach, P.G.L.: Lie and Noether counting theorems for one-dimensional systems. J. Math. Analy. App. **178**, 116–129 (1993)
358. Maldacena, J.: The gauge/gravity duality. In: Horowitz, G. (ed.) Black Holes in Higher Dimensions. Cambridge University Press (2012)
359. Mannel, T.H.: Effective Field Theories in Flavour Physics. Springer Tracts in Modern Physics 203. Springer, Berlin (2004)
360. Manohar, A.V., Wise, M.B.: Heavy Quark Physics. Cambridge University Press, Cambridge (2000)
361. Mardones, A., Zanelli, J.: Lovelock-Cartan theory of gravity. Class. Quantum Grav. **8**, 1545–1558 (1991)
362. Martin S. P.: A Supersymmetry Primer. In: Kane, G.L. (ed.) Perspectives on Supersymmetry II. World Scientific, Singapore (2010); also available as arXiv:hep-ph/9709356
363. Marshak R.E.: Conceptual Foundations of Modern Particle Physics. World Scientific (1993)
364. Martin B.R., Shaw, G.: Particle Physics, 3rd edn. Wiley (2008)
365. Mattingly, D.: Modern Tests of Lorentz Invariance. Living Reviews in Relativity, llr-2005-5
366. Maxwell, J.C.: On physical lines of force. Philos. Maga. (Fourth Series), **21**, 161–175, 281–291, 338–348 (1861); **23**, 12–24, 85–95 (1862)
367. Mermin, N.D.: Relativity without light. Am. J. Phys. **52**, 119–124 (1984)
368. Mittelstaedt, P., Weingartner, P.: Laws of Nature. Springer (2005)
369. Mohapatra, R.N.: Unification and Supersymmetry—The Frontiers of Quark-Lepton Physics, 3rd edn. Springer, New York (2003)
370. Mohapatra, R.N., Pal, P.B.: Massive Neutrinos in Physics and Astrophysics. World Scientific, Singapore (1991)
371. Morrissey, D.E., Plehn, T.H., Tait, T.M.P.: Physics searches at the LHC. Phys. Rep. **515**, 1–113 (2012)

372. Minkowski, H.: Raum und Zeit. Address to the 80th Assembly of German Natural Scientists and Physicians (Cologne/Köln, Sept 21, 1908), Engl. Translation in [348]
373. Misner, S.W., Thorne, K.S., Wheeler, J.A.: Gravitation. Freeman, San Francisco (1973)
374. Mukhanov, V., Wipf, A.: On the symmetries of Hamiltonian systems. Int. J. Mod. Phys. **A10**, 579–610 (1995)
375. Mukhopadhyay, A., Padmanabhan, T.: Holography of gravitational action functionals. Phys. Rev. **D74**, 124023 (2006)
376. Mukunda, N.: Symmetries and constraints in generalized Hamiltonian dynamics. Ann. Phys. **99**, 408–433 (1976); Generators for symmetry transformations of constrained Hamiltonian systems. Phys. Scripta **21**, 783–791 (1980)
377. Muñoz, G.: Lagrangian field theories and energy-momentum tensors. Am. J. Phys. **64**, 1153–1157 (1996)
378. Nahm, W.: Supersymmetries and their representations. Nucl. Phys. **B135**, 149–166 (1978)
379. Nash C., Sen, S.: Topology and Geometry for Physicists. Academic Press, London (1983); reprinted by Dover Public (2011)
380. Ne'eman, Y., Šijački, Dj.: SL(4, R) world spinors and gravity. Phys. Lett. **B157**, 275–279 (1985); Šijački Dj.: World spinors revisited. Acta Phys. Polonica **B29**, 1089–1097 (1998)
381. Neffe, J.: Einstein—A Biography. Farrar, Straus and Giroux, New York (2007)
382. Nester, J.M.: General pseudotensors and quasilocal quantities. Class. Quantum Grav. **21**, S261–S280 (2004)
383. Nester, J.M.: A manifestly covariant Hamiltonian formalism for dynamical geometry. Prog. Theor. Phys. Suppl. **172**, 30–39 (2008)
384. Nester, J.M., Meng, F.F., Chen, C.M.: Quasilocal center-of-mass. J. Korean Phys. Soc. **45**, S22–S25 (2004); arXiv:gr-qc/0403103
385. Newton, I.: Philosophiæ Naturalis Principia Mathematica. London, 1668; Isaac Newton: The Principia, Mathematical Principles of Natural Philosophy, a New Translation by I.B. Cohen, A. Whitman, together with A Guide to Newton's Principia by I.B. Cohen. University of California Press, (1999)
386. Ni, W-T.: Searches for the role of spin and polarization in gravity. Rep. Prog. Phys. **73**, 056901 (2010); reprinted in [48]
387. Nicolai, H.: From Grassmann to maximal ($N = 8$) supergravity. Ann. Phys. **19**, 150–160 (2010)
388. Nicolai, H., Schlotterer, O.: Infinite Dimensional Symmetries. www.itp.uni-hannover.de/lechtenf/Events/Lectures/nicolai.pdf (2009)
389. Niederer, U.: The maximal kinematical invariance group of the free Schrödinger equation. Helv. Phys. Acta **45**, 802–810 (1972)
390. Niederer, U.H., O'Raifeartaigh, L.: Realizations of the unitary representations of the inhomogeneous space-time groups II: Covariant realizations of the Poincaré group. Fortsch. Phys. **22**, 131–157 (1974)
391. Nielsen, H.B.: Field theories without fundamental gauge symmetries. Phil. Trans. R. Soc. Lond. **A310**, 261–272 (1983)
392. Nitsch, J.: The macroscopic limit of the Poincaré gauge theory of gravitation. In: Bergmann, P.G., de Sabbata, V. (eds.) Cosmology and Gravitation: Spin, Torsion. Rotation and Supergravity. Plenum, New York (1980)
393. Noether, E.: "Invariante Variationsprobleme", Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Math.-Phys. Klasse, 235–257 (1918); English translation by M.A. Tavel, Transp. Theory Stat. Phys. **1**(3), 183–207 (1971); available also as arXiv:physics/0503066
394. Norton, J.D.: What was Einstein's Principle of Equivalence?, in [282], pp. 5–47
395. Norton, J.D: General covariance and the foundations of general relativity: Eight decades of dispute. Rep. Prog. Phys. **56**, 791–858 (1993); General covariance, gauge theories, and the Kretschmann objection, in [57], pp. 110–123
396. Obukhov, Y.N., Rubilar, G.F.: Invariant conserved currents in gravity theories with local Lorentz and diffeomorphism symmetry. Phys. Rev. **D74**, 064002 (2006)

397. Obukhov, Y.N., Rubilar, G.F.: Covariance properties and regularization of conserved currents in tetrad gravity. Phys. Rev. **D73**, 124017 (2006); Obukhov Y.N., Rubilar G.F., Pereira, J.G.: Conserved currents in gravitational models with quasi-invariant Lagrangians—Application to teleparallel gravity. Phys. Rev. **D74**, 104007 (2006)
398. Ohnuki, Y.: Unitary Representations of the Poincare Group and Relativistic Wave Equations. World Scientific (1988)
399. Olver, P.J.: Applications of Lie Groups to Differential Equations, 2nd revised edn. Springer, New York (1986, 1993)
400. Overduin, J.M., Wesson, P.S.: Kaluza-Klein gravity. Phys. Rep. **283**, 303–378 (1997)
401. Padmanabhan, T.: Gravity: A new holographic perspective. Intern. J. Mod. Phys. **D15**, 1659–1676 (2006)
402. Padmanabhan, T.: From gravitons to gravity: Myths and reality. Intern. J. Mod. Phys. **D17**, 367–398 (2008)
403. Padmanabhan, T.: Thermodynamical aspects of gravity: New insights. Rep. Prog. Phys. **73**, 046901 (2010)
404. Padmanabhan, T.: Gravitation—Foundations and Frontiers. Cambridge University Press, Cambridge (2010)
405. Paschos, E.A.: Electroweak Theory. Cambridge University Press (2007)
406. Pauli, W.: Relativitätstheorie. Encyklopädie der mathematischen Wissenschaften, vol. 19, B.G. Teubner, Leipzig (1921); reprinted as: Theory of Relativity. Dover (1958)
407. Pauli, W.: On the connection between spin and statistics. Phys. Rev. **58**, 716–722 (1940)
408. Pauli, W., Weisskopf, V.: Über die Quantisierung der skalaren relativistischen Wellengleichung. Helv. Phys. Acta **7**, 709–731 (1934)
409. Peldán, P.: Actions for gravity, with generalizations: A review. Class. Quantum Grav. **11**, 1087–1132 (1994)
410. Penrose, R.: The Road to Reality. Alfred A. Knopf, New York (2005)
411. Penrose, R., Rindler, W.: Spinors and Space-Time. Cambridge University Press, Cambridge (1984)
412. Peskin, M.E., Schroeder, D.V.: An Introduction to Quantum Field Theory. Westview, Boulder (1995)
413. Petrov, A.N., Lompay, R.R.: Covariantized Noether identities and conservation laws for perturbations in metric theories of gravity. Gen. Rel. Grav. **45**, 545–579 (2013)
414. Picken, R.F., Sundermeyer, K.: Integration on supermanifolds and a generalized Cartan calculus. Comm. Math. Phys. **102**, 585–604 (1986)
415. Pickering, A.: Constructing Quarks: A Sociological History of Particle Physics. University of Chicago Press (1984)
416. Polchinski, J.: String Theory. Cambridge University. Press, Cambridge (1998)
417. Pons, J.M.: Plugging the gauge fixing into the Lagrangian. Int. J. Mod. Phys. **A11**, 975–988 (1996)
418. Pons, J.M.: On Dirac's incomplete analysis of gauge transformations. Stud. Hist. Philos. Mod. Phys. **36**, 491–518 (2005)
419. Pons, J.M.: Substituting fields within the action: Consistency issues and some applications. J. Math. Phys. **51**, 122903 (2010)
420. Pons, J.M.: Noether symmetries, energy-momentum tensors, and conformal invariance in classical field theory. J Math. Phys. **52**, 012904 (2011)
421. Pons, J.M., García, J.A.: Rigid and gauge Noether symmetries for constrained systems. Intern. J. Mod. Phys. **A15**, 4681–4721 (2000)
422. Pons, J.M., Salisbury, D.C.: The issue of time in generally covariant theories and the Komar-Bergmann approach to observables in general relativity. Phys. Rev. **D71**, 124012 (2005)
423. Pons, J.M., Salisbury, D.C., Shepley, L.C.: Gauge transformations in the Lagrangian and Hamiltonian formalisms of generally covariant systems. Phys. Rev. **D55**, 658–668 (1997)
424. Pons, J.M., Salisbury, D.C., Shepley, L.C.: Reduced phase space: Quotienting procedure for gauge theories. J. Math. Phys. A: Math. Gen. **32**, 419 (1999)

425. Pons, J.M., Salisbury, D.C., Shepley, L.C.: Gauge transformations in Einstein-Yang-Mills theories. J. Math. Phys. **41**, 5557–5571 (2000)
426. Pons, J.M., Salisbury, D.C., Sundermeyer, K.A.: Revisiting observables in generally covariant theories in the light of gauge fixing methods. Phys. Rev. **D80**, 084015 (2009); Observables in classical canonical gravity: Folklore demystified. J. Phys. Conf. Ser. **222**, 012018 (2010)
427. O'Raifeartaigh, L.: Unitary representations of Lie groups in quantum mechanics. In: Bargmann, V. (ed.) Group Representations in Mathematics and Physics. Springer, Berlin (1970)
428. O'Raifeartaigh, L.: Group Structure of Gauge Theories. Cambridge University Press, Cambridge (1986)
429. O'Raifeartaigh, L.: The Dawning of Gauge Theory. Princeton University Press (1997)
430. O'Raifeartaigh, L., Straumann, N.: Gauge theory: Historical origin and some modern developments. Rev. Mod. Phys. **72**, 1–23 (2000)
431. Ramond, P.: Field Theory—A Modern Primer, 4th edn. Addison-Wesley, Redwood City (1990)
432. Ramond, P.: Journeys Beyond the Standard Model. Perseus Books, Cambridge (1999)
433. Randall, L.: Warped Passages—Unraveling the Mysteries of the Universe's Hidden Dimensions. Harper Collins, New York (2005)
434. Rarita, W., Schwinger, J.: On a theory of particles with half-integral spin. Phys. Rev. **60**, 61 (1941)
435. Recami, E.: A hommage to E.C.G. Sudarshan: Superluminal objects and waves (An updated overview of the relevant experiments). arXiv:0804.1502
436. Rees, M.J.: Just Six Numbers—the Deep Forces that Shape the Universe. Weidenfeld, London (1999)
437. Regge, T., Teitelboim, C.: Improved Hamiltonian for general relativity. Phys. Lett. **53B**, 101–105 (1974); Role of surface integrals in the Hamiltonian formulation of general relativity. Ann. Phys. **88**, 286–318 (1974)
438. Renn, J. (ed.): Einstein's Annalen Papers, the Complete Collection 1901–1922. Wiley-VCH, Weinheim (2005)
439. Renn, J. (ed.): The Genesis of General Relativity. Springer-Verlag, Dordrecht (2007)
440. Renn, J., Stachel, J.: Hilbert's Foundation of Physics: From a Theory of Everything to a Constituent of General Relativity. Max Planck Institute for the History of Science. Preprint 118 (1999) ISSN 0948–9444
441. Rezende, D.J., Perez, A.: Four-dimensional Lorentzian Holst action with topological terms. Phys. Rev. **D79**, 064026 (2009)
442. Riemann, B.: Über die Hypothesen, welche der Geometrie zu Grunde liegen. Abh. Kgl. Ges. Wiss. Göttingen, Vol. 13 (1867). Translated by W.K. Clifford: On the Hypotheses which lie at the Bases of Geometry. Nature **8**(183, 184), 14–17, 36, 37
443. van Ritbergen, T., Vermaseren, A.M., Larin, S.A.: The four loop β-function in quantum chromodynamics. Phys. Lett. **B400**, 379–384 (1997)
444. Rodrigues, W.A., Vaz, J.: Spinors as differential forms, and applications to electromagnetism and quantum mechanics. In: Pronin, P., Sardanashvily, G. (eds.) Gravity, Particles and Space-Time, pp. 307–344. World Scientific, Singapore (1996)
445. Rose, M.: Introduction to Cosmology, 3rd edn. Wiley, Chichester (2003)
446. Rosen, G.: Galilean invariance and the general covariance of nonrelativistic laws. Am. J. Phys. **40**, 683–687 (1972)
447. Rosen, J.: Symmetry Rules—How Science and Nature are Founded on Symmetry. Springer-Verlag, Berlin (2008)
448. Rosen, N.: General relativity and flat space. I. Phys. Rev. **57**, 147–150 (1940)
449. Rosenfeld, L.: Zur Quantelung der Wellenfelder. Ann. Phys. **5**, 113 (1930). Translated and commented by D. Salisbury, On the quantization of wave fields. Preprint 381, Max Planck Institute for the History of Science, Berlin, (2009); Ann. Phys. (forthcoming)
450. Rosenfeld, L.: Sur le tenseur d'impulsion-énergie. Mem. Acad. Roy. Belg. Sci. **18**, 1–30 (1940). English translation: On the energy-momentum tensor. In: Cohen R.S., Stachel J.J., Selected Papers of Léon Rosenfeld, Reidel Publishing Company, Dordrecht, 711–735 (1979)

451. Rovelli, C.: Quantum Gravity. Cambridge University Press, Cambridge (2004)
452. Rovelli C.: Why Gauge?. Found. Phys. **44**, 91–104 (2014)
453. Rubakov, V.: Classical Theory of Gauge Fields. Princeton University Press, Princeton (2002)
454. Ruegg, H., Ruit-Altaba, M.: The Stueckelberg field. Int. J. Mod. Phys. **A19**, 3265–3348 (2004)
455. Ryan, M.P., Shepley, L.C.: Homogeneous Relativistic Cosmologies. Princeton University Press, Princeton (1975)
456. Ryder, L.H.: Quantum Field Theory. Cambridge University Press, Cambridge (1985)
457. Salam, A., Strathdee, J.: Super-gauge transformations. Nucl. Phys. **B76**, 477–82 (1974). Reprinted in [176]
458. Salam, A., Strathdee, J.: Super-symmetry and non-Abelian gauge. Phys. Lett. **51B**, 353–355 (1974). Reprinted in [176]
459. Saletan, E.J., Cromer, A.H.: Theoretical Mechanics. Wiley, New York (1971)
460. Salisbury, D.C.: Peter Bergmann and the invention of constrained Hamiltonian dynamics. Seventh international conference on the history of general relativity. Tenerife, Spain (2005). arXiv:physics/0608067
461. Salisbury, D.C.: Rosenfeld, Bergmann, Dirac and the Invention of Constrained Hamiltonian Dynamics. In: Kleinert, H., Jantzen, R.T., Ruffini, R. (eds.) Proceedings of the 11th Marcel Grossmann Meeting on General Relativity. World Scientific, Singapore (2008); also arXiv:physics/0701299
462. Salisbury, D.C., Sundermeyer, K.: Realization in phase space of general coordinate transformations. Phys. Rev. **D27**, 740–756 (1983); Local symmetries of the Einstein-Yang-Mills theory as phase-space transformations. Phys. Rev. **D27**, 757–763 (1983)
463. Sarlet, W., Cantrijn, F.: Generalizations of Noether's theorem in classical mechanics. SIAM Rev. **23**, 467–494 (1981)
464. Sauer, T.: Field equations in teleparallel space-time: Einstein's 'Fernparallelismus' approach toward unified field theory. Hist. Math. **33**, 399–439 (2006)
465. Sauer, T., Majer, U. (eds.): David Hilbert's Lectures on the Foundations of Physics 1915–1927. Springer (2009). Another volume, covering the years 1898–1914, is announced for 2015
466. Schellekens, A.N.: The emperor's last clothes? Overlooking the string theory landscape. Rep. Prog. Phys. **71**, 072201 (2008)
467. Schottenloher, M.: Geometrie und Symmetrie in der Physik. Vieweg, Braunschweig (1995)
468. Schouten, J.A.: Ricci Calculus. Springer-Verlag, Berlin (1954)
469. Schouten, J.A.: Tensor Analysis for Physicists, 2nd edn. Dover, New York (1989)
470. Schutz, B.: Geometrical Methods of Mathematical Physics. Cambridge University Press, Cambridge (1980)
471. Seiler, E., Stamatescu, I.-O. (eds.): Approaches to Fundamental Physics. Springer, Berlin (2007)
472. Sexl, R.U., Urbantke, H.K.: Relativity, Groups, Particles—Special Relativity and Relativistic Symmetry in Field and Particle Physics. Revised and translated from the German. Springer, Wien (2001)
473. Shie, K.-F., Nester, J.M., Yo, W.-J.: Torsion cosmology and the accelerating universe. Phys. Rev. **D78**, 023522 (2008). Reprinted in [50]
474. Silagadze, Z.K.: Relativity without tears. Acta Phys. Polonica **B39**, 811–885 (2008)
475. Slansky, R.: Group theory for unified model building. Phys. Rep. **79**, 1–128 (1981)
476. Smolin, L.: The Trouble with Physics. Houghton Mifflin Company, Boston (2006)
477. Smolin, L.: A perspective on the landscape problem. Found. Phys. **43**, 21–45 (2013)
478. Sotiriou, Th.P.: Hořava-Lifschitz gravity: A status report. J. Phys. Conf. Ser. **283**, 012034 (2011)
479. Sotiriou, Th.P., Faraoni, V.: f(R) theories of gravity. Rev. Mod. Phys. **82**, 451–497 (2010)
480. Sotiriou, Th.P., Liberati, S.: Theory of gravitation theories: A no-progress report. Int. J. Mod. Phys. **D17**, 399–423 (2008)
481. Srednicki, M.: Quantum Field Theory. Cambridge University Press, Cambridge (2007)
482. Stachel, J.: The meaning of general covariance: The hole story. In: Earman, J., Janis, A., Massey, G., Rescher, N. (eds.) Philosophical Problems of the Internal and External Worlds, pp. 129–160. University of Pittsburgh Press, Pittsburgh (1993)

483. Stachel, J.: Structure, individuality and quantum gravity. In: Rickles, D.P., French, S.R.D., Saatsi, J. (eds.) Structural Foundations of Quantum Gravity. Oxford University Press
484. Stelle, K.S., West, P.C.: Spontaneously broken de Sitter symmetry and the gravitational holonomy group. Phys. Rev. **D21**, 1466–1488 (1980). Reprinted in [50]
485. Stephani, H.: Differential Equations—Their Solution Using Symmetries. Cambridge University Press, Cambridge (1989)
486. Straumann, N.: On Pauli's Invention of Non-Abelian Kaluza-Klein Theory in 1953. arXiv:gr-qc/0012054
487. Straumann, N.: Unitary Representations of the Inhomogeneous Lorentz Group and Their Significance in Quantum Physics. Invited Talk at the Conference "Space and Time 100 Years after Minkowski" (2008). arXiv:0809.4942 (math-ph)
488. Straumann, N.: Problems with Modified Theories of Gravity, as Alternatives to Dark Energy. Invited talk at the conference "Beyond Einstein", Mainz (2008). arXiv:0809.5148 (gr-qc)
489. Straumann, N.: General Relativity, 2nd edn. Springer-Verlag, Berlin (2013)
490. Sudarshan, E.C.G., Mukunda, N.: Classical Dynamics—A Modern Perspective. Wiley, New York (1974)
491. Sundermeyer, K.: Constrained Dynamics—With Applications to Yang-Mills Theory. General Relativity, Classical Spin, Dual String Model, Lecture Notes in Physics 169. Springer-Verlag, Berlin (1982)
492. Sundermeyer, K.: Auxiliary 'fields' for the massive spinning relativistic particle. Lett. Nuovo Cim. **37**, 556–560 (1983)
493. Sundermeyer, K.: Extra 'gauge' symmetries in R+T^2gravitational theories. Phys. Lett. **134B**, 415–418
494. Susskind, L.: The Anthropic Landcape of String Theory. arXiv:hep-th/0302219; The Cosmic Landscape: String Theory and the Illusion of Intelligent Design. Little, New York (2005)
495. Szabados, L.B.: Quasi-local energy-momentum and angular momentum in GR: A review article. Living Rev. Relativ. **7**, 4 (2004). http://www.livingreviews.org/lrr-2004-4
496. Tanii, Y.: Introduction to Supergravities in Diverse Dimensions. arXiv:9802138 [hep-th]
497. Teitelboim, C.: How commutators of constraints reflect the spacetime structure. Ann. Phys. **79**, 542–557 (1973)
498. Teitelboim, C: The Hamiltonian Structure of Space-Time, in [267], vol. 1, 195–225
499. Thiemann, T.: Generalized boundary conditions for general relativity for the asymptotically flat case in terms of Ashtekar's variables. Class. Quant. Grav. **12**, 181–198 (1995) arXiv:gr-qc/9310039
500. 't Hooft, G.: Renormalization of massless Yang-Mills fields. Nucl. Phys. **B33**, 173–199 (1971). Renormalizable Lagrangians for massive Yang-Mills fields. Nucl. Phys. **B35**, 167–188 (1971)
501. 't Hooft, G.: Determinism Beneath Quantum Mechanics. arXiv:quant-ph/0212095
502. 't Hooft, G. (ed.): 50 Years of Yang-Mills Theory. World Scientific, New Jersey (2005)
503. 't Hooft, G.: Emergent quantum mechanics and emergent symmetries. AIP Conf. Proc. **957**, 154–163 (2007)
504. 't Hooft, G.: A Grand View of Physics. Int. J. Mod. Phys. **A23**, 3755–3759 (2008)
505. 't Hooft, G.: Unitarity in the Brout-Englert-Higgs mechanism for gravity. arXiv:0708.3184v4 [hep-th]
506. Toffoli, T.: What is the Lagrangian counting? Int. J. Theor. Phys. **42**, 363–381 (2003)
507. Tomboulis, E.T.: General relativity as the effective theory of $GL(4, \mathrm{R})$ spontanous symmetry breaking. Phys. Rev. **D84**, 084018 (2011)
508. Toretti, R.: Relativity and Geometry. Dover, New York (1996)
509. Trautman, A.: Conservation laws in general relativity. In: Witten, L. (ed.) Gravitation: An Introduction to Current Research. Wiley, New York (1962)
510. Trautman, A.: Noether equations and conservation laws. Comm. Math. Phys. **6**, 248–261 (1967)
511. Trautman, A.: Fiber Bundles, Gauge Fields, and Gravitation, in [267], pp. 287–308
512. Unruh, W.G.: Unimodular theory of canonical quantum gravity. Phys. Rev. **D40**, 1048–1052 (1989)

513. Utiyama, R.: Invariant Theoretical interpretation of Interaction. Phys. Rev. **101**, 1597–1607 (1956)
514. Utiyama, R.: Theory of invariant variation and the generalized canonical dynamics. Prog. Theor. Phys. Suppl. 19–44 (1959)
515. van Dam, H.: Jack, Ng.: Why 3+1 metrics rather than 4+0 or 2+2? Phys. Lett. **B250**, 159–162 (2001)
516. van Dongen, J.: Einstein and the Kaluza-Klein particle. Stud. Hist. Philos. Mod. Phys. **33**, 185–210 (2002)
517. van Holten, J.W.: Quantum theory of a massless spinning particle. Z. Phys. **C41**, 497–504 (1988)
518. van Fraasen, B.: Laws and Symmetries. Oxford University Press (1989)
519. Van Nieuwenhuizen, P.: Supergravity. Phys. Rep. **68**(4), 189–398 (1981)
520. Van der Waerden, B.L.: Spinoranalyse. Nachr. Ges. Wiss. Göttingen Math.-Phys. 100–109 (1929)
521. Velo, G., Zwanziger, D.: Propagation and quantization of Rarita-Schwinger waves in an external electromagnetic potential. Phys. Rev. **186**, 1337–1341 (1969); Noncausality and other defects of interaction Lagrangians for particles with spin one and higher. Phys. Rev. **188**, 2218–2222 (1969)
522. Verlinde, E.: On the origin of gravity and the laws of Newton. JEHP04 (2011) 029. arXiv:1001.0785
523. Von der Heyde, P.: A generalized Lovelock theorem for the gravitational field with torsion. Phys. Lett. **51A**, 381–382 (1975)
524. Von der Heyde, P.: The equivalence principle in the U_4 theory of gravitation. Nuovo Cim. Lett. **14**, 250 (1975)
525. von Neumann, J.: Mathematical Foundations of Quantum Mechanics. Princeton University Press, Princeton (1955)
526. Wald, R.M.: General Relativity. The University of Chicago Press, Chicago (1984)
527. Wald, R.M.: Black hole entropy is Noether charge. Phys. Rev. **D48**, R3427 (1993)
528. Wald, R.M., Zoupas, A.: General definition of 'conserved quantities' in general relativity and other theories of gravity. Phys. Rev. **D61**, 084027 (2000)
529. Wallner, R.P.: On hyperbolic U_4 manifolds with local duality. Acta Phys. Austriaca **55**, 67–87 (1983); On the structure of gravitational U_4-field equations. Gen. Rel. Grav., 1081–1107 (1985)
530. Walter, S.: Breaking in the 4-vectors: The four-dimensional movement in gravitation, 1905–1910. In: Renn, J., Schemmel, M. (eds.) Gravitation in the Twilight of Classical Physics: Between Mechanics, Field Theory, and Astronomy. The Genesis of General Relativity, vol. 3, pp. 193–252. Springer, Berlin (2007)
531. Weinberg, S.: Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations. Phys. Rev. **138**, B988–B1002 (1965)
532. Weinberg, S.: Dreams of a Final Theory—The Scientist's Search for the Ultimate Laws of Nature. Vintage, New York (1974)
533. Weinberg, S.: Ultraviolet divergences in quantum gravity. In: General Relativity—An Einstein Centenary Survey, pp. 790–831 Cambridge University Press (1979)
534. Weinberg, S.: Conceptual foundations of the unified theory of weak and electromagnetic interactions. Rev. Mod. Phys. **52**(3) (July 1980); since this is the Nobel prize talk Weinberg gave in Dec 1979 in Stockholm it can be found in various books collecting important articles on symmetries and unification
535. Weinberg, S.: Charges from extra dimensions. Phys. Lett. **125B**, 265–269 (1983)
536. Weinberg, S.: The Quantum Theory of Fields, vol. I Foundations (1995), vol. II Modern Applications (1996), vol. III Supersymmetry (2000). Cambridge University Press
537. Weinberg, S.: What is Quantum Field Theory and What Did We Think It is?, in [70]. arXiv:hep-th/9702027
538. Weinberg, S.: Changing Attitudes in the Standard Model, in [270], pp. 36–44
539. Weinberg, S.: Effective Field Theory, Past and Future. arXiv:0908 [hep-th] (1964)

540. Weinberg, S.: Pions in large N quantum chromodynamics. Phys. Rev. Lett. **105**, 261601 (2010)
541. Weinberg, S., Witten, E.: Limits on massless particles. Phys. Lett. **B96**, 59–62 (1980)
542. Weinstein, S., Wolpert, D., Fuchs, C. (eds.): Laws of nature—Their nature and knowability. http://pirsa.org/C10001
543. von Weizsäcker, C.F.: Die Einheit der Natur. Carl Hanser, München (1972)
544. Wess, J.: The conformal invariance in quantum field theory. Nuovo Cim. **18**, 1086–1107 (1960)
545. Wess, J., Bagger, J.: Supersymmetry and Supergravity. Princeton University Press (1992)
546. Wess, J., Zumino, B.: Supergauge transformations in four dimensions. Nucl. Phys. **B70**, pp. 39, (1974), reprinted in [176]
547. West, P.: Introduction to Supersymmetry and Supergravity. World Scientific, Singapore etc., extended second edition (1990)
548. Weyl, H.: Gravitation und Elektrizität. Situngsber. Preuss. Akad. Wissensch. 465–480 (1918); translated as Gravitation and Electricity, in [348], [429]
549. Weyl, H.: Raum, Zeit, Materie, 1st edn. (1918); Engl. transl.: Space Time Matter (1922), Methuen, reprinted 1952 Dover
550. Weyl, H.: Gruppentheorie und Quantenmechanik. Transl. by H. P. Robertson, The Theory of Groups and Quantum Mechanics (1931) reprint 1950 Dover
551. Weyl, H.: Elektron und Gravitation I. Z. Phys. **56**, 330–352 (1929), translated “Electron and Gravitation” in [429]
552. Weyl, H.: Symmetry. Princeton University Press (1952)
553. Wick, G.C., Wightman, A.S., Wigner, E.P.: The intrinsic parity of elementary particles. Phys. Rev. **88**, 101–105 (1952)
554. Wightman, A.S.: Superselection rules: Old and new. Nuovo Cim. **110B**, 751–769 (1995)
555. Wigner, E.P.: Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren. F. Vieweg, Braunschweig (1931), English translation: Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra by J.J. Griffin. Academic, New York (1959)
556. Wigner. E.P.: On unitary representations of the inhomogeneous Lorentz group. Ann. Math. **40**, 149 (1939)
557. Wigner, E.P.: Conservation laws in classical and quantum physics. Prog. Theor. Phys. **11**, 437–440 (1954)
558. Wigner, E.P.: Symmetries and Reflections. MIT Press, Cambridge (1970)
559. Wilczek, F.: Quantum field theory. Rev. Mod. Phys. **71**, S85–S95 (1999)
560. Wilczek, F.: Fantastic Realities. World Scientific, Singapore (2006)
561. Will, C.M.: Special Relativity: A Centenary Perspective. Einstein, 1905–2005: Poincaré Seminar 2005, Damour, T., Darrigol, O., Duplantier, B., Rivasseau, V. (eds.), pp. 33–58. Birkhauser, Basel (2006)
562. Will, C.M.: The confrontation between general relativity and experiment. Living. Rev. Rel. **9**, 3 (2005)
563. Wipf, A.: Hamilton’ Formalism for Systems with Constraints. In: Ehlers, J., Friedrich, H. (eds.) Canonical Gravity: From Classical to Quantum, Lecture Notes in Physics 434, pp. 22–58. Springer (1994)
564. Witten, E.: Dynamical breaking of supersymmetry. Nucl. Phys. **B185**, 513–554 (1981)
565. Witten, E.: Search for a realistic Kaluza-Klein theory. Nucl. Phys. **B186**, 412–428 (1981)
566. Woit, P.: Not Even Wrong. The Failure of String Theory and the Search for Unity in Physical Laws. Basic Books, New York (2006)
567. Woodard, R.P.: Avoiding dark energy with 1/R modifications of gravity. In: Papantonopoulos, L. (ed.) The Invisible Universe: Dark Matter and Dark Energy, Lecture Notes in Physics. 720, pp. 403, Springer, Berlin (2007); also available as arXiv:astro-ph/0601672
568. Wu, T.T., Yang, C.N.: Concept of nonintegrable phase factors and global formulation of gauge fields. Phys. Rev. **D12**, 3845–3857 (1975)
569. Yale, A., Padmanabhan, T.: Structure of Lanczos-Lovelock Lagrangians in critical dimensions. Gen. Rel. Grav. **43**, 1549–1570 (2011)

570. Yang, C.N., Mills, R.: Conservation of isotopic spin and isotopic gauge invariance. Phys. Rev. **96**, 191 (1954)
571. Yeomans, J.M.: Statistical Mechanics of Phase Transitions. Oxford University Press (1992)
572. Yo, H.-J., Nester, J.M.: Hamiltonian analysis of Poincaré gauge theory scalar modes. Int. J. Mod. Phys. **D8**, 459–479 (1999); Hamiltonian analysis of Poincaré gauge theory: Higher spin modes. Int. J. Mod. Phys. **D11**, 747–779 (2002)
573. Yourgrau, W., Mandelstam, S.: Variational Principles in Dynamics and Quantum Theory. Dover, New York (1979)
574. York Jr, J.W.: Role of conformal three-geometry in the dynamics of gravitation. Phys. Rev. Lett. **28**, 1082–1085 (1972)
575. York Jr, J.W.: Boundary terms in the action principles of general relativity. Found. Phys. **16**, 249–257 (1986)
576. Zakhary, E., McIntosh, C.B.G.: A complete set of Riemann invariants. Gen. Rel. Grav. **29**, 539–581 (1997)
577. Zanelli, J.: Chern-Simons forms in gravitation theories. Class. Quantum Grav. **29**, 133001 (2012); Lecture Notes on Chern-Simons (Super-) Gravities. 2nd edn. Feb. 2008, arXiv:hep-th/0502193
578. Zee, A.: Fearful Symmetry—The Search for Beauty in Modern Physics. Princeton University Press (1999)
579. Zee, A.: Quantum Field Theory in a Nutshell. Princeton University Press (2003)
580. Zeh, H.D.: The Physical Basis of the Direction of Time. Springer, Berlin-Heidelberg (2007)
581. Zinn-Justin, J.: The important role of functional integrals in modern physics. Theor. Math. Phys. **169**, 1380–1389 (2011)
582. Zumino, B.: Gravity theories in more than four dimensions. Phys. Rep. **137**, 109–114 (1986)
583. Zwiebach, B.: A First Course in String Theory. Cambridge University Press, Cambridge (2004)

Index

K. Sundermeyer, *Symmetries in Fundamental Physics*,
Fundamental Theories of Physics 176, DOI: 10.1007/978-3-319-06581-6,

Q

R

S