

Graduate Texts in Physics

Günter Ludyk

Einstein in Matrix Form

Exact Derivation of the Theory of Special
and General Relativity without Tensors



Springer

Graduate Texts in Physics

For further volumes:
www.springer.com/series/8431

Graduate Texts in Physics

Graduate Texts in Physics publishes core learning/teaching material for graduate- and advanced-level undergraduate courses on topics of current and emerging fields within physics, both pure and applied. These textbooks serve students at the MS- or PhD-level and their instructors as comprehensive sources of principles, definitions, derivations, experiments and applications (as relevant) for their mastery and teaching, respectively. International in scope and relevance, the textbooks correspond to course syllabi sufficiently to serve as required reading. Their didactic style, comprehensiveness and coverage of fundamental material also make them suitable as introductions or references for scientists entering, or requiring timely knowledge of, a research field.

Series Editors

Professor William T. Rhodes

Department of Computer and Electrical Engineering and Computer Science
Imaging Science and Technology Center
Florida Atlantic University
777 Glades Road SE, Room 456
Boca Raton, FL 33431
USA
wrhodes@fau.edu

Professor H. Eugene Stanley

Center for Polymer Studies Department of Physics
Boston University
590 Commonwealth Avenue, Room 204B
Boston, MA 02215
USA
hes@bu.edu

Professor Richard Needs

Cavendish Laboratory
JJ Thomson Avenue
Cambridge CB3 0HE
UK
rnl1@cam.ac.uk

Professor Martin Stutzmann

Walter Schottky Institut
TU München
85748 Garching
Germany
stutz@wsi.tu-muenchen.de

Günter Ludyk

Einstein in Matrix Form

Exact Derivation of the Theory of Special
and General Relativity without Tensors



Springer

Professor Dr. Günter Ludyk
Physics and Electrical Engineering
University of Bremen
Bremen, Germany
Guenter.Ludyk@nord-com.net

ISSN 1868-4513
Graduate Texts in Physics
ISBN 978-3-642-35797-8
DOI 10.1007/978-3-642-35798-5
Springer Heidelberg New York Dordrecht London

ISSN 1868-4521 (electronic)
ISBN 978-3-642-35798-5 (eBook)

Library of Congress Control Number: 2013942312

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

*To my grandchildren Ann-Sophie
and
Alexander
Hüttermann*

Preface

This book is an introduction to the theories of Special and General Relativity. The target audience are physicists, engineers and applied scientists who are looking for an understandable introduction to the topic—without too much new mathematics.

All necessary mathematical tools are provided either directly in the text or in the appendices. Also the appendices contain an introduction to vector or matrices: first, as a refresher of known fundamental algebra, and second, to gain new experiences, e.g. with the Kronecker-product of matrices and differentiation with respect to vectors and matrices.

The fundamental equations of Einstein’s theory of Special and General Relativity are derived using matrix calculus without the help of tensors. This feature makes the book special and a valuable tool for scientists and engineers with no experience in the field of tensor calculus. But physicists are also discovering that Einstein’s vacuum field equations can be expressed as a system of first-order differential-matrix equations, wherein the unknown quantity is a matrix. These matrix equations are also easy to handle when implementing numerical algorithms using standard software as, e.g. MATHEMATICA or MAPLE.

In Chap. 1, the foundations of Special Relativity are developed. Chapter 2 describes the structure and principles of General Relativity. Chapter 3 explains the Schwarzschild solution of spherical body gravity and examines the “Black Hole” phenomenon. Furthermore, two appendices summarize the basics of the matrix theory and differential geometry.

After completion of the book, I discovered the paper [37], where Einstein’s equations of a similar shape are derived.

I would like to thank *Claus Ascheron* (Springer) who has made great effort towards the publication of this book. Finally, I would like to thank my wife *Renate*, without her this book would have never been published!

Bremen

Günter Ludyk

Contents

1	Special Relativity	1
1.1	Galilei Transformation	1
1.1.1	Relativity Principle of Galilei	1
1.1.2	General Galilei Transformation	5
1.1.3	Maxwell's Equations and Galilei Transformation	6
1.2	Lorentz Transformation	7
1.2.1	Introduction	7
1.2.2	Determining the Components of the Transformation Matrix	8
1.2.3	Simultaneity at Different Places	12
1.2.4	Length Contraction of Moving Bodies	13
1.2.5	Time Dilation	15
1.3	Invariance of the Quadratic Form	15
1.3.1	Invariance with Respect to Lorentz Transformation	17
1.3.2	Light Cone	17
1.3.3	Proper Time	19
1.4	Relativistic Velocity Addition	20
1.4.1	Galilean Addition of Velocities	20
1.5	Lorentz Transformation of the Velocity	22
1.6	Momentum and Its Lorentz Transformation	25
1.7	Acceleration and Force	26
1.7.1	Acceleration	26
1.7.2	Equation of Motion and Force	28
1.7.3	Energy and Rest Mass	30
1.7.4	Emission of Energy	31
1.8	Relativistic Electrodynamics	32
1.8.1	Maxwell's Equations	32
1.8.2	Lorentz Transformation of the Maxwell's Equations	34
1.8.3	Electromagnetic Invariants	37
1.8.4	Electromagnetic Forces	39

1.9	The Energy–Momentum Matrix	41
1.9.1	The Electromagnetic Energy–Momentum Matrix	41
1.9.2	The Mechanical Energy–Momentum Matrix	43
1.9.3	The Total Energy–Momentum Matrix	47
1.10	The Most Important Definitions and Formulas in Special Relativity	48
2	Theory of General Relativity	51
2.1	General Relativity and Riemannian Geometry	51
2.2	Motion in a Gravitational Field	53
2.2.1	First Solution	54
2.2.2	Second Solution	55
2.2.3	Relation Between $\tilde{\Gamma}$ and \mathbf{G}	56
2.3	Geodesic Lines and Equations of Motion	57
2.3.1	Alternative Geodesic Equation of Motion	62
2.4	Example: Uniformly Rotating Systems	64
2.5	General Coordinate Transformations	67
2.5.1	Absolute Derivatives	67
2.5.2	Transformation of the Christoffel Matrix $\tilde{\Gamma}$	69
2.5.3	Transformation of the Christoffel Matrix $\hat{\Gamma}$	71
2.5.4	Coordinate Transformation and Covariant Derivative	72
2.6	Incidental Remark	76
2.7	Parallel Transport	78
2.8	Riemannian Curvature Matrix	80
2.9	Properties of the Riemannian Curvature Matrix	81
2.9.1	Composition of \mathbf{R} and $\tilde{\mathbf{R}}$	81
2.10	The Ricci Matrix and Its Properties	88
2.10.1	Symmetry of the Ricci Matrix \mathbf{R}_{Ric}	90
2.10.2	The Divergence of the Ricci Matrix	92
2.11	General Theory of Gravitation	93
2.11.1	The Einstein's Matrix \mathfrak{G}	93
2.11.2	Newton's Theory of Gravity	94
2.11.3	The Einstein's Equation with \mathfrak{G}	97
2.12	Summary	99
2.12.1	Covariance Principle	99
2.12.2	Einstein's Field Equation and Momentum	101
2.13	Hilbert's Action Functional	101
2.13.1	Effects of Matter	105
2.14	Most Important Definitions and Formulas	106
3	Gravitation of a Spherical Mass	109
3.1	Schwarzschild's Solution	109
3.1.1	Christoffel Matrix \mathbf{I}	110
3.1.2	Ricci Matrix \mathbf{R}_{Ric}	112
3.1.3	The Factors $A(r)$ and $B(r)$	114
3.2	Influence of a Massive Body on the Environment	116
3.2.1	Introduction	116

3.2.2	Changes to Length and Time	117
3.2.3	Redshift of Spectral Lines	118
3.2.4	Deflection of Light	120
3.3	Schwarzschild's Inner Solution	124
3.4	Black Holes	127
3.4.1	Astrophysics	127
3.4.2	Further Details about "Black Holes"	129
3.4.3	Singularities	131
3.4.4	Eddington's Coordinates	135
3.5	Rotating Masses	138
3.5.1	Ansatz for the Metric Matrix \mathbf{G}	138
3.5.2	Kerr's Solution in Boyer–Lindquist Coordinates	139
3.5.3	The Lense–Thirring Effect	139
3.6	Summary of Results for the Gravitation of a Spherical Mass	141
3.7	Concluding Remark	143
Appendix A	Vectors and Matrices	145
A.1	Vectors and Matrices	145
A.2	Matrices	147
A.2.1	Types of Matrices	147
A.2.2	Matrix Operations	148
A.2.3	Block Matrices	152
A.3	The Kronecker-Product	154
A.3.1	Definitions	154
A.3.2	Some Theorems	154
A.3.3	The Permutation Matrix $\mathbf{U}_{p \times q}$	156
A.3.4	More Properties of the Kronecker-Product	157
A.4	Derivatives of Vectors/Matrices with Respect to Vectors/Matrices	157
A.4.1	Definitions	157
A.4.2	Product Rule	158
A.4.3	Chain Rule	159
A.5	Differentiation with Respect to Time	159
A.5.1	Differentiation of a Function with Respect to Time	159
A.5.2	Differentiation of a Vector with Respect to Time	160
A.5.3	Differentiation of a 2×3 -Matrix with Respect to Time	161
A.5.4	Differentiation of an $n \times m$ -Matrix with Respect to Time	161
A.6	Supplements to Differentiation with Respect to a Matrix	162
Appendix B	Some Differential Geometry	165
B.1	Curvature of a Curved Line in Three Dimensions	165
B.2	Curvature of a Surface in Three Dimensions	166
B.2.1	Vectors in the Tangent Plane	166
B.2.2	Curvature and Normal Vectors	168
B.2.3	Theorema Egregium and the Inner Values g_{ij}	169

Appendix C Geodesic Deviation	179
Appendix D Another Ricci-Matrix	183
References	189
Index	191

Notation

Important definitions, facts and theorems are framed, and important *intermediate results* are double-underlined.

Scalars are written in normal typeface:

$$a, b, c, \alpha, \beta, \gamma, \dots;$$

Vectors in the 3-dimensional space (\mathbb{R}^3) are written in small bold typeface:

$$\mathbf{x}, \mathbf{v}, \mathbf{u}, \mathbf{a}, \dots;$$

Vectors in the 4-dimensional spacetime (\mathbb{R}^4) are written in small bold typeface with an arrow:

$$\vec{\mathbf{x}}, \vec{\mathbf{v}}, \vec{\mathbf{u}}, \vec{\mathbf{a}}, \dots;$$

Matrices are written in big bold typeface:

$$\mathbf{M}, \mathbf{G}, \mathbf{R}, \mathbf{I}, \dots.$$

The identity matrix \mathbf{I}_n of size $n \times n$ is the matrix in which all the elements on the main diagonal are equal to 1 and all other elements are equal to 0, e.g.,

$$\mathbf{I}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The *Derivative Operator* ∇ is the 3-dimensional column vector

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix},$$

and ∇^\top is the 3-dimensional row-vector

$$\nabla^\top = \left(\frac{\partial}{\partial x} \Big| \frac{\partial}{\partial y} \Big| \frac{\partial}{\partial z} \right).$$

The *Derivative Operator* $\vec{\nabla}$ is the 4-dimensional column vector

$$\vec{\nabla} = \gamma \begin{pmatrix} -\frac{1}{c} \frac{\partial}{\partial t} \\ \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}$$

with

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Remark The derivative operators ∇ and $\vec{\nabla}$ are column vector operators and can act both on the right and on the left! Example:

$$\vec{\nabla}^\top \vec{a} = \vec{a}^\top \vec{\nabla} = \gamma \left(-\frac{1}{c} \frac{\partial a_0}{\partial t} + \frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z} \right).$$

Chapter 1

Special Relativity

This chapter begins with the classical theorems of Galilei and Newton and the Galilei transformation. The special theory of relativity, developed by Einstein in 1905, leads to the four-dimensional spacetime of Minkowski and the Lorentz transformation. After that the relativity of simultaneity of events, the length contraction of moving bodies and the time dilation are discussed. This is followed by the velocity-addition formula and relativistic mechanics. The next topic is the mass–energy equivalence formula $E = mc^2$, where c is the speed of light in a vacuum. Then relativistic electromagnetism is treated and the invariance of special forms of the equations of dynamics and Maxwell’s electrodynamic with respect to the Lorentz transformation is shown. The energy–momentum matrix is introduced and discussed.

1.1 Galilei Transformation

1.1.1 Relativity Principle of Galilei

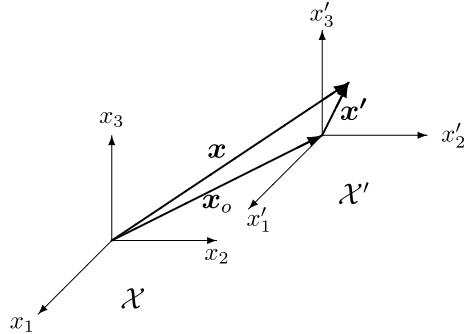
An *event* is anything that can happen in space and time, e.g. the emission of a flash of light in a room corner. Events happen at a single point. We assign to each event a set of four coordinates t, x_1, x_2 and x_3 , or with t and the three-dimensional vector

$$\boldsymbol{x} \stackrel{\text{def}}{=} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3.$$

The position vector \boldsymbol{x} and the time t form a *reference frame* \mathcal{X} . In this frame, Newton’s fundamental law of mechanics has the form

$$\frac{d\boldsymbol{p}}{dt} = \boldsymbol{f}$$

Fig. 1.1 Two against each other shifted reference systems



or, if the mass m in the momentum

$$\mathbf{p} = m \frac{d\mathbf{x}}{dt}$$

is constant,

$$m \frac{d^2\mathbf{x}}{dt^2} = \mathbf{f}. \quad (1.1)$$

An observer may now execute any motion, for example, he makes an experiment in a moving train. We want to find the equation that takes the place of

$$\frac{dp}{dt} = \mathbf{f}$$

for the moving observer. A coordinate system \mathcal{X}' is connected firmly with the moving observer; it should be axis-parallel to the original coordinate system \mathcal{X} . \mathbf{x}_o is the location of the origin of \mathcal{X}' measured in \mathcal{X} (Fig. 1.1). Then

$$\mathbf{x} = \mathbf{x}_o + \mathbf{x}',$$

or

$$\mathbf{x}' = \mathbf{x} - \mathbf{x}_o. \quad (1.2)$$

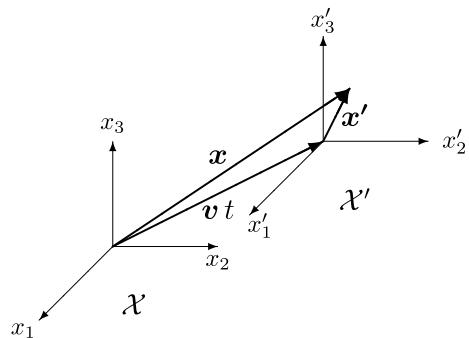
Here \mathbf{x} is the position vector of the event measured by an observer at rest in the reference system \mathcal{X} , and \mathbf{x}' is what an observer measures in the moving reference system \mathcal{X}' . Equation (1.2) differentiated with respect to time t ,

$$\frac{dx'}{dt} = \frac{d\mathbf{x}}{dt} - \frac{d\mathbf{x}_o}{dt}, \quad (1.3)$$

results in the speed addition theorem of classical mechanics:

$$\mathbf{v}'(t) = \mathbf{v}(t) - \mathbf{v}_0(t).$$

Fig. 1.2 Two reference systems moving against each other



For the acceleration one obtains

$$\frac{d^2 \mathbf{x}'}{dt^2} = \frac{d^2 \mathbf{x}}{dt^2} - \frac{d^2 \mathbf{x}_o}{dt^2}. \quad (1.4)$$

The force f acting on the mass m is independent of the chosen coordinate system, so $f' = f$. This and (1.1) used in (1.4) result in

$$f' - m \frac{d^2 \mathbf{x}_o}{dt^2} = m \frac{d^2 \mathbf{x}'}{dt^2}. \quad (1.5)$$

The fundamental law of mechanics has lost its validity! If the moving observer knows the external force f' , he can determine with measurements in \mathcal{X}' his acceleration with respect to the rest system \mathcal{X} . However, if the motion of \mathcal{X}' with respect to \mathcal{X} is uniform and rectilinear, i.e. $\mathbf{x}_o = \mathbf{v}_o t$ with a constant \mathbf{v}_o , then from (1.5)

$$f' = m \frac{d^2 \mathbf{x}'}{dt^2} \quad (1.6)$$

and the fundamental laws of mechanics have in \mathcal{X}' the same form as in \mathcal{X} . Such a uniform and rectilinear moving coordinate system is called an *inertial system*. The moving observer has no possibility to determine his own motion with respect to the coordinate system \mathcal{X} by a mechanical experiment. Any free particle moves in a straight line with constant speed. For example, Galilei considered a uniformly moving ship in a port, whose occupants cannot decide whether the ship moves with respect to the port or whether the port moves with respect to the ship. Today as an example one would take a train in a railway station. This is the *Relativity Principle* of Galilei:

Axiom: All natural laws are the same at every moment in all inertial systems.

All coordinate systems moving uniformly linearly with respect to an inertial system are themselves inertial systems.

If two reference systems \mathcal{X} and \mathcal{X}' move with a constant velocity v against each other (Fig. 1.2), then

$$\vec{x}' = \vec{x} - vt \quad \text{and} \quad t' = t. \quad (1.7)$$

With the four-dimensional column vector

$$\vec{x} \stackrel{\text{def}}{=} \begin{pmatrix} t \\ \vec{x} \end{pmatrix} \in \mathbb{R}^4$$

the two equations (1.7) can be summarized in one equation, and we get the Galilei transformation in a matrix form

$$\underline{\underline{\vec{x}' = \begin{pmatrix} 1 & \vec{o}^T \\ -v & I \end{pmatrix} \vec{x}}} = T_{\text{Galilei}} \vec{x}. \quad (1.8)$$

Looking at the inverse transformation of \mathcal{X}' to \mathcal{X} , one gets

$$t = t' \quad \text{and} \quad \vec{x} = vt + \vec{x}',$$

or

$$\vec{x} = \begin{pmatrix} 1 & \vec{o}^T \\ v & I \end{pmatrix} \vec{x}' = T'_{\text{Galilei}} \vec{x}'.$$

If both transformations are applied in series, the result is

$$T_{\text{Galilei}} T'_{\text{Galilei}} \vec{x}' = \begin{pmatrix} 1 & \vec{o}^T \\ -v & I \end{pmatrix} \begin{pmatrix} 1 & \vec{o}^T \\ v & I \end{pmatrix} \vec{x}' = \begin{pmatrix} 1 & \vec{o}^T \\ \vec{o} & I \end{pmatrix} \vec{x}' = \vec{x},$$

i.e. the matrix T'_{Galilei} is the inverse of the matrix T_{Galilei} .

For the time derivatives of the four-vector \vec{x} we obtain

$$\frac{d\vec{x}}{dt} = \begin{pmatrix} 1 \\ \frac{d\vec{x}}{dt} \end{pmatrix} \quad \text{and} \quad \frac{d^2\vec{x}}{dt^2} = \begin{pmatrix} 0 \\ \frac{d^2\vec{x}}{dt^2} \end{pmatrix}.$$

With $\vec{f} \stackrel{\text{def}}{=} \begin{pmatrix} 0 \\ f \end{pmatrix}$ we can write the fundamental equation of mechanics

$$\vec{f} = m \frac{d^2\vec{x}}{dt^2}. \quad (1.9)$$

This equation multiplied from the left with the transformation matrix T_{Galilei} , in fact, gives back the same form:

$$\underbrace{T_{\text{Galilei}} \vec{f}}_{\stackrel{\text{def}}{=} \vec{f}'} = m \underbrace{T_{\text{Galilei}} \frac{d^2\vec{x}}{dt^2}}_{\stackrel{\text{def}}{=} \frac{d^2\vec{x}'}{dt'^2}}, \quad (1.10)$$

$$\vec{f}' = m \frac{d^2\vec{x}'}{dt'^2}.$$

The fundamental equation of dynamics is thus *invariant* with respect to this Galilei transformation, i.e. it retains its shape regardless of the reference system.

1.1.2 General Galilei Transformation

Up to this point, the Galilei transformation was considered only under the uniform motion with the speed \mathbf{v} of the two inertial systems against each other and a fixed initial time $t_0 = 0$, a fixed initial point $\mathbf{x}_0 = \mathbf{0}$ of the new coordinate system \mathcal{X}' and with no rotation of the coordinate system.

It is now generally assumed that all laws of nature remain constant, therefore, are *invariant* with respect to *time shifting*. If $\mathbf{x}(t)$ is a solution of $m\ddot{\mathbf{x}} = \mathbf{f}$, then for all $t_0 \in \mathbb{R}$ is $\mathbf{x}(t + t_0)$ also a solution.

Next, it is assumed that the considered spaces are *homogeneous*, thus the same features are available at all points. So, if $\mathbf{x}(t)$ is again a solution of $m\ddot{\mathbf{x}} = \mathbf{f}$, then $\mathbf{x}(t) + \mathbf{b}$ is also a solution, but now for the starting point $\mathbf{x}_0 + \mathbf{b}$.

Moreover, it is assumed that the considered spaces are *isotropic*, i.e. there is no directional dependence of properties. So again, if $\mathbf{x}(t)$ is a solution of $m\ddot{\mathbf{x}} = \mathbf{f}$, then also $D\mathbf{x}(t)$ is a solution for the initial state $D\mathbf{x}_0$. Here, however, also the equality of the distances

$$\rho(\mathbf{x}_1, \mathbf{x}_2) = \rho(D\mathbf{x}_1, D\mathbf{x}_2)$$

must be true. For the rotation matrix D this means that it must be *orthogonal* because

$$\begin{aligned} \rho(D\mathbf{x}_1, D\mathbf{x}_2) &= \sqrt{(D\mathbf{x}_2 - D\mathbf{x}_1)^T (D\mathbf{x}_2 - D\mathbf{x}_1)} \\ &= \sqrt{(\mathbf{x}_2 - \mathbf{x}_1)^T D^T D (\mathbf{x}_2 - \mathbf{x}_1)} \stackrel{!}{=} \rho(\mathbf{x}_1, \mathbf{x}_2) \\ &= \sqrt{(\mathbf{x}_2 - \mathbf{x}_1)^T (\mathbf{x}_2 - \mathbf{x}_1)}, \end{aligned}$$

so

$$D^T D \stackrel{!}{=} I.$$

The time invariance, homogeneity and isotropy can be summarized in one transformation, the *general Galilei transformation*, as follows:

If t' is shifted versus t by t_0 , i.e. one has

$$t' = t_0 + t, \tag{1.11}$$

furthermore, the new coordinate system relative to the old is moved by \mathbf{x}_0 and rotated by the rotation matrix D , so for $\mathbf{v} = \mathbf{0}$ one gets

$$\mathbf{x}' = D\mathbf{x} + \mathbf{x}_0, \tag{1.12}$$

and so

$$\begin{pmatrix} t' \\ \mathbf{x}' \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{o}^\top \\ \mathbf{o} & \mathbf{D} \end{pmatrix} \begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} + \begin{pmatrix} t_o \\ \mathbf{x}_o \end{pmatrix}. \quad (1.13)$$

By moving the origin of the new coordinate system with the velocity \mathbf{v} as mentioned above, one finally gets the general Galilei transformation:

$$\vec{\mathbf{x}}' = \begin{pmatrix} 1 & \mathbf{o}^\top \\ -\mathbf{v} & \mathbf{D} \end{pmatrix} \vec{\mathbf{x}} + \vec{\mathbf{x}}_o, \quad (1.14)$$

with

$$\vec{\mathbf{x}} \stackrel{\text{def}}{=} \begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} \quad \text{and} \quad \vec{\mathbf{x}}_o \stackrel{\text{def}}{=} \begin{pmatrix} t_o \\ \mathbf{x}_o \end{pmatrix}.$$

This is an *affine mapping*, or an *affine transformation*.

One gets a *linear transformation* by introducing the extended vector:

$$\begin{pmatrix} t \\ \mathbf{x} \\ 1 \end{pmatrix} \in \mathbb{R}^5, \quad (1.15)$$

namely

$$\begin{pmatrix} t' \\ \mathbf{x}' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{o}^\top & t_0 \\ -\mathbf{v} & \mathbf{D} & \mathbf{x}_0 \\ 0 & \mathbf{o}^\top & 1 \end{pmatrix} \begin{pmatrix} t \\ \mathbf{x} \\ 1 \end{pmatrix}. \quad (1.16)$$

Newton's fundamental laws are invariant also with respect to such a transformation. A general Galilei transformation is determined by 10 parameters: $t_o, \mathbf{x}_o \in \mathbb{R}^3$, $\mathbf{v} \in \mathbb{R}^3$ and $\mathbf{D} \in \mathbb{R}^{3 \times 3}$. The rotation matrix \mathbf{D} , in fact, has only three main parameters since any general rotation consists of successive rotations performed around the x_1 -, x_2 - and x_3 -axis, so the whole rotation is characterized by the three angles φ_1, φ_2 and φ_3 , where, for example, the rotation around the x_1 -axis is achieved by the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_1 & \sin \varphi_1 \\ 0 & -\sin \varphi_1 & \cos \varphi_1 \end{pmatrix}.$$

1.1.3 Maxwell's Equations and Galilei Transformation

The situation is completely different with Maxwell's equations of electromagnetic dynamics. They are not invariant with respect to a Galilei transformation! Indeed, in an inertial system \mathcal{X} , a static charge q generates only a static electric field; in an inertial frame, moving with the speed \mathbf{v} , $q\mathbf{v}$ is an electric current which generates there a magnetic field!

In the nineteenth century, it was believed that all physical phenomena are mechanical and electromagnetic forces could be traced to the stress states of a world-aether, the Maxwell's tensions. It was assumed that even a vacuum must be filled with aether. This aether then is the carrier of the electromagnetic phenomena.

Suppose an *inertial frame* is a reference system in which Galilei's principle of inertia is valid. Then Einstein in his *general relativity principle* claims:

The laws of nature take on the same form in all inertial systems.

For the fundamental law of mechanics, this is derived above. The principle of relativity applies neither to electrodynamics nor to optics. How should the basic equations of electrodynamics be modified so that the relativity principle is valid? This is the content of Einstein's *Theory of Special Relativity*. He expanded it in the *Theory of General Relativity*. This theory treats how the natural laws must be modified so that they are also valid in *accelerating* or not uniformly against each other moving reference systems.

1.2 Lorentz Transformation

1.2.1 Introduction

At the end of the nineteenth century, experiments were conceived which should determine the velocity of the Earth with respect to the resting cosmic aether. This speed relative to the aether can only be measured by an electromagnetic effect, e.g. the light wave propagation. But in the Michelson–Morley experiment in 1881 and 1887 no drift velocity was found. Einstein concluded:¹

The speed of light c is always constant.

Independently of the movement of the light source and the observer, light has the same speed value in every inertial frame.

Assume a light pulse is produced at time $t = t' = 0$ in the two axis-parallel reference systems \mathcal{X} and \mathcal{X}' with a common origin. If the light spreads with the speed of light c in the reference system \mathcal{X}' , then, for example, it is true that $x'_1 = c \cdot t$. From (1.7) it follows for the x_1 -direction that if \mathbf{v} has the x_1 -directional component, then

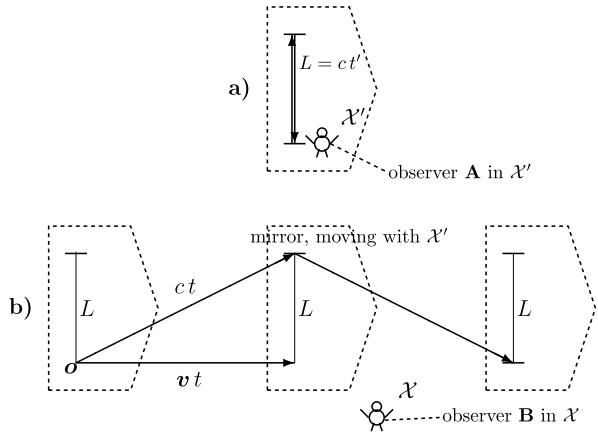
$$x_1 = x'_1 + vt = (c + v)t,$$

yielding, contrary to the Michelson–Morley experiment, a propagation velocity for light of $c + v > c$. Therefore, a transformation must be valid, different from the Galilei transformation. We try a linear transformation:

$$t = ft' + \mathbf{e}^\top \mathbf{x}', \quad (1.17)$$

¹ c from Latin *celeritas*: speed.

Fig. 1.3 Different ways of light: (a) seen by an observer in \mathcal{X}' , (b) seen by an observer in \mathcal{X}



$$\mathbf{x} = \mathbf{b}t' + \mathbf{A}\mathbf{x}', \quad (1.18)$$

i.e.

$$\vec{x} = \hat{\mathbf{L}}' \vec{x}' \in \mathbb{R}^4,$$

with

$$\hat{\mathbf{L}}' \stackrel{\text{def}}{=} \begin{pmatrix} f & \mathbf{e}^\top \\ \mathbf{b} & \mathbf{A} \end{pmatrix} \in \mathbb{R}^{4 \times 4}.$$

1.2.2 Determining the Components of the Transformation Matrix

That t' is different from t (according to Galilei this was not the case, he took $t' = t$) is shown by the following reasoning. We consider two observers; observer **A** moves relative to the observer **B** in a spaceship with velocity \mathbf{v} (Fig. 1.3). The spaceship with observer **A** has the inertial system \mathcal{X}' , and the observer **B** on Earth has the inertial system \mathcal{X} . A beam of light moves from the origin $\mathbf{x} = \mathbf{x}' = \mathbf{o}$ of the reference systems \mathcal{X} and \mathcal{X}' at time $t = t' = 0$ perpendicular to the velocity \mathbf{v} and reaches after t' seconds for the observer **A** in the moving reference system \mathcal{X}' a mirror moving also with the reference system \mathcal{X}' . For the observer **B** in the stationary reference system \mathcal{X} , the light beam reaches the mirror, which has been moved in the \mathbf{v} direction a distance of $\mathbf{v} \cdot t$ after t seconds. Since in all initial systems the speed of light is constant and equal to c , by Pythagorean theorem, one has

$$(ct)^2 = (vt)^2 + (L)^2 = (vt)^2 + (ct')^2,$$

or, after solving for t ,

$$\underline{\underline{t = \gamma t'}}$$
(1.19)

with

$$\boxed{\gamma \stackrel{\text{def}}{=} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.} \quad (1.20)$$

A comparison of (1.19) with (1.17) provides for $\mathbf{x}' = \mathbf{o}$

$$\boxed{\underline{\underline{f}} = \gamma.} \quad (1.21)$$

For $\mathbf{x}' = \mathbf{o}$ one has $\mathbf{x} = \mathbf{v}t$, and from (1.18) it follows that $\mathbf{x} = \mathbf{b}t'$. Thus, $\mathbf{b}t' = \mathbf{v}t$, i.e. $\mathbf{b} = \mathbf{v}\frac{t}{t'}$. From (1.19) it follows, on the other hand, that $\frac{t}{t'} = \gamma$, so

$$\boxed{\underline{\underline{\mathbf{b}}} = \gamma \mathbf{v}.} \quad (1.22)$$

Up to now, the following transformation equations were determined:

$$t = \gamma t' + \mathbf{e}^\top \mathbf{x}', \quad (1.23)$$

$$\mathbf{x} = \gamma \mathbf{v}t' + \mathbf{A}\mathbf{x}'. \quad (1.24)$$

Equations (1.23) and (1.24) deliver a transformation of \mathcal{X}' to \mathcal{X} . If one wants to invert this transformation, one must replace \mathbf{v} by $-\mathbf{v}$, \mathbf{x} by \mathbf{x}' , and vice versa, and t by t' and vice versa (as \mathbf{A} and \mathbf{e} may depend on \mathbf{v} , in the following we first write $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{e}}$):

$$t' = \gamma t + \tilde{\mathbf{e}}^\top \mathbf{x},$$

$$\mathbf{x}' = -\gamma \mathbf{v}t + \tilde{\mathbf{A}}\mathbf{x},$$

which can be combined to

$$\vec{\mathbf{x}}' = \begin{pmatrix} \gamma & \tilde{\mathbf{e}}^\top \\ -\gamma \mathbf{v} & \tilde{\mathbf{A}} \end{pmatrix} \vec{\mathbf{x}} \stackrel{\text{def}}{=} \hat{\mathbf{L}} \vec{\mathbf{x}}. \quad (1.25)$$

Both transformations performed one after another must result in the identity matrix:

$$\hat{\mathbf{L}}' \hat{\mathbf{L}} \stackrel{!}{=} \mathbf{I}. \quad (1.26)$$

For the top left $(1, 1)$ -element of the matrix product $\hat{\mathbf{L}}' \hat{\mathbf{L}}$ one gets:

$$(\gamma, \mathbf{e}^\top) \begin{pmatrix} \gamma \\ -\gamma \mathbf{v} \end{pmatrix} = \gamma^2 - \gamma \mathbf{e}^\top \mathbf{v} \stackrel{!}{=} 1.$$

Hence,

$$\gamma \mathbf{e}^\top \mathbf{v} = \gamma^2 - 1. \quad (1.27)$$

Taking for \mathbf{e}

$$\mathbf{e}^\top = \alpha \mathbf{v}^\top \quad (1.28)$$

and using (1.27) yields

$$\gamma\alpha v^2 = \gamma^2 - 1,$$

and moreover,

$$\gamma\alpha = \frac{1}{v^2} \left(\frac{c^2}{c^2 - v^2} - 1 \right) = \frac{1}{c^2 - v^2} = \frac{\gamma^2}{c^2},$$

i.e.

$$\alpha = \frac{\gamma}{c^2}. \quad (1.29)$$

Equation (1.29) used in (1.28) finally yields

$$\underline{\underline{\boldsymbol{e}^\top}} = \frac{\gamma}{c^2} \underline{\underline{\boldsymbol{v}^\top}}. \quad (1.30)$$

Thus, till now we have calculated:

$$\hat{\boldsymbol{L}}' = \begin{pmatrix} \gamma & \frac{\gamma}{c^2} \boldsymbol{v}^\top \\ \gamma \boldsymbol{v} & \boldsymbol{A} \end{pmatrix}.$$

Obviously,

$$\tilde{\boldsymbol{e}}^\top = -\frac{\gamma}{c^2} \boldsymbol{v}^\top.$$

Suppose now that in (1.25) we have $\tilde{\boldsymbol{A}} = \boldsymbol{A}$. Then for the matrix element in the lower right corner of the matrix product $\hat{\boldsymbol{L}}' \hat{\boldsymbol{L}}$ in (1.26) we obtain

$$(\gamma \boldsymbol{v}, \boldsymbol{A}) \begin{pmatrix} -\frac{\gamma}{c^2} \boldsymbol{v}^\top \\ \boldsymbol{A} \end{pmatrix} = -\frac{\gamma^2}{c^2} \boldsymbol{v} \boldsymbol{v}^\top + \boldsymbol{A}^2 \stackrel{!}{=} \boldsymbol{I},$$

i.e.

$$\boldsymbol{A}^2 = \boldsymbol{I} + \frac{\gamma^2}{c^2} \boldsymbol{v} \boldsymbol{v}^\top. \quad (1.31)$$

From (1.27), by inserting (1.30), follows

$$\frac{\gamma^2}{c^2} v^2 = \gamma^2 - 1.$$

Plugging this into (1.31) provides

$$\boldsymbol{A}^2 = \boldsymbol{I} + (\gamma^2 - 1) \frac{\boldsymbol{v} \boldsymbol{v}^\top}{v^2}. \quad (1.32)$$

Since $(\gamma - 1)^2 = \gamma^2 - 2(\gamma - 1) - 1$, for $\gamma^2 - 1$ one can write

$$\gamma^2 - 1 = (\gamma - 1)^2 + 2(\gamma - 1). \quad (1.33)$$

Equation (1.33) inserted into (1.32) yields

$$A^2 = I + 2(\gamma - 1) \frac{\mathbf{v}\mathbf{v}^\top}{v^2} + (\gamma - 1)^2 \frac{\mathbf{v}\mathbf{v}^\top}{v^2} = \left(I + (\gamma - 1) \frac{\mathbf{v}\mathbf{v}^\top}{v^2} \right)^2,$$

where

$$\frac{\mathbf{v}\mathbf{v}^\top \mathbf{v}\mathbf{v}^\top}{v^4} = \frac{\mathbf{v}(\mathbf{v}^\top \mathbf{v})\mathbf{v}^\top}{v^4} = \frac{\mathbf{v}\mathbf{v}^\top}{v^2}$$

was used. Therefore,

$$\underline{\underline{A}} = I + (\gamma - 1) \frac{\mathbf{v}\mathbf{v}^\top}{v^2}. \quad (1.34)$$

It is, in fact, true that $A(-\mathbf{v}) = A(\mathbf{v})$, i.e. the above assumption that $\tilde{A} = A$ is correct.

Thus, the matrix \hat{L} of the Lorentz transformation is determined completely as

$\hat{L} = \left(\begin{array}{c c} \gamma & -\frac{\gamma}{c^2} \mathbf{v}^\top \\ \hline -\gamma \mathbf{v} & I + (\gamma - 1) \frac{\mathbf{v}\mathbf{v}^\top}{v^2} \end{array} \right).$	(1.35)
---	--------

For $c \rightarrow \infty$ we get $\gamma = 1$ and the Lorentz transformation turns into the Galilei transformation with T_{Galilei} .

In the often in the textbooks treated special case when the velocity \mathbf{v} is towards the x_1 -axis, i.e.

$$\mathbf{v} = \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix},$$

one obtains

$$\hat{L} = \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c^2}(v, 0, 0) \\ \hline \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix} & I + \frac{(\gamma-1)}{v^2} \begin{pmatrix} v^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \right) = \begin{pmatrix} \gamma & -\frac{\gamma v}{c^2} & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

so

$\begin{aligned} t' &= \gamma t - \frac{\gamma}{c^2} v x_1, \\ x'_1 &= -\gamma v t + \gamma x_1, \\ x'_2 &= x_2, \\ x'_3 &= x_3. \end{aligned}$	(1.36)
---	--------

Introducing as the time component in the vector \vec{x} the time multiplied with the speed of light c , i.e. $x_0 = ct$, we obtain from (1.35) the transformation

$$\vec{x}' \stackrel{\text{def}}{=} \begin{pmatrix} ct' \\ \mathbf{x}' \end{pmatrix} = \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c} \mathbf{v}^\top \\ \hline -\frac{\gamma}{c} \mathbf{v} & \mathbf{I} + (\gamma - 1) \frac{\mathbf{v} \mathbf{v}^\top}{v^2} \end{array} \right) \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix}, \quad (1.37)$$

i.e. the new transformation matrix

$$\boxed{\mathbf{L}(\mathbf{v}) \stackrel{\text{def}}{=} \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c} \mathbf{v}^\top \\ \hline -\frac{\gamma}{c} \mathbf{v} & \mathbf{I} + (\gamma - 1) \frac{\mathbf{v} \mathbf{v}^\top}{v^2} \end{array} \right)} \quad (1.38)$$

is now a *symmetric matrix*, and so

$$\boxed{\begin{aligned} ct' &= \gamma ct - \frac{\gamma}{c} \mathbf{v}^\top \mathbf{x}, \\ \mathbf{x}' &= \mathbf{x} + (\gamma - 1) \frac{\mathbf{v}^\top \mathbf{x}}{v^2} \mathbf{v} - \gamma \mathbf{v} t. \end{aligned}} \quad (1.39)$$

1.2.3 Simultaneity at Different Places

It will be seen that events at different places which are simultaneous for an observer in \mathcal{X} are not in general simultaneous for a moving observer in \mathcal{X}' . This is caused by the finite speed of light. The reference system \mathcal{X}' may move towards the stationary reference system \mathcal{X} with velocity \mathbf{v} . If the two events E_1 and E_2 in \mathcal{X} have the coordinates \vec{x}_1 and \vec{x}_2 , then they are *simultaneous* if $t_1 = t_2$. Do these two events then happen also at the same time for an observer in the reference system \mathcal{X}' ? Due to (1.37),

$$ct'_1 = \gamma ct_1 - \frac{\gamma}{c} \mathbf{v}^\top \mathbf{x}_1$$

and

$$ct'_2 = \gamma ct_2 - \frac{\gamma}{c} \mathbf{v}^\top \mathbf{x}_2.$$

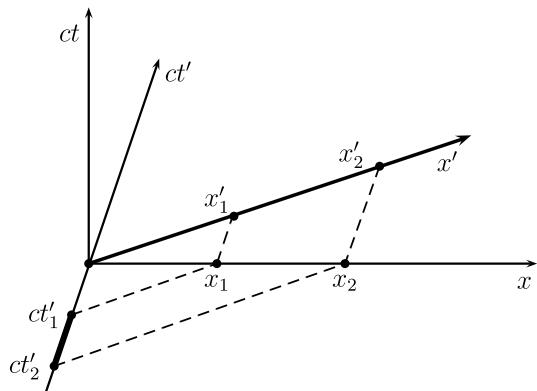
Dividing both equations by c and subtracting one from another yields

$$t'_1 - t'_2 = \gamma(t_1 - t_2) - \frac{\gamma}{c^2} \mathbf{v}^\top (\mathbf{x}_1 - \mathbf{x}_2),$$

so

$$\underline{\underline{t'_2 - t'_1 = \frac{\gamma}{c^2} \mathbf{v}^\top (\mathbf{x}_1 - \mathbf{x}_2)}}.$$

The events \vec{x}'_1 and \vec{x}'_2 are also simultaneous only if the velocity \mathbf{v} is perpendicular to the local difference $\mathbf{x}_1 - \mathbf{x}_2$. Conclusion:

Fig. 1.4 Simultaneity

Events at different locations, which in the reference system \mathcal{X} are simultaneous, need *not* be simultaneous when seen from the reference system \mathcal{X}' .

Example For the special case $\mathbf{v} = [v, 0, 0]^\top$ one only needs to consider the ct - and the x -coordinate (Minkowski diagram, Fig. 1.4). The Lorentz transformation does not change the x_2 - and x_3 -component. It is therefore possible to restrict the consideration to the two-dimensional transformations:

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\frac{\gamma}{c}v \\ -\frac{\gamma}{c}v & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}.$$

The event $\begin{pmatrix} 0 \\ x_1 \end{pmatrix}$ is transformed to

$$\begin{pmatrix} ct'_1 \\ x'_1 \end{pmatrix} = \begin{pmatrix} \gamma & -\frac{\gamma}{c}v \\ -\frac{\gamma}{c}v & \gamma \end{pmatrix} \begin{pmatrix} 0 \\ x_1 \end{pmatrix} = \begin{pmatrix} -\frac{\gamma}{c}vx_1 \\ \gamma x_1 \end{pmatrix}$$

and accordingly the event $\begin{pmatrix} 0 \\ x_2 \end{pmatrix}$, simultaneous in \mathcal{X} , is transformed to $\begin{pmatrix} -\frac{\gamma}{c}vx_2 \\ \gamma x_2 \end{pmatrix}$ in \mathcal{X}' . In Fig. 1.4, the difference on the ct' -axis is then

$$ct'_2 - ct'_1 = \frac{\gamma}{c}v(x_1 - x_2) \neq 0,$$

if $v \neq 0$ and $x_1 \neq x_2$, i.e. in the moving reference system \mathcal{X}' the two events E_1 and E_2 are no longer simultaneous.

1.2.4 Length Contraction of Moving Bodies

Einstein (1905) was the first who completely demonstrated that length contraction is an effect due to the change in the notions of space, time and simultaneity brought about by special relativity. Length contraction can simply be derived from

the Lorentz transformation. Let the reference system \mathcal{X} be stationary and the reference system \mathcal{X}' moving towards it with the speed v . A yardstick has in the stationary system the two endpoints \mathbf{x}_1 and \mathbf{x}_2 . Its resting length is

$$\mathbf{l}_0 = \mathbf{x}_2 - \mathbf{x}_1, \quad (1.40)$$

i.e.

$$l_0^2 = (\mathbf{x}_2 - \mathbf{x}_1)^\top (\mathbf{x}_2 - \mathbf{x}_1). \quad (1.41)$$

At time t' , the endpoints of the yardstick in the moving reference system \mathcal{X}' have the coordinates \mathbf{x}'_1 and \mathbf{x}'_2 . With (1.38) we obtain

$$\mathbf{l}_0 = \mathbf{x}_2 - \mathbf{x}_1 = \mathbf{A}(\mathbf{x}'_2 - \mathbf{x}'_1) \stackrel{\text{def}}{=} \mathbf{A}\mathbf{l}'. \quad (1.42)$$

It follows that

$$l_0^2 = \mathbf{l}_0^\top \mathbf{l}_0 = \mathbf{l}'^\top \mathbf{A}^2 \mathbf{l}'. \quad (1.43)$$

With (1.32) one gets

$$l_0^2 = \mathbf{l}'^\top \left(\mathbf{I} + (\gamma^2 - 1) \frac{\mathbf{v} \mathbf{v}^\top}{v^2} \right) \mathbf{l}' = \mathbf{l}'^\top \mathbf{l}' + (\gamma^2 - 1) \frac{(\mathbf{v}^\top \mathbf{l}')^2}{v^2}. \quad (1.44)$$

In the product $\mathbf{v}^\top \mathbf{l}'$, only the component \mathbf{l}'_{\parallel} of \mathbf{l}' parallel to the velocity \mathbf{v} comes into effect, i.e. it is true that $\mathbf{v}^\top \mathbf{l}' = \mathbf{v}^\top \mathbf{l}'_{\parallel}$. Thus,

$$l_0^2 = l'^2 + (\gamma^2 - 1) \frac{(\mathbf{v}^\top \mathbf{l}'_{\parallel})^2}{v^2} = l'^2 + (\gamma^2 - 1) l'_{\parallel}^2,$$

so

$$\underline{\underline{l'^2}} = l_0^2 - (\gamma^2 - 1) l'_{\parallel}^2. \quad (1.45)$$

Since always $\gamma^2 - 1 \geq 0$, it follows from (1.45) that

$$\underline{\underline{l'}} \leq l_0. \quad (1.46)$$

The result of the yardstick length measurement therefore depends on the reference system in which the length measurement was made.

If the yardstick is parallel to the velocity \mathbf{v} , then $\mathbf{l}' = \mathbf{l}'_{\parallel}$ and (1.45) becomes $\gamma l' = l_0$, or

$$l' = \frac{l_0}{\gamma} = l_0 \sqrt{1 - \frac{v^2}{c^2}},$$

(1.47)

i.e. as $v \rightarrow c$ one gets $l' \rightarrow 0$. For example, if the velocity is $v = 0.8c$ (i.e. 80 % of the speed of light), then the length is $l' = 0.6l_0$.

1.2.5 Time Dilation

Time dilation is an actual difference of elapsed time between two events as measured by observers moving relative to each other. Using (1.19), it is

$$\Delta t = \gamma \Delta t', \quad (1.48)$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \geq 1.$$

Let t' be the time of a light clock, measured by an observer in the moving inertial frame \mathcal{X}' after covering the distance between the mirrors. Then t is the time required for the light to cover the distance between the two mirrors of the *moving* clock, which is measured by an observer at rest in \mathcal{X} . The faster the clock moves, i.e. the greater v is, the longer this time. If, for example, for the moving observer $t' = 1$ second is elapsed, for the stationary observer $t = \gamma \geq 1$ seconds have passed. If v so large that $\gamma = 20$, then, for example, for the stationary observer $t = 20$ years have passed, and for the moving observer only $t' = 1$ year is gone!

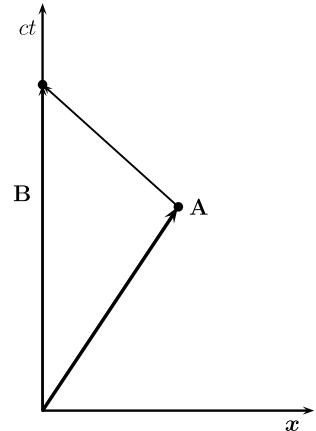
Twin Paradox There are two twins **A** and **B** on Earth. **A** starts in a rocket and flies away with high speed, while **B** stays on Earth. During the flight, **A** is aging more slowly than **B**. After some time the rocket is slowed down and returns with **A** at high speed back to Earth. During the flight **A** has aged less than **B**, which remained at rest on Earth, Fig. 1.5. Now comes the paradox: The velocities are relative! One could take also the twin **A** in its entrained coordinate system as stationary and consider **B** as moving with large speeds. That's right; but one key difference is that the twin **A** is not always in the same uniformly moving inertial frame since, at the turning point where the return begins, the inertial system changes! This is not the case for **B**. He always stays in the same inertial frame. Therefore, there is no paradox.

1.3 Invariance of the Quadratic Form

The Michelson–Morley experiment says that in any reference system the light propagates in all directions at the same speed c . If at the origin $\mathbf{x} = \mathbf{o}$ of \mathcal{X} a flash of light is ignited, it propagates with the speed of light c spherically. After the time t , the light signal reaches all points of the sphere of radius ct . To points on the sphere applies:

$$x_1^2 + x_2^2 + x_3^2 = (ct)^2, \quad \text{i.e.} \quad (ct)^2 - x_1^2 - x_2^2 - x_3^2 = 0. \quad (1.49)$$

If the origins of the two reference systems \mathcal{X} and \mathcal{X}' at the ignition time $t = t' = 0$ of the light flash are at the same space point $\mathbf{x}(t = 0) = \mathbf{X}'(t' = 0) = \mathbf{o}$, the light in

Fig. 1.5 Twin paradox

the reference system \mathcal{X}' is also spreading according to the following law:

$$(ct')^2 - x_1'^2 - x_2'^2 - x_3'^2 = 0. \quad (1.50)$$

So this quantity is *invariant*.

One can imagine (1.49) generated by the quadratic form

$\vec{x}^\top M \vec{x} = 0,$

(1.51)

with the Minkowski matrix

$$M \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and the *four-dimensional* vector

$$\vec{x} \stackrel{\text{def}}{=} \begin{pmatrix} ct \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Minkowski was the first who represented Einstein's Special Theory of Relativity using four-dimensional *spacetime* vectors.

1.3.1 Invariance with Respect to Lorentz Transformation

Now the invariance of the quadratic form (1.51) to a Lorentz transformation is examined. If $\vec{x}' = L\vec{x}$, then

$$\vec{x}'^\top M \vec{x}' = \vec{x}^\top L^\top M L \vec{x}. \quad (1.52)$$

Then for the matrix product $L^\top M L$ with the help of (1.31) and (1.34) one obtains

$$\begin{aligned} L^\top M L &= \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c} v^\top \\ \hline -\frac{\gamma}{c} v & A \end{array} \right) \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c} v^\top \\ \hline \frac{\gamma}{c} v & -A \end{array} \right) \\ &= \left(\begin{array}{c|c} \gamma^2 - \frac{\gamma^2}{c^2} v^\top v & \frac{\gamma}{c} v^\top A - \frac{\gamma^2}{c} v^\top \\ \hline \frac{\gamma}{c} A v - \frac{\gamma^2}{c} v & \frac{\gamma^2}{c^2} v v^\top - A^2 \end{array} \right) = \begin{pmatrix} 1 & o^\top \\ o & -I \end{pmatrix} = M. \end{aligned}$$

So for the quadratic forms, in fact, the following is valid:

$$\vec{x}'^\top M \vec{x}' = \vec{x}^\top M \vec{x},$$

i.e. they are invariant with respect to a Lorentz transformation!

When dealing with the quadratic form, the propagation of light was considered till now. For this the quadratic form $\vec{x}^\top M \vec{x}$ is equal to zero. Considering, however, the movement of a particle, the light will spread faster than the particle, i.e. it will always be true that

$$(ct)^2 > \vec{x}^\top \vec{x}$$

or

$$(ct)^2 - \vec{x}^\top \vec{x} = \vec{x}^\top M \vec{x} > 0.$$

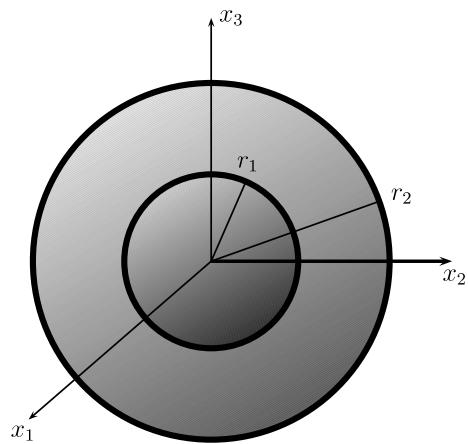
If we denote by $\Delta \vec{x}$ the travelled way between two events and the elapsed time by Δt , we will obtain the four-dimensional spacetime interval Δs from

$$\Delta s^2 = \Delta \vec{x}^\top M \Delta \vec{x}. \quad (1.53)$$

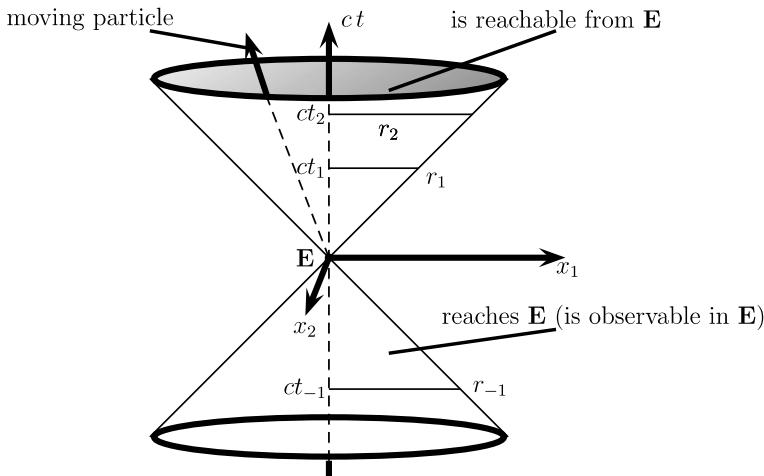
The “distance” Δs between the two events is an *invariant interval* in the four-dimensional spacetime. Since the right-hand side of (1.53) is invariant with respect to a Lorentz transformation, Δs is independent of the chosen inertial frame, it has always the same length. So the theory of relativity does not relativise everything! Δs^2 is negative when the distance $\Delta \vec{x}$ is so big that no light signal can traverse the distance in finite time. This possibility will be examined in the following section.

1.3.2 Light Cone

A flash of light at time $t_0 = 0$ is spreading spherically in the three-dimensional space with the speed of light c . At the time $t_1 > t_0$, the light has reached the surface of a

Fig. 1.6 Light propagation

sphere of radius $r_1 = ct_1$. At time $t_2 > t_1$, a spherical surface with a larger radius $r_2 = ct_2$ is reached, and so forth as shown in Fig. 1.6. In this two-dimensional image, the light wave through space is illustrated by a circle that expands with the speed of light. One can implement this movement of light waves in a space-time diagram in which the time coordinate ct is vertical and two of the three spatial coordinates, e.g. x_1 and x_2 , are displayed horizontally (Fig. 1.7). Due to the fact that the time coordinate is represented by the time t multiplied by the speed of light c , the photons move in straight lines on this diagram, which are sloped at 45° . For photons the possible paths are in an open top cone whose walls have a slope of 45° . The speed of a moving particle is always less than the speed of light, therefore the path must always be run within the light cone with a slope which is always less than 45° to the time axis.

**Fig. 1.7** Light cone

The photons which reach the point \mathbf{E} at a time before $t_0 = 0$, e.g. in the time interval of $t_{-1} < t_0$, all come from a ball of radius $ct_{-1} = r_{-1}$, etc. Altogether we obtain again a cone as in Fig. 1.7, the cone of the events which can reach the event \mathbf{E} , i.e. these events are observable from \mathbf{E} .

In the Minkowski spacetime of special relativity, in any event the light cones are aligned in parallel; the central axes of all light cones are parallel to the time axis. In the general theory of relativity, this is no longer always the case due to the curvature of space, i.e. the central axes of the light cones are not always parallel to the time axis.

1.3.3 Proper Time

If one looks at a completely arbitrarily moving clock from any inertial frame, then one can interpret this movement as uniform at any instant of time. If we introduce a coordinate system which is permanently connected to the clock, then this is again an inertial frame. In the infinitesimal time interval dt , measured with the clock of the observer in the inertial system \mathcal{X} , the moving clock covers a distance of $(dx_1^2 + dx_2^2 + dx_3^2)^{1/2}$. In the inertial system \mathcal{X}' , permanently connected to the clock, the clock does not move, so $dx'_1 = dx'_2 = dx'_3 = 0$, but the elapsed displayed time is dt' . The invariant quadratic form

$$ds^2 = d\vec{x}^\top \mathbf{M} d\vec{x} = d\vec{x}'^\top \mathbf{M} d\vec{x}'$$

computed for

$$d\vec{x}^\top = [c dt, dx_1, dx_2, dx_3] \quad \text{and} \quad d\vec{x}'^\top = [c dt', 0, 0, 0]$$

gives

$$ds^2 = c^2 dt^2 - dx_1^2 - dx_2^2 - dx_3^2 = c^2 dt'^2, \quad (1.54)$$

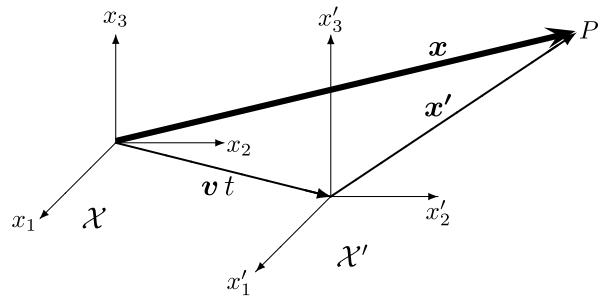
so

$$\begin{aligned} dt' &= \frac{1}{c} ds = \frac{1}{c} \sqrt{c^2 dt^2 - dx_1^2 - dx_2^2 - dx_3^2} \\ &= dt \sqrt{1 - \frac{dx_1^2 + dx_2^2 + dx_3^2}{c^2 dt^2}}, \end{aligned}$$

and taking

$$v^2 = \frac{dx_1^2 + dx_2^2 + dx_3^2}{dt^2},$$

Fig. 1.8 Relativistic velocity addition



where v is the velocity of the moving clock relative to the observer, finally yields the relationship (1.19)

$$dt' = dt \sqrt{1 - \frac{v^2}{c^2}}. \quad (1.55)$$

So, if the stationary clock of the observer shows the time interval $t_2 - t_1$, then the moving clock shows the interval $t'_2 - t'_1$ of the *proper time*,

$$t'_2 - t'_1 = \int_{t_1}^{t_2} \sqrt{1 - \frac{v(t)^2}{c^2}} dt. \quad (1.56)$$

The proper time interval of a moving mass is due to (1.55) and (1.56) always smaller than the time interval in the stationary system. In general, the proper time of a moving mass is called τ instead of t' . Due to (1.54), the proper time is thus

$$\underline{\underline{d\tau = ds/c}}. \quad (1.57)$$

1.4 Relativistic Velocity Addition

1.4.1 Galilean Addition of Velocities

Galilei observed that if a ship is moving relative to the shore at velocity v , and a sailor is moving with velocity u measured on the ship, calculating the velocity of the sailor measured on the shore is what is meant by the addition of the velocities v and u . When both the sailor and the ship are moving slowly compared to light, it is accurate enough to use the vector sum

$$\mathbf{w} = \mathbf{v} + \mathbf{u}$$

where \mathbf{w} is the velocity of the sailor relative to the shore.

Consider the two inertial systems \mathcal{X} and \mathcal{X}' in Fig. 1.8. The reference system \mathcal{X}' is moving with respect to the other reference system \mathcal{X} with the velocity \mathbf{v} . Suppose

that the vector \mathbf{x}' describes the motion of a point P in the inertial system \mathcal{X}' , and this point is moving with respect to the reference system \mathcal{X}' with the velocity $\mathbf{u} = \frac{d\mathbf{x}'}{dt'}$. What is the speed $\mathbf{w} \stackrel{\text{def}}{=} \frac{d\mathbf{x}}{dt}$ of the point P relative to the reference system \mathcal{X} ?

It is

$$\mathbf{w} = \frac{d\mathbf{x}}{dt} = \frac{d\mathbf{x}}{dt'} \left(\frac{dt}{dt'} \right)^{-1}. \quad (1.58)$$

Between \vec{x} and \vec{x}' the relationship $\vec{x}' = \mathbf{L}(v)\vec{x}$ holds, or, solving for \vec{x} ,

$$\vec{x} = \mathbf{L}^{-1}(v)\vec{x}'. \quad (1.59)$$

In detail,

$$t = \gamma_v t' + \frac{\gamma_v}{c^2} \mathbf{v}^\top \mathbf{x}', \quad (1.60)$$

$$\mathbf{x} = \mathbf{x}' + (\gamma_v - 1) \frac{\mathbf{v}^\top \mathbf{x}'}{v^2} \mathbf{v} + \gamma_v \mathbf{v} t'. \quad (1.61)$$

From (1.60) follows

$$\frac{dt}{dt'} = \gamma_v + \frac{\gamma_v}{c^2} \mathbf{v}^\top \frac{d\mathbf{x}'}{dt'} = \gamma_v \left(1 + \frac{\mathbf{v}^\top \mathbf{u}}{c^2} \right), \quad (1.62)$$

and from (1.61)

$$\begin{aligned} \frac{d\mathbf{x}}{dt'} &= \frac{d\mathbf{x}'}{dt'} + (\gamma_v - 1) \frac{\mathbf{v} \mathbf{v}^\top}{v^2} \frac{d\mathbf{x}'}{dt'} + \gamma_v \mathbf{v} \\ &= \mathbf{u} + (\gamma_v - 1) \frac{\mathbf{v}^\top \mathbf{u}}{v^2} \mathbf{v} + \gamma_v \mathbf{v}. \end{aligned} \quad (1.63)$$

Equations (1.62) and (1.63) used in (1.58) provide

$$\mathbf{w} = \frac{\mathbf{v} + \frac{1}{\gamma_v} \mathbf{u} + (1 - \frac{1}{\gamma_v}) \frac{\mathbf{v}^\top \mathbf{u}}{v^2} \mathbf{v}}{1 + \frac{\mathbf{v}^\top \mathbf{u}}{c^2}}, \quad (1.64)$$

or, after addition of $\mathbf{u} - \mathbf{u} = \mathbf{o}$ in the numerator,

$$\mathbf{w} = \frac{\mathbf{v} + \mathbf{u} + (\frac{1}{\gamma_v} - 1)(\mathbf{u} - \frac{\mathbf{v}^\top \mathbf{u}}{v^2} \mathbf{v})}{1 + \frac{\mathbf{v}^\top \mathbf{u}}{c^2}}. \quad (1.65)$$

So this is the speed of the point P with respect to the reference system \mathcal{X} . For a double vectorial product one has

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a}^\top \mathbf{c}) \mathbf{b} - (\mathbf{a}^\top \mathbf{b}) \mathbf{c}. \quad (1.66)$$

This allows one to transform the last parenthesis in the numerator of (1.65) as follows:

$$\mathbf{u} - \frac{\mathbf{v}^\top \mathbf{u}}{v^2} \mathbf{v} = \frac{\mathbf{v}^\top \mathbf{v}}{v^2} \mathbf{u} - \frac{\mathbf{v}^\top \mathbf{u}}{v^2} \mathbf{v} = \frac{1}{v^2} (\mathbf{v} \times (\mathbf{u} \times \mathbf{v})),$$

i.e. instead of (1.65) one can also write

$$\mathbf{w} = \frac{\mathbf{v} + \mathbf{u} + \frac{1}{v^2} (\frac{1}{\gamma_v} - 1) (\mathbf{v} \times (\mathbf{u} \times \mathbf{v}))}{1 + \frac{\mathbf{v}^\top \mathbf{u}}{c^2}}.$$

(1.67)

If the two velocities \mathbf{v} and \mathbf{u} are parallel, then $\mathbf{v} \times \mathbf{u} = \mathbf{o}$ and from (1.67) it is clear that the sum of the two velocities is

$$\mathbf{w} = \frac{\mathbf{v} + \mathbf{u}}{1 + \frac{\mathbf{v}^\top \mathbf{u}}{c^2}}.$$
(1.68)

If, in contrast, the two velocities \mathbf{v} and \mathbf{u} are perpendicular to each other, then $\mathbf{v}^\top \mathbf{u} = \mathbf{o}$ and from (1.65) it follows that the sum of the two velocities is

$$\mathbf{w} = \mathbf{v} + \frac{1}{\gamma_v} \mathbf{u}.$$
(1.69)

In a further special case which is almost exclusively treated in textbooks and in which both the vector \mathbf{u} and the vector \mathbf{v} have only one component in the x_1 -direction, i.e.

$$\mathbf{u} = \begin{pmatrix} u_1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ 0 \\ 0 \end{pmatrix},$$

one has $w_2 = w_3 = 0$ and

$$\mathbf{w}_1 = \frac{v_1 + u_1}{1 + \frac{u_1 v_1}{c^2}}.$$
(1.70)

1.5 Lorentz Transformation of the Velocity

How does one write the basic laws of mechanics so that they remain invariant under a Lorentz transformation? We start from the transformation equation

$$\vec{x}' = \mathbf{L} \vec{x} \tag{1.71}$$

with the four-dimensional spacetime vector $\vec{x} \stackrel{\text{def}}{=} \begin{pmatrix} ct \\ x \end{pmatrix}$ and the transformation matrix (1.38)

$$\mathbf{L} = \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c} \mathbf{v} \\ \hline -\frac{\gamma}{c} \mathbf{v}^\top & \mathbf{I} + (\gamma - 1) \frac{\mathbf{v} \mathbf{v}^\top}{v^2} \end{array} \right).$$

Differentiating (1.71) with respect to time t' , we obtain

$$\frac{d\vec{x}'}{dt'} = \left(\frac{c}{\frac{dx'}{dt'}} \right) = L \frac{d\vec{x}}{dt} \frac{dt}{dt'}.$$

Using (1.35) with $\mathbf{u} \stackrel{\text{def}}{=} \frac{dx}{dt}$ (which now has a different \mathbf{u} than in the previous section),

$$\frac{dt'}{dt} = \frac{d}{dt} \left(-\frac{\gamma}{c^2} \mathbf{v}^\top \mathbf{x} + \gamma t \right) = -\frac{\gamma}{c^2} \mathbf{v}^\top \mathbf{u} + \gamma. \quad (1.72)$$

Thus we obtain

$$\begin{pmatrix} c \\ \mathbf{u}' \end{pmatrix} = L \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \frac{1}{\gamma} \frac{1}{1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2}}. \quad (1.73)$$

In (1.73), we see that the velocity vector $\begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}$ is *not* transformed into the velocity vector $\begin{pmatrix} c \\ \mathbf{u}' \end{pmatrix}$ using a Lorentz transformation matrix! For this to be the case, the definition of the velocity has to be modified. For this purpose, a short interim statement:

The second block row of (1.73) provides

$$\mathbf{u}' = \frac{1}{\gamma(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2})} (\mathbf{A}\mathbf{u} - \gamma \mathbf{v}).$$

Taking the scalar product of this vector with itself, we obtain

$$\mathbf{u}'^2 \stackrel{\text{def}}{=} \mathbf{u}'^\top \mathbf{u}' = \left(\frac{1}{\gamma(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2})} \right)^2 (\mathbf{u}^\top \mathbf{A}^\top - \gamma \mathbf{v}^\top)(\mathbf{A}\mathbf{u} - \gamma \mathbf{v}).$$

Since $\mathbf{A} = \mathbf{A}^\top$ and $\mathbf{A}^2 = \mathbf{I} + \frac{\gamma^2}{c^2} \mathbf{v} \mathbf{v}^\top$, the above is

$$\begin{aligned} \mathbf{u}'^2 &= \frac{1}{\gamma^2(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2})^2} \left(u^2 - \frac{\gamma^2}{c^2} (\mathbf{u}^\top \mathbf{v})^2 - 2\gamma^2 (\mathbf{u}^\top \mathbf{v})^2 + \gamma^2 v^2 \right) \\ &= \frac{1}{(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2})^2} \left(\frac{1}{\gamma^2} u^2 - \frac{1}{c^2} (\mathbf{u}^\top \mathbf{v})^2 - 2(\mathbf{u}^\top \mathbf{v})^2 + v^2 \right). \end{aligned} \quad (1.74)$$

With

$$\frac{u^2}{\gamma^2} = \left(1 - \frac{v^2}{c^2} \right) u^2 = u^2 - \frac{v^2 u^2}{c^2}$$

from (1.74) (without γ) one gets

$$\mathbf{u}'^2 = \frac{1}{(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2})^2} \left(u^2 + \frac{-(\mathbf{u}^\top \mathbf{v})^2 - v^2 u^2}{c^2} - 2(\mathbf{u}^\top \mathbf{v})^2 + v^2 \right). \quad (1.75)$$

With the help of (1.75) one obtains

$$\begin{aligned} 1 - \frac{u'^2}{c^2} &= \frac{1}{(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2})^2} \left(1 - \frac{u^2}{c^2} - \frac{v^2}{c^2} + \frac{v^2 u^2}{c^4} \right) \\ &= \frac{1}{(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2})^2} \left(1 - \frac{u^2}{c^2} \right) \left(1 - \frac{v^2}{c^2} \right), \end{aligned}$$

and this implies

$$\sqrt{1 - \frac{u'^2}{c^2}} = \sqrt{1 - \frac{u^2}{c^2}} \sqrt{1 - \frac{v^2}{c^2}} / \left(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2} \right),$$

or with

$$\gamma_u \stackrel{\text{def}}{=} \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}$$

and

$$\gamma_{u'} \stackrel{\text{def}}{=} \frac{1}{\sqrt{1 - \frac{u'^2}{c^2}}},$$

we finally get

$$\underline{\underline{\gamma \left(1 - \frac{\mathbf{v}^\top \mathbf{u}}{c^2} \right)}} = \frac{\gamma_{u'}}{\gamma_u}. \quad (1.76)$$

Inserting (1.76) into (1.73), we obtain

$$\underline{\underline{\left(\begin{array}{c} c \\ \mathbf{u}' \end{array} \right)}} = \mathbf{L} \left(\begin{array}{c} c \\ \mathbf{u} \end{array} \right) \frac{\gamma_u}{\gamma_{u'}}, \quad (1.77)$$

or

$$\left(\begin{array}{c} \gamma_{u'} c \\ \gamma_{u'} \mathbf{u}' \end{array} \right) = \mathbf{L} \left(\begin{array}{c} \gamma_u c \\ \gamma_u \mathbf{u} \end{array} \right).$$

(1.78)

The thus modified new velocity vector

$$\vec{\mathbf{u}} \stackrel{\text{def}}{=} \gamma_u \left(\begin{array}{c} c \\ \mathbf{u} \end{array} \right) \quad (1.79)$$

now is transformed by a Lorentz transformation \mathbf{L} into the velocity vector $\vec{\mathbf{u}}'$:

$$\vec{\mathbf{u}}' = \mathbf{L} \vec{\mathbf{u}}.$$

(1.80)

The so-defined velocity $\vec{\mathbf{u}}$ is much better suited for the formulation of physics laws, as they have the same shape in every inertial frame. With \mathbf{u} this would not be the

case. For clarification we re-emphasize that a particle or a focal point moves with the velocity \mathbf{u} in an inertial system \mathcal{X}' , which itself moves or can move at the speed \mathbf{v} with respect to another reference system. That is the difference between \mathbf{u} and \mathbf{v} !

By the way, the quadratic form for the velocity

$$\vec{\mathbf{u}}^\top \mathbf{M} \vec{\mathbf{u}} = \gamma_u^2 c^2 - \gamma_u^2 \mathbf{u}^\top \mathbf{u} = \frac{c^4}{c^2 - u^2} - \frac{c^2 u^2}{c^2 - u^2} = c^2 \quad (1.81)$$

is, of course, invariant with respect to a Lorentz transformation because—as one can easily show—even $\vec{\mathbf{u}}'^\top \mathbf{M} \vec{\mathbf{u}}' = c^2$ holds.

In (1.57), the proper time $d\tau = \frac{1}{c} ds$ was introduced. We have

$$ds^2 = \vec{\mathbf{x}}^\top \mathbf{M} \vec{\mathbf{x}} = c^2 dt^2 - \mathbf{x}^\top \mathbf{x} = c^2 dt^2 \left(1 - \frac{1}{c^2} \frac{d\mathbf{x}^\top}{dt} \cdot \frac{d\mathbf{x}}{dt} \right),$$

so with $\frac{d\mathbf{x}}{dt} = \mathbf{u}$

$$dt = dt \left(1 - \frac{1}{c^2} \mathbf{u}^\top \mathbf{u} \right)^{\frac{1}{2}} = dt \left(1 - \frac{u^2}{c^2} \right)^{\frac{1}{2}},$$

or with $\gamma_u = (1 - \frac{u^2}{c^2})^{-\frac{1}{2}}$

$$\underline{\underline{dt}} = \gamma_u d\tau. \quad (1.82)$$

This is the same relationship as in the time dilation in (1.73). τ is therefore the time that a comoving clock displays, while t is the time that an observer at rest measures. However, the moving clock must no longer move rectilinearly and uniformly!

In (1.79), one has $\mathbf{u} = \frac{dx}{dt}$. Replacing dt by (1.82) in it gives $\mathbf{u} = \frac{1}{\gamma_u} \frac{dx}{d\tau}$, or equivalently, $\gamma_u \mathbf{u} = \frac{dx}{d\tau}$. Furthermore,

$$d\vec{\mathbf{x}} = \left(\begin{array}{c} c dt \\ d\mathbf{x} \end{array} \right) = \left(\begin{array}{c} c \gamma_u d\tau \\ \gamma_u \frac{dx}{d\tau} \end{array} \right),$$

so

$$\underline{\underline{\frac{d\vec{\mathbf{x}}}{d\tau}}} = \left(\begin{array}{c} \gamma_u c \\ \gamma_u \frac{dx}{d\tau} \end{array} \right) = \left(\begin{array}{c} \gamma_u c \\ \gamma_u \mathbf{u} \end{array} \right) = \underline{\underline{\mathbf{u}}}.$$

When the trajectory in spacetime is parameterized by the proper time τ , $\vec{\mathbf{x}} = \vec{\mathbf{x}}(\tau)$, then $\vec{\mathbf{u}} = \frac{d\vec{\mathbf{x}}}{d\tau}$ is the four-velocity along the trajectory.

1.6 Momentum and Its Lorentz Transformation

Multiplying the equation $\vec{\mathbf{u}}' = \mathbf{L} \vec{\mathbf{u}}$ with the rest mass m_0 , we obtain

$$\left(\begin{array}{c} m_0 \gamma_{u'} c \\ m_0 \gamma_{u'} \mathbf{u}' \end{array} \right) = \mathbf{L} \left(\begin{array}{c} m_0 \gamma_u c \\ m_0 \gamma_u \mathbf{u} \end{array} \right). \quad (1.84)$$

Herein we define as usual the *momentum*

$$\vec{p} \stackrel{\text{def}}{=} m_0 \gamma_u \vec{u} = m_u \vec{u} = m_u \frac{d\vec{x}}{dt} \quad (1.85)$$

by

$$m_u \stackrel{\text{def}}{=} m_0 \gamma_u = \frac{m_0}{\sqrt{1 - \frac{u^2}{c^2}}}.$$

The momentum vector

$$\vec{p} \stackrel{\text{def}}{=} \begin{pmatrix} m_u c \\ \vec{p} \end{pmatrix} = m_0 \vec{u} = m_0 \gamma_u \begin{pmatrix} c \\ \vec{u} \end{pmatrix} \quad (1.86)$$

is due to (1.84) transformed as

$$\boxed{\vec{p}' = L \vec{p}.} \quad (1.87)$$

Also, the quadratic form associated with the momentum vector

$$\vec{p}^\top M \vec{p} = m_0^2 \vec{u}^\top M \vec{u} = m_0^2 c^2 \quad (1.88)$$

is invariant with respect to a Lorentz transformation because also $\vec{p}'^\top M \vec{p}' = m_0^2 c^2$.

1.7 Acceleration and Force

1.7.1 Acceleration

The acceleration is generally defined as the time derivative of speed. Differentiating the modified velocity vector $\vec{u} = \begin{pmatrix} \gamma_u c \\ \gamma_u \vec{u} \end{pmatrix} \in \mathbb{R}^4$ with respect to time t , one receives as the derivative of the second component in this vector

$$\frac{d}{dt}(\gamma_u \vec{u}) = \frac{d\gamma_u}{dt} \vec{u} + \gamma_u \frac{d\vec{u}}{dt}. \quad (1.89)$$

In particular, for $\frac{d\gamma_u}{dt}$ we obtain with

$$\vec{a} \stackrel{\text{def}}{=} \frac{d\vec{u}}{dt} \in \mathbb{R}^3$$

and

$$\frac{du^2}{dt} = \frac{d\vec{u}^\top \vec{u}}{dt} = \frac{d\vec{u}^\top}{dt} \vec{u} + \vec{u}^\top \frac{d\vec{u}}{dt} = 2\vec{u}^\top \vec{a}$$

the result

$$\begin{aligned}\frac{d\gamma_u}{dt} &= \frac{d}{dt} \left(1 - \frac{u^2}{c^2} \right)^{-1/2} = -\frac{1}{2} \left(1 - \frac{u^2}{c^2} \right)^{-3/2} \cdot \frac{d}{dt} \left(1 - \frac{u^2}{c^2} \right) \\ &= -\frac{1}{2} \gamma_u^3 \cdot \frac{-2\mathbf{u}^\top}{c^2} \cdot \mathbf{a},\end{aligned}\quad (1.90)$$

i.e.

$$\underline{\underline{\frac{d\gamma_u}{dt}}} = \underline{\underline{\frac{\gamma_u^3}{c^2} \cdot \mathbf{u}^\top \mathbf{a}}}.\quad (1.91)$$

With (1.91) one obtains for (1.89)

$$\underline{\underline{\frac{d}{dt}(\gamma_u \mathbf{u})}} = \underline{\underline{\frac{\gamma_u^3}{c^2} \cdot \mathbf{u}^\top \mathbf{a} \cdot \mathbf{u} + \gamma_u \cdot \mathbf{a}}}.\quad (1.92)$$

Differentiating the velocity transformation equation (1.78),

$$\begin{pmatrix} \gamma_{u'} c \\ \gamma_{u'} \mathbf{u}' \end{pmatrix} = \mathbf{L} \begin{pmatrix} \gamma_u c \\ \gamma_u \mathbf{u} \end{pmatrix},$$

with respect to time t' , we obtain

$$\frac{d}{dt'} \begin{pmatrix} \gamma_{u'} c \\ \gamma_{u'} \mathbf{u}' \end{pmatrix} = \mathbf{L} \cdot \frac{d}{dt} \begin{pmatrix} \gamma_u c \\ \gamma_u \mathbf{u} \end{pmatrix} \cdot \frac{dt}{dt'}.\quad (1.93)$$

From (1.72) and (1.76) follows

$$\frac{dt}{dt'} = \frac{\gamma_u}{\gamma_{u'}}.\quad (1.94)$$

This, used in (1.93), results for the newly defined four-dimensional acceleration vector

$\bar{\mathbf{a}} \stackrel{\text{def}}{=} \gamma_u \cdot \frac{d}{dt} \vec{\mathbf{u}} \in \mathbb{R}^4$

(1.95)

in the Lorentz transformation of the acceleration vector $\vec{\mathbf{a}}$

$\bar{\mathbf{a}}' = \mathbf{L} \bar{\mathbf{a}}.$

(1.96)

The acceleration vector defined in (1.95),

$$\bar{\mathbf{a}} = \gamma_u \cdot \frac{d}{dt} \begin{pmatrix} \gamma_u c \\ \gamma_u \mathbf{u} \end{pmatrix},\quad (1.97)$$

is therefore suitable to formulate physical laws in relativistic form! With $dt = \gamma_u d\tau$ and $\vec{u} = d\vec{x}/d\tau$ one can write for (1.97)

$$\vec{a} = \frac{d\vec{u}}{d\tau} = \frac{d^2\vec{x}}{d\tau^2}. \quad (1.98)$$

The vector \vec{a} is also obtained using (1.91) and (1.92) as

$$\vec{a} = \left(\begin{array}{c} \frac{\gamma_u^4}{c} \cdot \mathbf{u}^\top \mathbf{a} \\ \frac{\gamma_u^4}{c^2} \cdot \mathbf{u}^\top \mathbf{a} \cdot \mathbf{u} + \gamma_u^2 \cdot \mathbf{a} \end{array} \right). \quad (1.99)$$

If $\vec{u}^\top \mathbf{M} \vec{u} = c^2$ is differentiated with respect to the proper time τ , we obtain

$$\frac{d\vec{u}^\top}{d\tau} \mathbf{M} \vec{u} + \vec{u}^\top \mathbf{M} \frac{d\vec{u}}{d\tau} = 2\vec{u}^\top \mathbf{M} \frac{d\vec{u}}{d\tau} = 2\vec{u}^\top \mathbf{M} \vec{a} = 0,$$

i.e. the two four-dimensional vectors \vec{u} and $\mathbf{M}\vec{a}$ are orthogonal in \mathbb{R}^4 !

For each point of time of an arbitrarily accelerated motion, one can always specify a reference system \mathcal{X}' which is an inertial system, named “local inertial system”. We obtain the corresponding Lorentz transformation by selecting $\mathbf{L}(\mathbf{u})$ as the transformation matrix. Then, with

$$\mathbf{A}(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{I} + (\gamma_u - 1) \frac{\mathbf{u}\mathbf{u}^\top}{u^2},$$

it is indeed true that

$$\begin{aligned} \vec{a}' &= \mathbf{L}(\mathbf{u})\vec{a} = \left(\begin{array}{c|c} \gamma_u & -\frac{\gamma_u}{c} \mathbf{u}^\top \\ \hline -\frac{\gamma_u}{c} \mathbf{u} & \mathbf{A}(\mathbf{u}) \end{array} \right) \left(\begin{array}{c} \frac{\gamma_u^4}{c} \cdot \mathbf{u}^\top \mathbf{a} \\ \gamma_u^2 \cdot \mathbf{a} + \frac{\gamma_u^4}{c^2} \cdot \mathbf{u}^\top \mathbf{a} \cdot \mathbf{u} \end{array} \right) \\ &= \left(\begin{array}{c} -\gamma_u^3 \frac{\mathbf{u}^\top \mathbf{a}}{c} - \gamma_u^5 \frac{(\mathbf{u}^\top \mathbf{a})(\mathbf{u}^\top \mathbf{u})}{c^3} + \gamma_u^5 \frac{\mathbf{u}^\top \mathbf{a}}{c} \\ \gamma_u^2 \mathbf{A}(\mathbf{u}) \mathbf{a} + \gamma_u^4 \frac{\mathbf{u}(\mathbf{u}^\top \mathbf{a})}{c^2} + (\gamma_u - 1) \gamma_u^4 \frac{\mathbf{u}(\mathbf{u}^\top \mathbf{u})(\mathbf{u}^\top \mathbf{a})}{c^2 u^2} - \gamma_u^5 \frac{(\mathbf{u}^\top \mathbf{a}) \mathbf{u}}{c^2} \end{array} \right) \\ &= \left(\begin{array}{c} 0 \\ \gamma_u^2 \mathbf{a} + \gamma_u^2 (\gamma_u - 1) \frac{\mathbf{u}^\top \mathbf{a}}{u^2} \mathbf{u} \end{array} \right) = \left(\begin{array}{c} 0 \\ \mathbf{a}' \end{array} \right). \end{aligned}$$

1.7.2 Equation of Motion and Force

The relativistic equation of motion for a particle has to be Lorentz-invariant, and, in the inertial system of the considered particle, Newton's equation of motion has to be true:

$$m_0 \frac{d\mathbf{u}}{dt} = \mathbf{f} \in \mathbb{R}^3. \quad (1.100)$$

Let the accompanying inertial system be \mathcal{X} . Furthermore, suppose \mathcal{X}' is the inertial system which moves relative to \mathcal{X} with the constant speed $\mathbf{u}(t_0)$. The particle rests momentarily at time $t = t_0$ in \mathcal{X}' . The equation of motion (1.100) refers to a point of time and its neighbourhood. For this neighbourhood, $t = t_0 \pm dt$ is the desired arbitrary small speed in \mathcal{X}' . For speeds $v \ll c$ we have (1.100). Hence in \mathcal{X}'

$$m_0 \frac{d\mathbf{u}'}{dt'} = \mathbf{f}' \in \mathbb{R}^3 \quad (1.101)$$

also holds exactly. From (1.101), the relativistic equations of motion in an arbitrary reference system may be derived. In (1.101), m_0 is the rest mass and \mathbf{f}' the three-dimensional force in \mathcal{X}' . Expand the vector \mathbf{f}' in (1.101) to a four-vector and call the result $\vec{\mathbf{f}}'$:

$$m_0 \frac{d}{dt'} \begin{pmatrix} c \\ \mathbf{u}' \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{f}' \end{pmatrix} \stackrel{\text{def}}{=} \vec{\mathbf{f}}'. \quad (1.102)$$

Thus $\vec{\mathbf{f}}'$ is specified in the resting system \mathcal{X}' . In the inertial system \mathcal{X} , in which the mass particle moves with the velocity \mathbf{u} , $\vec{\mathbf{f}}$ is obtained by a Lorentz transformation with $\mathbf{L}(-\mathbf{u})$:

$$\vec{\mathbf{f}} = \mathbf{L}(-\mathbf{u}) \begin{pmatrix} 0 \\ \mathbf{f}' \end{pmatrix} = \begin{pmatrix} \frac{\gamma_u}{c} \mathbf{u}^\top \mathbf{f}' \\ A(\mathbf{u}) \mathbf{f}' \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} f_0 \\ \mathbf{f} \end{pmatrix}. \quad (1.103)$$

The equation

$$m_0 \gamma \frac{d}{dt} \begin{pmatrix} \gamma c \\ \gamma \mathbf{u} \end{pmatrix} = \vec{\mathbf{f}} = \begin{pmatrix} f_0 \\ \mathbf{f} \end{pmatrix},$$

i.e.

$m_0 \vec{\mathbf{a}} = \vec{\mathbf{f}}$

(1.104)

possesses all the desired properties! The four-vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{f}}$ are Lorentz-invariant and, in the inertial frame of the particle, this equation is reduced to Newton's equation of motion

$$m_0 \begin{pmatrix} 0 \\ \frac{d\mathbf{u}'}{dt'} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{f}' \end{pmatrix}.$$

For the last three components of the equation of motion (1.104),

$$\frac{d(m_u \mathbf{u})}{dt} = \frac{1}{\gamma_u} \mathbf{f} \quad (1.105)$$

with the velocity-dependent mass

$$m_u \stackrel{\text{def}}{=} \gamma_u m_0. \quad (1.106)$$

In the theory of relativity, the time derivative of the momentum $m_u \mathbf{u}$ is also interpreted as force. The components f_i of the relativistic equation of motion are thus,

according to (1.103) and (1.104),

$$\underline{\underline{f_0 = \gamma_u \frac{d}{dt} (m_u c)}} = \underline{\underline{\frac{\gamma_u}{c} \mathbf{u}^\top \mathbf{f}'}} \quad (1.107)$$

and

$$\underline{\underline{\mathbf{f} = \gamma_u \frac{d}{dt} (m_u \mathbf{u})}} = \underline{\underline{A(\mathbf{u}) \mathbf{f}'}}. \quad (1.108)$$

1.7.3 Energy and Rest Mass

Equation (1.107) multiplied with c/γ_u provides

$$\frac{d}{dt} (m_u c^2) = \mathbf{u}^\top \mathbf{f}. \quad (1.109)$$

In (1.109), $\mathbf{u}^\top \mathbf{f}$ is the instantaneous power. This is the work per unit of time, done by the force \mathbf{f} . So the left-hand side of (1.109) must be the temporal change of energy, i.e. $m_u c^2 = \gamma_u m_0 c^2$ is energy. We obtain for the *relativistic energy* the most renowned formula of the theory of relativity:

$$\boxed{E = m_u c^2}. \quad (1.110)$$

When $\mathbf{u} = \mathbf{o}$, i.e. when the particle is at rest, $\gamma_u = 1$, and

$$\boxed{E_0 = m_0 c^2} \quad (1.111)$$

is the “rest energy”, Einstein’s famous formula. Weinberg says in [36] about the content of this formula: “If some mass is destroyed (as in radioactive decay or fusion, or fission), then very large quantities of kinetic energy will be liberated, with consequences of well-known importance.”

The four-dimensional momentum vector $\vec{\mathbf{p}}$ is then recognised as the combination of energy and momentum:

$$\vec{\mathbf{p}} = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}. \quad (1.112)$$

For the quadratic form in (1.88)

$$\vec{\mathbf{p}}^\top \mathbf{M} \vec{\mathbf{p}} = m_0^2 c^2$$

we now obtain

$$\vec{\mathbf{p}}^\top \mathbf{M} \vec{\mathbf{p}} = (E/c, \mathbf{p}^\top) \begin{pmatrix} E/c \\ -\mathbf{p} \end{pmatrix} = E^2/c^2 - p^2 = m_0^2 c^2,$$

i.e.

$$E = \sqrt{(m_0 c^2)^2 + p^2 c^2}. \quad (1.113)$$

For high velocities the momentum term $p^2 c^2$ dominates here: $E = pc$ as for neutrinos and particles in accelerators (CERN). For small velocities $u \ll c$, we may use the approximation

$$E \approx m_0 c^2 + \frac{1}{2} m_0 u^2.$$

The first term is the rest energy E_0 . The second term is the classical kinetic energy E_{kin} . The relativistic kinetic energy is

$$E_{\text{kin}} = E - m_0 c^2 = (\gamma_u - 1) m_0 c^2. \quad (1.114)$$

If $v \rightarrow c$, then $E_{\text{kin}} \rightarrow \infty$. Equation (1.114) gives a reason why mass particles cannot be accelerated up to the velocity of light!

1.7.4 Emission of Energy

A body, at rest in \mathcal{X}' with the rest mass $m_{0,\text{before}}$, radiates at a certain time the energy E'_{Emission} in form of light or heat radiation. This radiation is emitted symmetrically so that the total momentum of the radiated energy in \mathcal{X}' is zero, i.e. the body remains at rest during the radiation process. The energy-momentum vector of the radiation in \mathcal{X}' is therefore

$$\begin{pmatrix} \frac{1}{c} E'_{\text{Emission}} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (1.115)$$

Compared with the inertial system \mathcal{X} , the body is moving with the velocity \mathbf{v} . In this inertial system, the total momentum is equal to the momentum of the body $\mathbf{p}_{\text{before}} = \gamma_u m_{0,\text{before}} \mathbf{u}$. After radiation, the body has the momentum $\mathbf{p}_{\text{after}} = \gamma_u m_{0,\text{after}} \mathbf{u}$, and the momentum of radiation is calculated using (1.115) with the Lorentz transformation $\mathbf{L}(-\mathbf{u})$ from \mathcal{X}' to \mathcal{X} :

$$\vec{\mathbf{p}} = \mathbf{L}(-\mathbf{u}) \begin{pmatrix} \frac{1}{c} E'_{\text{Emission}} \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_u \frac{1}{c} E'_{\text{Emission}} \\ \gamma_u \mathbf{u} \frac{1}{c^2} E'_{\text{Emission}} \\ \gamma_u \mathbf{u} \frac{1}{c^2} E'_{\text{Emission}} \\ \gamma_u \mathbf{u} \frac{1}{c^2} E'_{\text{Emission}} \end{pmatrix}.$$

During the emission of radiation, the body has given away the momentum $\gamma_u \mathbf{u} \frac{1}{c^2} E'_{\text{Emission}}$, without changing its speed. This is only possible because the body

has changed its rest mass! Because of the momentum conservation law, one has

$$\gamma_u m_{0,\text{before}} \mathbf{u} = \gamma_u m_{0,\text{after}} \mathbf{u} + \gamma_u \mathbf{u} \frac{1}{c^2} E'_{\text{Emission}}.$$

It follows that

$$m_{0,\text{after}} = m_{0,\text{before}} - \frac{1}{c^2} E'_{\text{Emission}}. \quad (1.116)$$

In Einstein's words: "If a body emits the energy E' in the form of radiation, its mass is reduced by E'/c^2 ."

1.8 Relativistic Electrodynamics

1.8.1 Maxwell's Equations

The magnetic field associated with the induction \mathbf{b} and the electric field with the field strength² \mathbf{e} , the sources of the charge q and the current \mathbf{j} satisfy the Maxwell's equations:³

$$\nabla \times \mathbf{b} = \frac{1}{c} \left(\frac{\partial \mathbf{e}}{\partial t} + \mathbf{j} \right), \quad (1.117)$$

$$\nabla^\top \mathbf{e} = \rho, \quad (1.118)$$

$$\nabla \times \mathbf{e} = -\frac{1}{c} \frac{\partial \mathbf{b}}{\partial t}, \quad (1.119)$$

$$\nabla^\top \mathbf{b} = 0. \quad (1.120)$$

The *Derivative Operator* ∇ is the 3-dimensional column vector

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}$$

and ∇^\top is the 3-dimensional row-vector

$$\nabla^\top = \left(\frac{\partial}{\partial x} \middle| \frac{\partial}{\partial y} \middle| \frac{\partial}{\partial z} \right),$$

²Since the magnetic induction and the electric field strength are *vectors*, they are marked by bold small letters \mathbf{b} and \mathbf{e} .

³In other symbols, $\nabla \times \mathbf{b} = \text{curl } \mathbf{b}$ and $\nabla^\top \mathbf{e} = \text{div } \mathbf{e}$.

so that

$$\nabla^\top \mathbf{e} = \frac{\partial e_x}{\partial x} + \frac{\partial e_y}{\partial y} + \frac{\partial e_z}{\partial z}, \quad (1.121)$$

and

$$\nabla \times \mathbf{b} = -\mathbf{B}_\times \nabla \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial b_z}{\partial y} - \frac{\partial b_y}{\partial z} \\ -\frac{\partial b_z}{\partial x} + \frac{\partial b_x}{\partial z} \\ \frac{\partial b_y}{\partial x} - \frac{\partial b_x}{\partial y} \end{pmatrix}, \quad (1.122)$$

with⁴

$$\mathbf{B}_\times \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -b_z & b_y \\ b_z & 0 & -b_x \\ -b_y & b_x & 0 \end{pmatrix}. \quad (1.123)$$

Equations (1.117) and (1.118) can be summed up after minor changes in the following four equations:

$$\begin{aligned} \frac{\partial e_x}{\partial x} + \frac{\partial e_y}{\partial y} + \frac{\partial e_z}{\partial z} &= \rho, \\ -\frac{1}{c} \frac{\partial e_x}{\partial t} - \frac{\partial b_z}{\partial y} + \frac{\partial b_y}{\partial z} &= \frac{1}{c} j_1, \\ -\frac{1}{c} \frac{\partial e_y}{\partial t} + \frac{\partial b_z}{\partial x} - \frac{\partial b_x}{\partial z} &= \frac{1}{c} j_2, \\ -\frac{1}{c} \frac{\partial e_z}{\partial t} - \frac{\partial b_y}{\partial x} + \frac{\partial b_x}{\partial y} &= \frac{1}{c} j_3, \end{aligned} \quad (1.124)$$

or in matrix form,⁵

$$\underbrace{\begin{pmatrix} 0 & \mathbf{e}^\top \\ -\mathbf{e} & \mathbf{B}_\times \end{pmatrix}}_{\stackrel{\text{def}}{=} \mathbf{F}_{B,e}} \underbrace{\gamma \left(-\frac{1}{c} \frac{\partial}{\partial t} \right)}_{\stackrel{\text{def}}{=} \vec{\nabla}} \underbrace{\begin{pmatrix} c\rho \\ -\mathbf{j} \end{pmatrix}}_{\stackrel{\text{def}}{=} \vec{j}} \in \mathbb{R}^4. \quad (1.125)$$

The skew-symmetric matrix $\mathbf{F}_{B,e}$, composed of \mathbf{B}_\times and \mathbf{e} , is called Faraday's matrix. A similarly structured field strength matrix is obtained if (1.119) and (1.120)

⁴An equation $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ cannot in this form be linearly transformed with an invertible transformation matrix \mathbf{T} . But with $\mathbf{A}_\times \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{pmatrix}$ this is possible in the form of the equation $\mathbf{c} = \mathbf{A}_\times \mathbf{b}$: $(\mathbf{T}\mathbf{c}) = (\mathbf{T}\mathbf{A}_\times \mathbf{T}^{-1})(\mathbf{T}\mathbf{b})$.

⁵The factor γ has been added on both sides, so that no difficulties arise later in the invariance of this equation with respect to a Lorentz transformation.

are gathered and summarized in the following equations:

$$\begin{aligned} -\frac{\partial b_x}{\partial x} - \frac{\partial b_y}{\partial y} - \frac{\partial b_z}{\partial z} &= 0, \\ +\frac{1}{c} \frac{\partial b_x}{\partial t} - \frac{\partial e_z}{\partial y} + \frac{\partial e_y}{\partial z} &= 0, \\ +\frac{1}{c} \frac{\partial b_y}{\partial t} + \frac{\partial e_z}{\partial x} - \frac{\partial e_x}{\partial z} &= 0, \\ +\frac{1}{c} \frac{\partial b_z}{\partial t} - \frac{\partial e_y}{\partial x} + \frac{\partial e_x}{\partial y} &= 0, \end{aligned} \quad (1.126)$$

or in matrix form,

$$\underbrace{\begin{pmatrix} 0 & -\mathbf{b}^\top \\ \mathbf{b} & \mathbf{E}_\times \end{pmatrix}}_{\text{def } \mathbf{F}_{E,b}} \underbrace{\gamma \left(-\frac{1}{c} \frac{\partial}{\nabla} \right)}_{\vec{\nabla}} = \mathbf{o} \in \mathbb{R}^4. \quad (1.127)$$

The matrix $\mathbf{F}_{E,b}$ is called the Maxwell's matrix.

So (1.125) and (1.127) contain the Maxwell's equations in a new four-dimensional form:

$$\mathbf{F}_{B,e} \vec{\nabla} = \frac{1}{c} \vec{j} \quad \text{and} \quad \mathbf{F}_{E,b} \vec{\nabla} = \mathbf{o}. \quad (1.128)$$

This form has a great advantage of being invariant when transitioning to another reference system by a Lorentz transformation, i.e. in every inertial system its keeps the same external form. But the quantities appearing in it take different values in each reference system. This is shown in the following.

1.8.2 Lorentz Transformation of the Maxwell's Equations

Multiplying (1.125) from the left with \mathbf{L}^{-1} and inserting $\mathbf{L}^{-1}\mathbf{L} = \mathbf{I}$, we obtain

$$\underbrace{\mathbf{L}^{-1} \mathbf{F}_{B,e} \mathbf{L}^{-1}}_{\mathbf{F}'_{B',e'}} \mathbf{L} \vec{\nabla} = \frac{1}{c} \underbrace{\mathbf{L}^{-1} \vec{j}}_{\vec{j}'} \quad (1.129)$$

For the new Faraday's matrix $\mathbf{F}'_{B',e'}$ we get

$$\begin{aligned} \mathbf{F}'_{B',e'} &= \begin{pmatrix} 0 & e'^\top \\ -\mathbf{e}' & \mathbf{B}'_\times \end{pmatrix} \\ &= \begin{pmatrix} \gamma_v & \frac{\gamma_v}{c} \mathbf{v}^\top \\ \frac{\gamma_v}{c} \mathbf{v} & \mathbf{A}(\mathbf{v}) \end{pmatrix} \begin{pmatrix} 0 & \mathbf{e}^\top \\ -\mathbf{e} & \mathbf{B}_\times \end{pmatrix} \begin{pmatrix} \gamma_v & \frac{\gamma_v}{c} \mathbf{v}^\top \\ \frac{\gamma_v}{c} \mathbf{v} & \mathbf{A}(\mathbf{v}) \end{pmatrix}. \end{aligned} \quad (1.130)$$

After multiplying the three matrices, we obtain for the bottom left corner vector \mathbf{e}' with $A = A(\mathbf{v})$ and $\gamma = \gamma_v$:

$$\begin{aligned}
-\mathbf{e}' &= \frac{\gamma}{c} \mathbf{A} \mathbf{B}_\times \mathbf{v} + \frac{\gamma^2}{c^2} \mathbf{v} \mathbf{e}^\top \mathbf{v} - \gamma \mathbf{A} \mathbf{e} \\
&= \frac{\gamma}{c} \left(\mathbf{B}_\times \mathbf{v} + (\gamma - 1) \underbrace{\frac{\mathbf{v} \mathbf{v}^\top \mathbf{B}_\times \mathbf{v}}{v^2}}_0 \right) + \frac{\gamma^2}{c^2} \mathbf{v} \mathbf{e}^\top \mathbf{v} - \gamma \mathbf{e} - \frac{\gamma(\gamma - 1)}{v^2} \mathbf{v} \mathbf{v}^\top \mathbf{e} \\
&= \frac{\gamma}{c} \mathbf{B}_\times \mathbf{v} - \underbrace{\left(-\frac{\gamma^2}{c^2} + \frac{\gamma^2}{v^2} - \frac{\gamma}{v^2} \right)}_{\frac{1}{v^2}} \mathbf{v}^\top \mathbf{e} \mathbf{v} - \gamma \mathbf{e}, \\
\mathbf{e}' &= \gamma \left(\mathbf{e} - \frac{1}{c} \mathbf{B}_\times \mathbf{v} \right) + \frac{(1 - \gamma)}{v^2} (\mathbf{v}^\top \mathbf{e}) \mathbf{v},
\end{aligned} \tag{1.131}$$

i.e.

$$\boxed{\mathbf{e}' = \gamma \left(\mathbf{e} + \frac{1}{c} \mathbf{v} \times \mathbf{b} \right) + (1 - \gamma) \frac{\mathbf{v}^\top \mathbf{e}}{v^2} \mathbf{v}.}$$

(1.132)

In the matrix $\mathbf{F}'_{B'_x, e'}$ in (1.130), for the right bottom 3×3 sub-matrix \mathbf{B}'_\times after performing the matrix multiplication we obtain

$$\mathbf{B}'_\times = \mathbf{B}_\times + \frac{(\gamma - 1)}{v^2} (\mathbf{B}_\times \mathbf{v} \mathbf{v}^\top + \mathbf{v} \mathbf{v}^\top \mathbf{B}_\times) + \frac{\gamma}{c} (\mathbf{v} \mathbf{e}^\top - \mathbf{e} \mathbf{v}^\top). \tag{1.133}$$

From this matrix equation, the components of the magnetic induction vector \mathbf{b}' can be filtered out with the help of the row or column vectors of the unit matrix (\mathbf{i}_j is the j th column of the 3×3 identity matrix \mathbf{I}_3) as follows:

$$\mathbf{B}'_\times \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -b'_z & b'_y \\ b'_z & 0 & -b'_x \\ -b'_y & b'_x & 0 \end{pmatrix}. \tag{1.134}$$

We thus obtain

$$b'_x = \mathbf{i}_3^\top \mathbf{B}'_\times \mathbf{i}_2, \tag{1.135}$$

$$b'_y = \mathbf{i}_1^\top \mathbf{B}'_\times \mathbf{i}_3, \tag{1.136}$$

$$b'_z = \mathbf{i}_2^\top \mathbf{B}'_\times \mathbf{i}_1. \tag{1.137}$$

Using (1.135) for (1.133) results in

$$\begin{aligned}
b'_x &= \mathbf{i}_3^\top \mathbf{B}'_\times \mathbf{i}_2 = b_x + \frac{(\gamma - 1)}{v^2} [(-b_y v_1 + b_x v_2) v_2 + v_3 (-b_z v_1 + b_x v_3)] \\
&\quad + \frac{\gamma}{c} (v_3 e_y - e_z v_2).
\end{aligned} \tag{1.138}$$

Accordingly, one obtains for the two remaining components

$$\begin{aligned} b'_y &= \mathbf{i}_1^\top \mathbf{B}'_\times \mathbf{i}_3 = b_y + \frac{(\gamma - 1)}{v^2} [(-b_z v_2 + b_y v_3) v_3 + v_1 (b_y v_1 - b_x v_2)] \\ &\quad + \frac{\gamma}{c} (v_1 e_z - e_x v_3) \end{aligned} \quad (1.139)$$

and

$$b'_z = \mathbf{i}_2^\top \mathbf{B}'_\times \mathbf{i}_1 = b_z + \frac{(\gamma - 1)}{v^2} [(b_z v_1 - b_x v_3) v_1 + v_2 (b_z v_2 - b_y v_3)] + \frac{\gamma}{c} (v_2 e_x - e_y v_1). \quad (1.140)$$

Summarizing the last terms in (1.138), (1.139) and (1.140) results in the vector product

$$\frac{\gamma}{c} \mathbf{e} \times \mathbf{v}. \quad (1.141)$$

For the second to last summands one obtains the vector

$$\frac{(\gamma - 1)}{v^2} \begin{pmatrix} (-b_y v_1 + b_x v_2) v_2 + v_3 (-b_z v_1 + b_x v_3) \\ (-b_z v_2 + b_y v_3) v_3 + v_1 (b_y v_1 - b_x v_2) \\ (b_z v_1 - b_x v_3) v_1 + v_2 (b_z v_2 - b_y v_3) \end{pmatrix},$$

and using $0 = b_x v_1^2 - b_x v_1^2$ in the first component, $0 = b_y v_2^2 - b_y v_2^2$ in the second, and $0 = b_z v_3^2 - b_z v_3^2$ in the third,

$$\begin{aligned} &\frac{(\gamma - 1)}{v^2} \begin{pmatrix} (-b_x v_1^2 - b_y v_1 v_2 - b_z v_1 v_3) + (b_x v_1^2 + b_x v_2^2 + b_x v_3^2) \\ (-b_x v_1 v_2 - b_y v_2^2 - b_z v_2 v_3) + (b_y v_1^2 + b_y v_2^2 + b_y v_3^2) \\ (-b_x v_1 v_3 - b_y v_2 v_3 - b_z v_3^2) + (b_z v_1^2 + b_z v_2^2 + b_z v_3^2) \end{pmatrix} \\ &= \frac{(\gamma - 1)}{v^2} (-(\mathbf{b}^\top \mathbf{v}) \mathbf{v} + v^2 \mathbf{b}). \end{aligned} \quad (1.142)$$

Equations (1.141) and (1.142), together with (1.138), (1.139) and (1.140), yield the final result

$$\mathbf{b}' = \gamma \left(\mathbf{b} - \frac{1}{c} \mathbf{v} \times \mathbf{e} \right) + (1 - \gamma) \frac{\mathbf{v}^\top \mathbf{b}}{v^2} \mathbf{v}. \quad (1.143)$$

One gets the same result for \mathbf{e}' and \mathbf{b}' by a Lorentz transformation of $\mathbf{F}_{E,b}$!

The formulas (1.132) and (1.143) are simplified considerably for small velocities $v \ll c$ because then $\gamma \approx 1$ and one obtains

$$\mathbf{e}' = \mathbf{e} + \frac{1}{c} \mathbf{v} \times \mathbf{b}$$

and

$$\mathbf{b}' = \mathbf{b} - \frac{1}{c} \mathbf{v} \times \mathbf{e}.$$

The decomposition of the electromagnetic field in an electric and magnetic field has no absolute significance. If there exists, for example, in a reference system \mathcal{X} a purely electrostatic field with $\mathbf{b} = \mathbf{0}$, then, due to (1.143), a magnetic field $\mathbf{b}' = \frac{\gamma}{c} \mathbf{v} \times \mathbf{e} \neq \mathbf{0}$ exists in a reference system \mathcal{X}' that moves with the speed \mathbf{v} relative to the reference system \mathcal{X} . Physically this means that all charges rest in \mathcal{X} . But these charges move relative to \mathcal{X}' with the velocity \mathbf{v} . So there exists a current in \mathcal{X}' which generates a magnetic field in \mathcal{X}' .

The equations (1.132) and (1.143) can be summarized in one equation:

$$\boxed{\begin{pmatrix} \mathbf{b}' \\ \mathbf{e}' \end{pmatrix} = \underbrace{\begin{pmatrix} \gamma \mathbf{I} + (1 - \gamma) \frac{\mathbf{v} \mathbf{v}^T}{v^2} & -\frac{\gamma}{c} \mathbf{V}_\times \\ \frac{\gamma}{c} \mathbf{V}_\times & \gamma \mathbf{I} + (1 - \gamma) \frac{\mathbf{v} \mathbf{v}^T}{v^2} \end{pmatrix}}_{\mathbf{P}(\mathbf{v})} \begin{pmatrix} \mathbf{b} \\ \mathbf{e} \end{pmatrix} \in \mathbb{R}^6}, \quad (1.144)$$

with

$$\mathbf{V}_\times \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}.$$

The symmetric 6×6 -matrix $\mathbf{P}(\mathbf{v})$ occurring in (1.144) has a formal similarity with the Lorentz matrix \mathbf{L} !

It is, of course, true that $\mathbf{P}(-\mathbf{v}) = \mathbf{P}^{-1}(\mathbf{v})$, for, with

$$\mathbf{V}_\times^2 = \mathbf{v} \mathbf{v}^T - v^2 \mathbf{I}, \quad (1.145)$$

one obtains easily $\mathbf{P}(\mathbf{v}) \mathbf{P}(-\mathbf{v}) = \mathbf{I}$.

In the special case when only one component of the velocity vector in the x -direction is nonzero, i.e. $\mathbf{v} = [v, 0, 0]^T$, one obtains the matrix

$$\mathbf{P}(\mathbf{v}) = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 & 0 & \frac{\gamma v}{c} \\ 0 & 0 & \gamma & 0 & -\frac{\gamma v}{c} & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{\gamma v}{c} & 0 & \gamma & 0 \\ 0 & \frac{\gamma v}{c} & 0 & 0 & 0 & \gamma \end{array} \right). \quad (1.146)$$

This special symmetric 6×6 -matrix can also be found in [33].

1.8.3 Electromagnetic Invariants

For the electromagnetic field quantities \mathbf{e} and \mathbf{b} one can form *invariants* which do not change during the transition to another inertial system. For the scalar product of

the electric field strength \mathbf{e} and the magnetic induction \mathbf{b} we obtain

$$\begin{aligned}\mathbf{e}'^\top \mathbf{b}' &= \left(\gamma \left(\mathbf{e}^\top - \frac{1}{c} (\mathbf{v} \times \mathbf{b})^\top \right) + (1 - \gamma) \frac{\mathbf{v}^\top \mathbf{e}}{v^2} \mathbf{v}^\top \right) \left(\gamma \left(\mathbf{b} + \frac{1}{c} \mathbf{v} \times \mathbf{e} \right) \right. \\ &\quad \left. + (1 - \gamma) \frac{\mathbf{v}^\top \mathbf{b}}{v^2} \mathbf{v} \right) \\ &= \gamma^2 \mathbf{e}^\top \mathbf{b} + [2\gamma(1 - \gamma) + (1 - \gamma^2)] \frac{\mathbf{e}^\top \mathbf{v} \mathbf{v}^\top \mathbf{b}}{v^2} - \frac{\gamma^2}{c^2} \underbrace{(\mathbf{v} \times \mathbf{b})^\top (\mathbf{v} \times \mathbf{e})}_{\mathbf{b}^\top \mathbf{V}_x^\top \mathbf{V}_x \mathbf{e}}.\end{aligned}$$

With $\mathbf{V}_x^\top = -\mathbf{V}_x$ one obtains due to (1.145)

$$\mathbf{V}_x^\top \mathbf{V}_x = -\mathbf{V}_x^2 = v^2 \mathbf{I} - \mathbf{v} \mathbf{v}^\top.$$

This gives

$$\underline{\underline{\mathbf{e}'^\top \mathbf{b}'}} = \gamma^2 \mathbf{e}^\top \mathbf{b} - \frac{\gamma^2}{c^2} (\mathbf{v}^\top \mathbf{e} \mathbf{v}^\top \mathbf{b} + v^2 \mathbf{b}^\top \mathbf{e} - \mathbf{b}^\top \mathbf{v} \mathbf{v}^\top \mathbf{e}) = \left(\gamma^2 - \frac{\gamma^2 v^2}{c^2} \right) \mathbf{e}^\top \mathbf{b} = \underline{\underline{\mathbf{e}^\top \mathbf{b}}},$$

so the scalar product of the electric field intensity \mathbf{e} and the magnetic induction \mathbf{b} is invariant with respect to a Lorentz transformation!

Using a slightly long calculation, one can show that the difference of the squares of the field intensities is also invariant:

$$\underline{\underline{\mathbf{b}'^2 - \mathbf{e}'^2}} = \mathbf{b}^2 - \mathbf{e}^2. \quad (1.147)$$

You can also arrive at electromagnetic invariants in another way, namely with the help of the invariant Faraday's matrix

$$\mathbf{F}_{B,e} = \begin{pmatrix} 0 & \mathbf{e}^\top \\ -\mathbf{e} & \mathbf{B}_x \end{pmatrix}$$

and the invariant Maxwell's matrix

$$\mathbf{F}_{E,b} = \begin{pmatrix} 0 & -\mathbf{b}^\top \\ \mathbf{b} & \mathbf{E}_x \end{pmatrix}.$$

With

$$\mathbf{F}_{B,e}^* \stackrel{\text{def}}{=} \mathbf{M} \mathbf{F}_{B,e} \mathbf{M} = \begin{pmatrix} 0 & -\mathbf{e}^\top \\ \mathbf{e} & \mathbf{B}_x \end{pmatrix},$$

one gets, e.g.

$$\mathbf{F}_{B,e}^* \mathbf{F}_{B,e} = \begin{pmatrix} 0 & -\mathbf{e}^\top \\ \mathbf{e} & \mathbf{B}_x \end{pmatrix} \begin{pmatrix} 0 & \mathbf{e}^\top \\ -\mathbf{e} & \mathbf{B}_x \end{pmatrix} = \begin{pmatrix} \mathbf{e}^\top \mathbf{e} & -\mathbf{e}^\top \mathbf{B}_x \\ -\mathbf{B}_x \mathbf{e} & \mathbf{e} \mathbf{e}^\top + \mathbf{B}_x \mathbf{B}_x \end{pmatrix}$$

$$= \begin{pmatrix} e^2 & & & & s^\top \\ s & \begin{pmatrix} e_x^2 - b_z^2 - b_y^2 & \cdots & & & s^\top \\ \cdots & e_y^2 - b_z^2 - b_x^2 & \cdots & & \cdots \\ \cdots & \cdots & e_z^2 - b_y^2 - b_x^2 & & \end{pmatrix} \end{pmatrix},$$

with the Poynting vector $s \stackrel{\text{def}}{=} \mathbf{e} \times \mathbf{b}$. By forming the trace of this matrix product, i.e. taking the sum of the matrix elements on the main diagonal, one obtains

$$\text{trace}(\mathbf{F}_{B,e}^* \mathbf{F}_{B,e}) = 2e^2 - 2b^2,$$

i.e. the above invariant in the form

$$-\frac{1}{2} \text{trace}(\mathbf{F}_{B,e}^* \mathbf{F}_{B,e}) = b^2 - e^2.$$

(1.148)

One obtains the second invariant $\mathbf{e}^\top \mathbf{b}$ above by taking the trace of the product of the modified Faraday's matrix $\mathbf{F}_{B,e}^*$ and the Maxwell's matrix $\mathbf{F}_{E,b}$:

$$-\frac{1}{4} \text{trace}(\mathbf{F}_{B,e}^* \mathbf{F}_{E,e}) = \mathbf{e}^\top \mathbf{b}.$$

(1.149)

Making invariants of this type later plays a role in the consideration of the singularities of the Schwarzkopf's solution of General Relativity equations.

1.8.4 Electromagnetic Forces

We want to determine the force acting on a charged particle with the charge q , which is in an electromagnetic field moving with the velocity \mathbf{u} relative to an inertial frame \mathcal{X} . Let \mathcal{X}' be the inertial frame in which the particle rests at the moment. In this system, due to $\mathbf{u}' = \mathbf{o}$ and $\mathbf{u}' \times \mathbf{b}' = \mathbf{o}$,

$$m_0 \frac{d\mathbf{u}'}{dt'} = q\mathbf{e}' \in \mathbb{R}^3.$$
(1.150)

Generally, due to the relativity principle, the Lorentz force in \mathcal{X} is

$$\mathbf{f} = q \left(\mathbf{e} + \frac{1}{c} \mathbf{u} \times \mathbf{b} \right) = \frac{q}{c} [\mathbf{e} \mid -\mathbf{B}_\times] \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}^3.$$
(1.151)

The following law expresses how an electromagnetic field acts on a stationary charge q and current \mathbf{j} :

$$\mathbf{f} = q\mathbf{e} + \frac{1}{c} \mathbf{j} \times \mathbf{b} = \frac{1}{c} [\mathbf{e} \mid -\mathbf{B}_\times] \begin{pmatrix} cq \\ \mathbf{j} \end{pmatrix} \in \mathbb{R}^3.$$
(1.152)

\vec{f} is completed to a four-vector \vec{f} as in (1.108)

$$\vec{f} \stackrel{\text{def}}{=} \gamma_u \begin{pmatrix} \mathbf{u}^\top f/c \\ f \end{pmatrix} \in \mathbb{R}^4. \quad (1.153)$$

Equation (1.151) applied to (1.153) yields (with $\mathbf{u}^\top (\mathbf{u} \times \mathbf{b}) = 0$)

$$\begin{aligned} \vec{f} &= q \gamma_u \left(\mathbf{e} + \frac{1}{c} \mathbf{u} \times \mathbf{b} \right) = q \frac{\gamma_u}{c} \underbrace{\begin{pmatrix} 0 & \mathbf{e}^\top \\ \mathbf{e} & -\mathbf{B}_\times \end{pmatrix}}_{\mathbf{M}\mathbf{F}_{B,e}} \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \\ &= \frac{q}{c} \mathbf{M}\mathbf{F}_{B,e} \underbrace{\gamma_u \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}}_{\vec{u}}, \end{aligned} \quad (1.154)$$

so

$$\boxed{\vec{f} = \frac{q}{c} \mathbf{M}\mathbf{F}_{B,e} \vec{u}.} \quad (1.155)$$

With (1.152) one can write instead of (1.155)

$$\boxed{\vec{f} = \frac{1}{c} \mathbf{F}_{B,e}^* \vec{j},} \quad (1.156)$$

in which again

$$\vec{j} = \gamma \begin{pmatrix} cq \\ -\mathbf{j} \end{pmatrix} \in \mathbb{R}^4.$$

Subjecting (1.155) to a linear transformation using the Lorentz matrix $\mathbf{L}(\mathbf{v})$, one obtains

$$\vec{f}' = \mathbf{L}(\mathbf{v}) \vec{f} = \frac{q}{c} \underbrace{\mathbf{L}(\mathbf{v}) \mathbf{M}\mathbf{F}_{B,e}}_{\mathbf{M}\mathbf{F}'_{B',e'}} \underbrace{\mathbf{L}^{-1}(\mathbf{v})}_{\vec{u}'} \underbrace{\mathbf{L}(\mathbf{v}) \vec{u}}_{\vec{u}'}, \quad (1.157)$$

i.e.

$$\boxed{\vec{f}' = \frac{q}{c} \mathbf{M}\mathbf{F}'_{B',e'} \vec{u}'.} \quad (1.158)$$

The power equation is also invariant with respect to a Lorentz transformation!

1.9 The Energy–Momentum Matrix

1.9.1 The Electromagnetic Energy–Momentum Matrix

We want to derive one equation which contains the fundamental dynamic equations of the theory of electricity. This equation should also include the energy theorem and the momentum theorems of electrodynamics. We will find again the therein contained energy–momentum matrix in the main equation of the theory of general relativity.

Due to (1.128), one has

$$\mathbf{F}_{B,e} \vec{\nabla} = \frac{1}{c} \vec{j}$$

and because of (1.156)

$$\vec{f} = \frac{1}{c} \mathbf{F}_{B,e}^* \vec{j}.$$

Is there a matrix such that

$$\vec{f} = \mathbf{T}_{b,e} \vec{\nabla} ? \quad (1.159)$$

We try $\mathbf{T}_{b,e} = \mathbf{F}_{B,e}^* \mathbf{F}_{B,e}$. Then

$$\begin{aligned} \mathbf{T}_{b,e} \vec{\nabla} &= (\mathbf{F}_{B,e}^* \mathbf{F}_{B,e}) \vec{\nabla} = \begin{pmatrix} \mathbf{e}^\top \mathbf{e} & -\mathbf{e}^\top \mathbf{B}_\times \\ -\mathbf{B}_\times \mathbf{e} & \mathbf{e} \mathbf{e}^\top + \mathbf{B}_\times^2 \end{pmatrix} \vec{\nabla} \\ &= \begin{pmatrix} e^2 & s^\top \\ s & \mathbf{e} \mathbf{e}^\top + \mathbf{b} \mathbf{b}^\top - b^2 \mathbf{I}_3 \end{pmatrix} \vec{\nabla} \stackrel{!}{=} \vec{f}, \end{aligned} \quad (1.160)$$

by considering the Poynting vector

$$\mathbf{s} \stackrel{\text{def}}{=} \mathbf{e} \times \mathbf{b} \quad (1.161)$$

and the relation $\mathbf{B}_\times^2 = \mathbf{b} \mathbf{b}^\top - b^2 \mathbf{I}$ in conformance with (1.145). The Poynting vector gives the magnitude and direction of the energy transport in electromagnetic fields. What is the result when differentiating the first line of (1.160)? It is

$$-\frac{1}{c} \frac{\partial(e^2)}{\partial t} + \nabla^\top \mathbf{s} \stackrel{!}{=} \frac{\rho \mathbf{u}^\top \mathbf{f}}{c}.$$

$\frac{1}{c} \frac{\partial(e^2)}{\partial t}$ is proportional to the temporal change of the energy density of the electromagnetic field when no magnetic field is present. But then also $\mathbf{b} = \mathbf{0}$, i.e. $\mathbf{s} = \mathbf{0}$, and thus the whole equation is without a statement. The matter would be different if instead of e^2 in top left corner of the matrix $\mathbf{T}_{b,e}$, one would have the expression $(e^2 + b^2)/2$, as then

$$\frac{1}{2c} \frac{\partial(e^2 + b^2)}{\partial t} = \frac{1}{c} \frac{\partial w}{\partial t},$$

namely, the temporal change of the energy density $w = (e^2 + b^2)/2$ of the electromagnetic field. By adding $(b^2 - e^2)/2$ to the upper left, there arises, in fact, $(e^2 + b^2)/2$! Therefore, now the

Definition: The *electromagnetic energy-momentum matrix* $\mathbf{T}_{b,e}$ has the form

$$\mathbf{T}_{b,e} \stackrel{\text{def}}{=} \mathbf{F}_{B,e}^* \mathbf{F}_{B,e} + \frac{1}{2}(b^2 - e^2) \mathbf{I}_4 = \begin{pmatrix} w & s^\top \\ s & (ee^\top + bb^\top - w\mathbf{I}_3) \end{pmatrix}, \quad (1.162)$$

with

$$w \stackrel{\text{def}}{=} \frac{1}{2}(e^2 + b^2) \quad (1.163)$$

and

$$\mathbf{T}_{b,e} \cdot \vec{\nabla} = \frac{1}{\gamma_v} \vec{f}. \quad (1.164)$$

While doing this, $\vec{\nabla}$ again has the form

$$\vec{\nabla} = \gamma_u \left(-\frac{1}{c} \frac{\partial}{\partial t} \right)$$

and

$$\vec{f} = \gamma_u \left(\frac{\rho}{c} \mathbf{u}^\top \mathbf{f} \right).$$

For the first row of (1.164) now one obtains

$$-\frac{1}{c} \frac{\partial w}{\partial t} + s^\top \nabla = \frac{\rho}{c} \mathbf{u}^\top \mathbf{f},$$

i.e.

$$c \operatorname{div} s = \frac{\partial w}{\partial t} + \rho \mathbf{u}^\top \mathbf{f}. \quad (1.165)$$

On the left is the infinitely small volume unit of the entering or exiting energy flow, and it consists of the temporal change of the energy density $\frac{\partial w}{\partial t}$ and the conversion of the electromagnetic energy into mechanical energy per unit time and volume $\rho \mathbf{u}^\top \mathbf{f}$. So the whole is the *Energy Theorem of Electrodynamics*.

Next we get for the second to fourth components of (1.164):

$$-\frac{1}{c} \frac{\partial s}{\partial t} + (ee^\top) \nabla + (bb^\top) \nabla - (w\mathbf{I}_3) \nabla = \mathbf{f}. \quad (1.166)$$

With the help of Maxwell's equations, we obtain for the first term on the left-hand side:

$$-\frac{1}{c} \frac{\partial s}{\partial t} = \frac{1}{c} \frac{\partial \mathbf{e} \times \mathbf{b}}{\partial t} = \frac{1}{c} \frac{\partial \mathbf{e}}{\partial t} \times \mathbf{b} + \frac{1}{c} \mathbf{e} \times \frac{\partial \mathbf{b}}{\partial t}$$

$$\begin{aligned}
&= \left(\mathbf{curl} \mathbf{b} - \frac{1}{c} \mathbf{j} \right) \times \mathbf{b} + \mathbf{e} \times (-\mathbf{curl} \mathbf{e}) \\
&= -\frac{1}{c} \mathbf{j} \times \mathbf{b} - \mathbf{b} \times \mathbf{curl} \mathbf{b} - \mathbf{e} \times \mathbf{curl} \mathbf{e}.
\end{aligned} \tag{1.167}$$

For the first component, the x -component of the three-dimensional vector $(\mathbf{e}\mathbf{e}^\top)\nabla - \frac{1}{c}(e^2 \mathbf{I}_3)\nabla$ on the left-hand side, one obtains

$$\begin{aligned}
&\left[e_x^2 - \frac{1}{2} e^2 |e_x e_y| e_x e_z \right] \nabla \\
&= 2e_x \frac{\partial e_x}{\partial x} - \left(e_x \frac{\partial e_x}{\partial x} + e_y \frac{\partial e_y}{\partial y} + e_z \frac{\partial e_z}{\partial z} \right) + \frac{\partial e_x}{\partial y} e_y + e_x \frac{\partial e_y}{\partial y} + \frac{\partial e_x}{\partial z} e_z + e_x \frac{\partial e_z}{\partial z} \\
&= e_x \left(\frac{\partial e_x}{\partial x} + \frac{\partial e_y}{\partial y} + \frac{\partial e_z}{\partial z} \right) + e_z \frac{\partial e_x}{\partial z} + e_y \frac{\partial e_x}{\partial y} - e_z \frac{\partial e_z}{\partial x} - e_y \frac{\partial e_y}{\partial x} \\
&= e_x \operatorname{div} \mathbf{e} - (\mathbf{e} \times \mathbf{curl} \mathbf{e})_x = e_x \cdot \rho - (\mathbf{e} \times \mathbf{curl} \mathbf{e})_x.
\end{aligned} \tag{1.168}$$

Accordingly, one receives for the y - and z -component together

$$(\mathbf{e}\mathbf{e}^\top)\nabla - \frac{1}{2}(e^2 \mathbf{I}_3)\nabla = \mathbf{e} \cdot \rho - \mathbf{e} \times \mathbf{curl} \mathbf{e}. \tag{1.169}$$

Also we obtain with $\operatorname{div} \mathbf{b} = 0$:

$$(\mathbf{b}\mathbf{b}^\top)\nabla - \frac{1}{2}(b^2 \mathbf{I}_3)\nabla = -\mathbf{b} \times \mathbf{curl} \mathbf{b}. \tag{1.170}$$

Overall, (1.165), (1.168) and (1.170) result for the second to fourth rows of (1.164) in

$$\mathbf{e} \cdot \rho + \frac{1}{c} \mathbf{j} \times \mathbf{b} = \mathbf{f}. \tag{1.171}$$

But this is the *Momentum Theorem* of Lorentz! Thus the above assertion (1.164) is completely proved.

1.9.2 The Mechanical Energy–Momentum Matrix

A particle at rest at the point \mathbf{x}_0 has the speed

$$\vec{u} = \frac{d\vec{x}}{d\tau} = \frac{d}{d\tau} \begin{pmatrix} c\tau \\ \mathbf{x}_0 \end{pmatrix} = \begin{pmatrix} c \\ \mathbf{0} \end{pmatrix}$$

and the momentum $\vec{p} = m_0 \begin{pmatrix} c \\ \mathbf{0} \end{pmatrix}$. The zeroth component of \vec{p} is the rest energy of the particle divided by c . For a moving particle, with $E = \gamma_u m_0 c^2$, one gets

$$\vec{p} = \gamma_v m_0 \begin{pmatrix} c \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} E/c \\ \gamma_v \mathbf{p} \end{pmatrix}.$$

This equation expresses the fact that in the theory of relativity energy and momentum are the temporal and spatial components of the four-vector \vec{p} . They preserve this distinction even after a Lorentz transformation just like the four-vector $\vec{x} = \begin{pmatrix} ct \\ x \end{pmatrix}$: its zeroth component is always the time component and the rest represents the space components.

We now go over to a distributed matter, such as that in a perfect fluid, i.e. a liquid without internal friction, but of quite variable density. It is described by the two scalar fields, density ρ and pressure p , and the velocity vector field \vec{u} . The aim of this derivation of the energy-momentum matrix is that this matrix somehow represents the energy content of the liquid and the transition to the curved world of the theory of general relativity, which can serve as a source of the field of gravity. The *continuity equation* describes the conservation of mass. The conservation of mass requires that the change of mass $dV \delta\rho/\delta t$ in the unit of time, associated with the local compression $\delta\rho/\delta t$, must be equal to the difference between the entering and exiting masses per unit of time. In the x -direction, for this difference the following is valid:

$$\rho u_x(x) \cdot dy \cdot dz - \left(\rho u_x + \frac{\partial \rho u_x}{\partial x} dx \right) dy dz = -\frac{\partial \rho u_x}{\partial x} dV.$$

One gets similar expressions for the y - and z -direction, making a total for the conservation of mass

$$\frac{\partial \rho}{\partial t} dV = - \left(\frac{\partial \rho u_x}{\partial x} + \frac{\partial \rho u_y}{\partial y} + \frac{\partial \rho u_z}{\partial z} \right) dV.$$

So this is the final continuity equation in differential form:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_x}{\partial x} + \frac{\partial \rho u_y}{\partial y} + \frac{\partial \rho u_z}{\partial z} = \underline{\underline{\frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u})}} = 0. \quad (1.172)$$

The *dynamic* behaviour is described by the Euler's equation. By using Newton's law on the mass contained in a volume element of a perfect fluid, we obtain the Euler equation of motion, initially only in the x -direction:

$$dm \frac{du_x}{dt} = \rho dx dy dz \frac{du_x}{dt} = dx dy dz f_{D,x} - \left(\frac{\partial p}{\partial x} dx \right) dy dz,$$

which implies

$$\rho \frac{du_x}{dt} = f_{D,x} - \frac{\partial p}{\partial x}, \quad (1.173)$$

where $f_{D,x}$ is the x -component of the force per unit volume (power density), f_D , e.g. the gravitational force. Using the total differential of Δu_x ,

$$\Delta u_x = \frac{\partial u_x}{\partial t} \Delta t + \frac{\partial u_x}{\partial x} \Delta x + \frac{\partial u_x}{\partial y} \Delta y + \frac{\partial u_x}{\partial z} \Delta z,$$

dividing by Δt and passing to the limit $\Delta t \rightarrow 0$, we obtain

$$\frac{du_x}{dt} = \frac{\partial u_x}{\partial t} + \frac{\partial u_x}{\partial x} u_x + \frac{\partial u_x}{\partial y} u_y + \frac{\partial u_x}{\partial z} u_z. \quad (1.174)$$

With the corresponding equations for the y - and z -direction, we obtain in total

$$\rho \left(\frac{\partial u_x}{\partial t} + \frac{\partial u_x}{\partial x} u_x + \frac{\partial u_x}{\partial y} u_y + \frac{\partial u_x}{\partial z} u_z \right) = f_{D,x} - \frac{\partial p}{\partial x},$$

$$\rho \left(\frac{\partial u_y}{\partial t} + \frac{\partial u_y}{\partial x} u_x + \frac{\partial u_y}{\partial y} u_y + \frac{\partial u_y}{\partial z} u_z \right) = f_{D,y} - \frac{\partial p}{\partial y},$$

$$\rho \left(\frac{\partial u_z}{\partial t} + \frac{\partial u_z}{\partial x} u_x + \frac{\partial u_z}{\partial y} u_y + \frac{\partial u_z}{\partial z} u_z \right) = f_{D,z} - \frac{\partial p}{\partial z},$$

summarized in

$$\underline{\underline{\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{u}}{\partial \mathbf{x}^\top} \mathbf{u} \right) + \mathbf{grad} p = \mathbf{f}_D.}} \quad (1.175)$$

This is the Euler's equation in modern form. The first term in the parentheses is called the local and the second the convective change.

The relativistic generalization of the hydrodynamic equations (1.172) and (1.175) will now be established. ρ_0 is the *rest density*, defined as the rest mass per rest volume. With $\vec{x}^\top = [ct|\mathbf{x}^\top]$ and $\vec{u}^\top = \gamma_u[c|\mathbf{u}^\top]$, the Euler's equation (1.175) can also be written as

$$\rho \frac{\partial \mathbf{u}}{\partial \vec{x}^\top} \vec{u} = \mathbf{f}_D - \mathbf{grad} p. \quad (1.176)$$

With regard to the subsequent application of the operator $\vec{\nabla}$, we now start for the energy–momentum matrix with

$$\mathbf{T}_{\text{mech},1} \stackrel{\text{def}}{=} \rho_0 \vec{u} \vec{u}^\top. \quad (1.177)$$

This matrix is symmetric and built up from the two values ρ_0 and \vec{u} , which describe completely the dynamics of a perfect fluid with the pressure p and the acting external forces \mathbf{f}_D . For the use of the operator $\vec{\nabla}$ and the further investigation of the result, it is advantageous to first divide the matrix $\mathbf{T}_{\text{mech},1}$ similarly as the matrix $\mathbf{T}_{b,e}$:

$$\mathbf{T}_{\text{mech},1} = \rho_0 \gamma_u^2 \begin{pmatrix} c^2 & c \mathbf{u}^\top \\ c \mathbf{u} & \mathbf{u} \mathbf{u}^\top \end{pmatrix}. \quad (1.178)$$

In a liquid moving with the velocity \mathbf{u} , the volume decreases with γ_u while the mass increases with γ_u , thus one obtains the total density $\rho = \gamma_u^2 \rho_0$. With this we now

define

$$\mathbf{T}_{\text{mech},1} \stackrel{\text{def}}{=} \begin{pmatrix} \rho c^2 & \rho c \mathbf{u}^\top \\ \rho c \mathbf{u} & \rho \mathbf{u} \mathbf{u}^\top \end{pmatrix}. \quad (1.179)$$

Multiplying the first line in (1.179) from the right with the operator $\vec{\nabla}$ yields

$$-c \frac{\partial \rho}{\partial t} + c(\rho \mathbf{u}^\top) \nabla.$$

By setting this expression equal to zero, we obtain with $(\rho \mathbf{u}^\top) \nabla = \text{div}(\rho \mathbf{u})$ the classical continuity equation (1.172)!

Multiplying now the second row of the matrix in (1.179) from the right with the operator $\vec{\nabla}$, one obtains

$$\begin{aligned} -\frac{\partial \rho \mathbf{u}}{\partial t} + (\rho \mathbf{u} \mathbf{u}^\top) \nabla &= -\rho \frac{\partial \mathbf{u}}{\partial t} - \frac{\partial \rho}{\partial t} \mathbf{u} + \rho \frac{\partial \mathbf{u}}{\partial \mathbf{x}^\top} \mathbf{u} + \mathbf{u} \text{div}(\rho \mathbf{u}) \\ &= \left(-\frac{\partial \rho}{\partial t} \mathbf{u} + \text{div}(\rho \mathbf{u}) \right) \mathbf{u} + \rho \left(-\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial \mathbf{u}}{\partial \mathbf{x}^\top} \mathbf{u} \right). \end{aligned} \quad (1.180)$$

The term inside the first parentheses is equal to zero, due to the continuity equation in the non-relativistic case, and the term in the second parentheses contains the force- and pressure-free Euler's equation!

The pressure p must now be incorporated. Presuppose an isotropic liquid, then the pressure p is direction-independent. In the Euler's equation, the pressure appears in the form of **grad** p which can be written as

$$\mathbf{grad} p = \begin{pmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{pmatrix} \nabla. \quad (1.181)$$

If this is taken into account in the matrix \mathbf{T}_{mech} , it must be remembered that (1.181) holds for a reference system \mathcal{X}' moving with the liquid; therefore, this approach makes sense:

$$\mathbf{T}'_{\text{mech},2} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \quad (1.182)$$

This matrix can now with the help of the Lorentz matrix

$$\mathbf{L}(-\mathbf{u}) = \begin{pmatrix} \gamma_u & \frac{\gamma_u}{c} \mathbf{u}^\top \\ \frac{\gamma_u}{c} \mathbf{u} & \mathbf{I}_3 + (\gamma_u - 1) \frac{\mathbf{u} \mathbf{u}^\top}{u^2} \end{pmatrix}$$

be transformed back into the resting inertial system \mathcal{X} :

$$\mathbf{T}_{\text{mech},2} = \mathbf{L}(-\mathbf{u}) \mathbf{T}'_{\text{mech},2} \mathbf{L}^\top(-\mathbf{u}) = p \begin{pmatrix} \frac{\gamma_u^2 u^2}{c^2} & \frac{\gamma_u^2}{c} \mathbf{u}^\top \\ \frac{\gamma_u^2}{c} \mathbf{u} & \mathbf{I}_3 + \frac{\gamma_u^2}{c^2} \mathbf{u} \mathbf{u}^\top \end{pmatrix}.$$

Considering that $p\gamma^2 u^2/c^2 = p(\gamma^2 - 1)$, this can also be written as

$$\mathbf{T}_{\text{mech},2} = \frac{p}{c^2} \bar{\mathbf{u}} \bar{\mathbf{u}}^\top + p \begin{pmatrix} -1 & \mathbf{o}^\top \\ \mathbf{o} & \mathbf{I}_3 \end{pmatrix}.$$

For the sum of the two matrices $\mathbf{T}_{\text{mech},1}$ and $\mathbf{T}_{\text{mech},2}$, we finally obtain with the Minkowski matrix \mathbf{M}

$$\mathbf{T}_{\text{mech}} \stackrel{\text{def}}{=} \left(\rho_0 + \frac{p}{c^2} \right) \bar{\mathbf{u}} \bar{\mathbf{u}}^\top - p \mathbf{M}. \quad (1.183)$$

We now summarize everything together to the relativistic generalization of the hydrodynamic equations

$$\mathbf{T}_{\text{mech}} \vec{\nabla} = \vec{\mathbf{f}}_D.$$

(1.184)

By the way, one obtains, when the energy–momentum matrix \mathbf{T}_{mech} from the right is multiplied with $\mathbf{M}\bar{\mathbf{u}}$, the four-vector momentum density $\rho_0\bar{\mathbf{u}}$ multiplied by c^2 :

$$\mathbf{T}_{\text{mech}} \mathbf{M} \bar{\mathbf{u}} = \left(\rho_0 + \frac{p}{c^2} \right) \bar{\mathbf{u}} \underbrace{\gamma^2(c^2 - v^2)}_{c^2} - p \bar{\mathbf{u}} = c^2 \rho_0 \bar{\mathbf{u}}. \quad (1.185)$$

1.9.3 The Total Energy–Momentum Matrix

The derived energy–momentum matrices include the conservation laws of energy and momentum of a *closed* system. For example, if a force density \mathbf{f}_D acts from the outside on the fluid, so that an electromagnetic field acts on the electrically charged liquid, then

$$\mathbf{f}_D = -\mathbf{T}_{b,e} \vec{\nabla} = \mathbf{T}_{\text{mech}} \vec{\nabla}, \quad (1.186)$$

or combined,

$$\mathbf{T}_{\text{total}} \stackrel{\text{def}}{=} (\mathbf{T}_{\text{mech}} + \mathbf{T}_{b,e}), \quad (1.187)$$

so

$$\mathbf{T}_{\text{total}} \vec{\nabla} = \mathbf{f}_{\text{total}}.$$

(1.188)

The conservation laws now apply to the whole system, which is fluid plus the electromagnetic field. Since the individual matrices are symmetric, the total energy–momentum matrix $\mathbf{T}_{\text{total}}$ is symmetric. If there are other components in the considered system, they can be incorporated in the symmetric total energy–momentum matrix $\mathbf{T}_{\text{total}}$ in a similar way as described above, and one obtains (1.188). This form of mathematical representation of the dynamic behaviour of physical systems will play a major role in the main equations of Einstein’s theory of General Relativity!

1.10 The Most Important Definitions and Formulas in Special Relativity

For inertial systems, i.e. reference systems that are uniformly moving against each other, the fundamental physical laws are linked with the Lorentz transformation and are invariant. In the Theory of Special Relativity, we derived (see (1.20))

$$\gamma \stackrel{\text{def}}{=} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

and the Lorentz transformation matrix (see (1.38))

$$\mathbf{L}(\mathbf{v}) \stackrel{\text{def}}{=} \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c} \mathbf{v}^\top \\ \hline -\frac{\gamma}{c} \mathbf{v} & \mathbf{I} + (\gamma - 1) \frac{\mathbf{v} \mathbf{v}^\top}{v^2} \end{array} \right)$$

with the following (see (1.39)) equations:

$$ct' = \gamma ct - \frac{\gamma}{c} \mathbf{v}^\top \mathbf{x}, \quad \mathbf{x}' = \mathbf{x} + (\gamma - 1) \frac{\mathbf{v}^\top \mathbf{x}}{v^2} \mathbf{v} - \gamma \mathbf{v} t.$$

In (1.65), we derived the relativistic velocity addition:

$$\mathbf{w} = \frac{\mathbf{v} + \mathbf{u} + (\frac{1}{\gamma v} - 1)(\mathbf{u} - \frac{\mathbf{v}^\top \mathbf{u}}{v^2} \mathbf{v})}{1 + \frac{\mathbf{v}^\top \mathbf{u}}{c^2}}.$$

If the two velocities \mathbf{v} and \mathbf{u} are parallel, then (see (1.68))

$$\mathbf{w} = \frac{\mathbf{v} + \mathbf{u}}{1 + \frac{\mathbf{v}^\top \mathbf{u}}{c^2}}.$$

The modified velocity vector in (1.79)

$$\mathbf{u} \stackrel{\text{def}}{=} \frac{d\mathbf{x}}{dt} \in \mathbb{R}^3, \quad \gamma_u \stackrel{\text{def}}{=} \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}}, \quad \vec{\mathbf{u}} \stackrel{\text{def}}{=} \gamma_u \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}^4$$

is transformed by a Lorentz transformation \mathbf{L} into the velocity vector $\vec{\mathbf{u}}'$ (see (1.80))

$$\vec{\mathbf{u}}' = \mathbf{L} \vec{\mathbf{u}}.$$

With (1.95),

$$\vec{\mathbf{a}} \stackrel{\text{def}}{=} \gamma_u \cdot \frac{d}{dt} \vec{\mathbf{u}} \in \mathbb{R}^4$$

is the Lorentz transformation of the acceleration vector \vec{a} (see (1.96)):

$$\vec{a}' = \mathbf{L} \vec{a}.$$

Einstein's famous formula (1.111) reads

$$E_0 = m_0 c^2.$$

The invariance of the fundamental equation of mechanics (m_0 is the rest mass) is documented in

$$m_o \vec{a} = \vec{f}$$

$$\mathbf{L} \Rightarrow \Leftarrow \mathbf{L}^{-1}$$

$$m_o \vec{a}' = \vec{f}'$$

and in electrodynamics

$$\begin{aligned} \mathbf{F}_{B,e} \vec{\nabla} &= \frac{1}{c} \vec{j} \\ \mathbf{F}_{E,b} \vec{\nabla} &= \mathbf{o} \end{aligned}$$

$$\mathbf{L} \Rightarrow \Leftarrow \mathbf{L}^{-1}$$

$$\begin{aligned} \mathbf{F}'_{B',e'} \vec{\nabla}' &= \frac{1}{c} \vec{j}' \\ \mathbf{F}'_{E',b'} \vec{\nabla}' &= \mathbf{o} \end{aligned}$$

and

$$\vec{f} = \frac{q}{c} \mathbf{M} \mathbf{F}_{B,e} \vec{u}$$

$$\mathbf{L} \Rightarrow \Leftarrow \mathbf{L}^{-1}$$

$$\vec{f}' = \frac{q}{c} \mathbf{M} \mathbf{F}'_{B',e'} \vec{u}'$$

with

$$\mathbf{F}_{B,e} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & \mathbf{e}^\top \\ -\mathbf{e} & \mathbf{B}_\times \end{pmatrix}, \quad \mathbf{F}_{E,b} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -\mathbf{b}^\top \\ \mathbf{b} & \mathbf{E}_\times \end{pmatrix} \text{ and } \vec{\nabla} \stackrel{\text{def}}{=} \left(-\frac{1}{c} \frac{\partial}{\partial t} \right).$$

With the electromagnetic energy-momentum matrix

$$\mathbf{T}_{b,e} \stackrel{\text{def}}{=} \begin{pmatrix} w & s^\top \\ s & (\mathbf{e}\mathbf{e}^\top + \mathbf{b}\mathbf{b}^\top - w\mathbf{I}_3) \end{pmatrix}$$

where

$$w \stackrel{\text{def}}{=} 1/2(e^2 + b^2),$$

one gets

$$\mathbf{T}_{b,e} \cdot \vec{\nabla} = \frac{1}{\gamma_v} \vec{f},$$

and with the mechanical energy-momentum matrix

$$\mathbf{T}_{\text{mech}} \stackrel{\text{def}}{=} \left(\rho_0 + \frac{p}{c^2} \right) \vec{u} \vec{u}^\top - p \mathbf{M},$$

we obtain the relativistic generalization of the hydrodynamic equations

$$\boxed{\boldsymbol{T}_{\text{mech}} \vec{\nabla} = \vec{f}_D.}$$

Remark The operator $\vec{\nabla}$ was somewhat unusual in the above formulas—it was written to the right to the approaching object, e.g. in $\boldsymbol{T}_{\text{mech}} \vec{\nabla}$. This was done so that *column* vectors appear on both sides of the equation. The usual order would be if one transposes the eligible equations. Then there would be *row* vectors on the left- and on the right-hand side, and the operators, though provided with a transpose sign, would have their usual place as, e.g. in

$$\vec{\nabla}^\top \boldsymbol{T}_{\text{mech}} = \vec{f}_D^\top.$$

(Since the matrix $\boldsymbol{T}_{\text{mech}}$ is symmetric, it need not be transposed.)

Chapter 2

Theory of General Relativity

The chapter begins with the introduction of the metric matrix \mathbf{G} and the effect of a homogeneous field of gravitation on a mass particle. Then the motion on geodesic lines in a gravitational field is considered. The general transformation of coordinates leads to the Christoffel matrix and the Riemannian curvature matrix. With the help of the Ricci matrix, the Theory of General Relativity of Einstein can then be formulated.

2.1 General Relativity and Riemannian Geometry

In the theory of general relativity, the invariance of the equations with respect to any coordinate transformation is required. Especially the theorems of physics are to remain unchanged if one transforms them from one coordinate system in another coordinate system by the general transformation equations

$$x_i = x_i(x'_0, x'_1, x'_2, x'_3) \quad \text{for } i = 0, 1, 2 \text{ and } 3. \quad (2.1)$$

Consider an arbitrary coordinate system \mathcal{K} on the infinitely small neighbourhood of the point P , in which also a field of gravity may be present. In an infinitely small space and an infinitely small time interval, or in other words, in an infinitely small spacetime interval, the coordinate system \mathcal{K} can be replaced by a coordinate system \mathcal{K}' which is accelerated to it and in which no field of gravitation is present. \mathcal{K}' is the *local* spacetime coordinate system in the neighbourhood of a point, and \mathcal{K} the *general* coordinate system. It is now assumed that

For all local coordinate systems \mathcal{K}' the Special Relativity is valid in any infinitely small four-dimensional neighbourhood.

The point P' may be infinitely close to the point P , and have the real coordinates dx_0, dx_1, dx_2 and dx_3 in the Cartesian (rectangular) coordinate system. For a line element, one has

$$ds^2 = dx_0^2 - (dx_1^2 + dx_2^2 + dx_3^2) \quad (2.2)$$

where x_0 is the time coordinate ct . If ds^2 is positive, then P' is reached from P by a movement with a velocity smaller than the velocity of light. Equation (2.2) can be written using the matrix

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

as a quadratic form

$$ds^2 = d\vec{x}^\top \mathbf{M} d\vec{x}. \quad (2.3)$$

Going on to the coordinate system \mathcal{K}' , one gets with

$$dx_i = \frac{\partial x_i}{\partial x'_0} dx'_0 + \frac{\partial x_i}{\partial x'_1} dx'_1 + \frac{\partial x_i}{\partial x'_2} dx'_2 + \frac{\partial x_i}{\partial x'_3} dx'_3$$

and the Jacobi matrix

$$\mathbf{J} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial x_0}{\partial x'_0} & \cdots & \frac{\partial x_0}{\partial x'_3} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_3}{\partial x'_0} & \cdots & \frac{\partial x_3}{\partial x'_3} \end{pmatrix} = \frac{\partial \vec{x}}{\partial \vec{x}'^\top} \quad (2.4)$$

the connection

$$d\vec{x} = \mathbf{J} d\vec{x}'. \quad (2.5)$$

This is used in (2.3), yielding

$$ds^2 = d\vec{x}'^\top \mathbf{J}^\top \mathbf{M} \mathbf{J} d\vec{x}' = d\vec{x}'^\top \mathbf{G} d\vec{x}'. \quad (2.6)$$

So the *metric matrix* \mathbf{G} is defined as

$\mathbf{G} \stackrel{\text{def}}{=} \mathbf{J}^\top \mathbf{M} \mathbf{J} \in \mathbb{R}^{4 \times 4}.$

(2.7)

The matrix elements g_{ik} are functions of the parameters x'_i . They may change their values from point to point. In Special Relativity, $\mathbf{G} = \mathbf{M}$ in any finite region. An unforced particle moves in such a region straightforwardly and uniformly.

But if the particle is in a gravitational field, the motion is curvilinear and nonuniform. Depending on the nature of the gravitational field, the g_{ik} are functions of other parameters. At most ten different elements g_{ik} of the symmetric 4×4 -matrix \mathbf{G} describe the field of gravitation at every point in the coordinate system. In every *local* coordinate system the g_{ik} are constant and can be transformed by a similarity transformation into the form \mathbf{M} . Because if $\mathbf{G} = \mathbf{J}^\top \mathbf{M} \mathbf{J}$, then one immediately obtains with the transformation $d\vec{x}' = \mathbf{J}^{-1} d\vec{\xi}$

$$d\vec{x}'^\top \mathbf{G} d\vec{x}' = d\vec{\xi}^\top \mathbf{J}^{-1} \mathbf{J}^\top \mathbf{M} \mathbf{J} \mathbf{J}^{-1} d\vec{\xi} = d\vec{\xi}^\top \mathbf{M} d\vec{\xi}.$$

But this is true only for a point because the matrices \mathbf{J} are different from point to point. There is *no* transformation matrix \mathbf{J} which is valid globally.

2.2 Motion in a Gravitational Field

What influence does a gravitational field have on a mass particle? Following the relativity principle, in the local inertial system, which means in a coordinate system moving with the particle, the laws of special relativity are valid. For the movement of a particle with no external forces acting on it,

$$\frac{d^2\vec{\xi}}{d\tau^2} = \mathbf{0}. \quad (2.8)$$

The proper time τ follows from

$$ds^2 = c^2 d\tau^2 = d\vec{\xi}^\top \mathbf{M} d\vec{\xi}. \quad (2.9)$$

By integrating (2.8), one gets with the initial position $\vec{\xi}(0)$ and the initial velocity $\dot{\vec{\xi}}(0)$

$$\vec{\xi}(\tau) = \vec{\xi}(0) + \dot{\vec{\xi}}(0)\tau.$$

A photon also moves in a straight line in the local inertial system. But τ is then not the proper time of the photon. A photon has no proper time because for light $ds = 0 = c d\tau$. Therefore, we introduce the parameter λ such that

$$\frac{d^2\vec{\xi}}{d\lambda^2} = \mathbf{0}$$

is the equation of motion of the photon in the local inertial system. Now we move from the local inertial system with the spacetime vector $\vec{\xi}$ to the global inertial system with the spacetime vector \vec{x} . With

$$\mathbf{J} \stackrel{\text{def}}{=} \frac{\partial \vec{\xi}}{\partial \vec{x}^\top}$$

one gets for (2.9)

$$ds^2 = c^2 d\tau^2 = d\vec{x}^\top \mathbf{J}^\top \mathbf{M} \mathbf{J} d\vec{x} = d\vec{x}^\top \mathbf{G} d\vec{x}. \quad (2.10)$$

For light this becomes

$$d\vec{x}^\top \mathbf{G} d\vec{x} = 0. \quad (2.11)$$

2.2.1 First Solution

From (2.8) one obtains for the motion of a particle with

$$\dot{\vec{x}} \stackrel{\text{def}}{=} \frac{d\vec{x}}{d\tau},$$

$$\frac{d}{d\tau} \left(\frac{\partial \vec{\xi}}{\partial \vec{x}^\top} \frac{d\vec{x}}{d\tau} \right) = \frac{d}{d\tau} (\mathbf{J} \dot{\vec{x}}) = \frac{d}{d\tau} (\mathbf{J}) \dot{\vec{x}} + \mathbf{J} \ddot{\vec{x}} = \mathbf{0}, \quad (2.12)$$

and with (A.90)

$$\frac{d}{d\tau} (\mathbf{J}(\vec{x}(\tau))) = (\dot{\vec{x}}^\top \otimes \mathbf{I}_4) \frac{\partial \mathbf{J}}{\partial \vec{x}}$$

from (2.12)

$$\ddot{\vec{x}} = -\mathbf{J}^{-1} (\dot{\vec{x}}^\top \otimes \mathbf{I}_4) \frac{\partial \mathbf{J}}{\partial \vec{x}} \dot{\vec{x}}, \quad (2.13)$$

or

$$\ddot{\vec{x}} = -(\dot{\vec{x}}^\top \otimes \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}} \dot{\vec{x}} = -(\dot{\vec{x}}^\top \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}} \dot{\vec{x}},$$

that is,

$$\ddot{\vec{x}} = -(\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}} \dot{\vec{x}}. \quad (2.14)$$

With $\mathbf{J}_k \stackrel{\text{def}}{=} \frac{\partial \mathbf{J}}{\partial x_k} \in \mathbb{R}^{4 \times 4}$ and

$$\hat{\mathbf{F}} = \begin{pmatrix} \hat{\mathbf{F}}_0 \\ \vdots \\ \hat{\mathbf{F}}_3 \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}} = \mathbf{U}_{4 \times 4} \begin{pmatrix} \mathbf{J}^{-1} \mathbf{J}_0 \\ \vdots \\ \mathbf{J}^{-1} \mathbf{J}_3 \end{pmatrix} \in \mathbb{R}^{16 \times 4} \quad (2.15)$$

one can write in place of (2.14) the compact equation

$$\ddot{\vec{x}} = -(\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \hat{\mathbf{F}} \dot{\vec{x}}. \quad (2.16)$$

With (2.15) and (2.16) one gets for the individual vector components \ddot{x}_k according to the form of the permutation matrix $\mathbf{U}_{4 \times 4}$ in the appendix ($\mathbf{j}_k^{-\top} \in \mathbb{R}^4$ is row k of \mathbf{J}^{-1})

$$\ddot{x}_k = -\dot{\vec{x}}^\top (\mathbf{I}_4 \otimes \mathbf{j}_k^{-\top}) \frac{\partial \mathbf{J}}{\partial \vec{x}} \dot{\vec{x}}. \quad (2.17)$$

From (2.15) and (2.17) we can read directly

$$\hat{\mathbf{F}}_k = (\mathbf{I}_4 \otimes \mathbf{j}_k^{-\top}) \frac{\partial \mathbf{J}}{\partial \vec{x}}. \quad (2.18)$$

The so-called Christoffel matrices $\hat{\Gamma}$ can be calculated directly from the Jacobi matrix \mathbf{J} , that is, from the transformation matrix for the transition from the local inertial system to the accelerated non-inertial system (the coordinate system with a gravitational field). For the motion of a photon, we obtain in the same manner

$$\boxed{\frac{d^2 \vec{x}}{d\lambda^2} = -\left(\mathbf{I}_4 \otimes \frac{d\vec{x}^\top}{d\lambda}\right) \hat{\Gamma} \frac{d\vec{x}}{d\lambda}.} \quad (2.19)$$

2.2.2 Second Solution

An alternative solution is obtained from the second form in (A.90), namely

$$\frac{d}{d\tau}(\mathbf{J}) = \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} (\dot{\vec{x}} \otimes \mathbf{I}_4). \quad (2.20)$$

Thus we obtain from (2.12)

$$\ddot{\vec{x}} = -\mathbf{J}^{-1} \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} (\dot{\vec{x}} \otimes \mathbf{I}_4) \dot{\vec{x}} = -\mathbf{J}^{-1} \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} (\dot{\vec{x}} \otimes \dot{\vec{x}}), \quad (2.21)$$

that is, with

$$\boxed{\tilde{\Gamma} \stackrel{\text{def}}{=} \mathbf{J}^{-1} \frac{\partial \mathbf{J}}{\partial \vec{x}^\top}} \quad (2.22)$$

written completely as

$$\tilde{\Gamma} = \mathbf{J}^{-1} \left[\begin{array}{c|c|c|c} \frac{\partial \mathbf{J}}{\partial x_0} & \frac{\partial \mathbf{J}}{\partial x_1} & \frac{\partial \mathbf{J}}{\partial x_2} & \frac{\partial \mathbf{J}}{\partial x_3} \end{array} \right] \in \mathbb{R}^{4 \times 16}, \quad (2.23)$$

so with $\dot{\vec{x}} \otimes \dot{\vec{x}} \in \mathbb{R}^{16}$,

$$\boxed{\ddot{\vec{x}} = -\tilde{\Gamma}(\dot{\vec{x}} \otimes \dot{\vec{x}}).} \quad (2.24)$$

This is an alternative representation of the relation (2.16)!

If one defines

$$\tilde{\gamma}_k^\top \stackrel{\text{def}}{=} \mathbf{j}_k^{-\top} \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} \in \mathbb{R}^{16}, \quad (2.25)$$

then for each vector component the following scalar vector product is obtained:

$$\ddot{x}_k = -\tilde{\gamma}_k^\top \cdot (\dot{\vec{x}} \otimes \dot{\vec{x}}). \quad (2.26)$$

2.2.3 Relation Between $\tilde{\Gamma}$ and G

Since $G = J^\top M J$ and $\tilde{\Gamma} = J^{-1} \frac{\partial J}{\partial x^\top}$, the matrix $\tilde{\Gamma}$ must depend on $\frac{\partial G}{\partial x}$. This indeed is the case.

On the one hand,

$$G\tilde{\Gamma} = J^\top M \frac{\partial J}{\partial x^\top}, \quad (2.27)$$

and, on the other hand, $g_{\mu\nu} = j_\mu^\top M j_\nu$, so

$$\frac{\partial g_{\mu\nu}}{\partial x_\lambda} = \frac{\partial j_\mu^\top}{\partial x_\lambda} M j_\nu + j_\mu^\top M \frac{\partial j_\nu}{\partial x_\lambda}. \quad (2.28)$$

Furthermore,

$$\frac{\partial g_{\lambda\nu}}{\partial x_\mu} = \frac{\partial j_\lambda^\top}{\partial x_\mu} M j_\nu + j_\lambda^\top M \frac{\partial j_\nu}{\partial x_\mu} \quad (2.29)$$

and

$$\frac{\partial g_{\mu\lambda}}{\partial x_\nu} = \frac{\partial j_\mu^\top}{\partial x_\nu} M j_\lambda + j_\mu^\top M \frac{\partial j_\lambda}{\partial x_\nu}. \quad (2.30)$$

If we add (2.28) and (2.29) and then subtract (2.30), we obtain

$$\frac{\partial g_{\mu\nu}}{\partial x_\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x_\mu} - \frac{\partial g_{\mu\lambda}}{\partial x_\nu} = 2 \frac{\partial j_\mu^\top}{\partial x_\lambda} M j_\nu = 2 j_\nu^\top M \frac{\partial j_\mu}{\partial x_\lambda}. \quad (2.31)$$

Calling $G\tilde{\Gamma} \stackrel{\text{def}}{=} \check{\tilde{\Gamma}}$, then we obtain with (2.27) and (2.31) as the element $\check{\tilde{\Gamma}}_{\nu\mu}^\lambda$ in the ν th row and the μ th column of $\check{\tilde{\Gamma}}_\lambda$

$$\check{\tilde{\Gamma}}_{\nu\mu}^\lambda = \frac{1}{2} \left(\frac{\partial g_{\mu\nu}}{\partial x_\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x_\mu} - \frac{\partial g_{\mu\lambda}}{\partial x_\nu} \right). \quad (2.32)$$

Since $\tilde{\Gamma} = G^{-1} G \tilde{\Gamma} = G^{-1} \check{\tilde{\Gamma}}$,

$$\tilde{\Gamma}_\lambda = G^{-1} \check{\tilde{\Gamma}}_\lambda = G^{-1} J^\top M \frac{\partial J}{\partial x_\lambda},$$

and one finally obtains with the α th row $g_\alpha^{[-T]}$ of the matrix G^{-1} and the ν th element $g_{\alpha\nu}^{[-1]}$ of this row vector the following relation between the elements in the α th row and the μ th column of $\tilde{\Gamma}_\lambda$ and the elements of G

$$\tilde{\Gamma}_{\alpha\mu}^\lambda = g_\alpha^{[-T]} J^\top M \frac{\partial j_\mu}{\partial x_\lambda} = \sum_{\nu=0}^3 \frac{g_{\alpha\nu}^{[-1]}}{2} \left(\frac{\partial g_{\mu\nu}}{\partial x_\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x_\mu} - \frac{\partial g_{\mu\lambda}}{\partial x_\nu} \right). \quad (2.33)$$

This is the desired relationship between the matrix elements of $\tilde{\Gamma}$ and G !

2.3 Geodesic Lines and Equations of Motion

The motion of photons and particles in a gravitational field will again be considered, but now using the calculus of variations. The same results as in (2.16) are expected. This is to be seen by using the same Christoffel matrix Γ in the results. In special relativity, the motion of a photon is given by $c^2 t^2 = \mathbf{x}^\top \mathbf{x}$, that means $s^2 = c^2 t^2 - \mathbf{x}^\top \mathbf{x} = 0$, or $ds^2 = dx_0^2 - dx^\top dx = d\vec{x}^\top \mathbf{M} d\vec{x} = 0$ for any small distance dx . If paths were straight lines, that would mean the shortest possible connection between the points P_1 and P_2 . Also the theory of general relativity demands that light and particles move on straightest possible paths. These paths are the so-called *geodesic curves* for which the length has an extreme value:

$$\delta \int_{P_1}^{P_2} ds = 0. \quad (2.34)$$

This produces a system of four differential equations. For the variation of ds^2 we obtain

$$\begin{aligned} \delta(ds^2) &= \delta(d\vec{x}^\top \mathbf{G} d\vec{x}), \\ 2(\delta ds) ds &= (\delta d\vec{x}^\top) \mathbf{G} d\vec{x} + d\vec{x}^\top (\delta \mathbf{G}) d\vec{x} + d\vec{x}^\top \mathbf{G} (\delta d\vec{x}), \end{aligned} \quad (2.35)$$

but since \mathbf{G} is symmetric, $\mathbf{G} = \mathbf{G}^\top$, we get

$$2(\delta ds) ds = 2d\vec{x}^\top \mathbf{G} (\delta d\vec{x}) + d\vec{x}^\top (\delta \mathbf{G}) d\vec{x}. \quad (2.36)$$

Equation (2.36) divided by $2 ds$ results, with $d(\delta \vec{x}) = \delta d\vec{x}$ and $\frac{d\vec{x}}{ds} \stackrel{\text{def}}{=} \dot{\vec{x}}$, in

$$\delta ds = \dot{\vec{x}}^\top \mathbf{G} d(\delta \vec{x}) + \frac{1}{2} \dot{\vec{x}}^\top (\delta \mathbf{G}) \dot{\vec{x}}. \quad (2.37)$$

Multiplying the right-hand side by ds provides

$$\delta ds = \left[\dot{\vec{x}}^\top \mathbf{G} \frac{d(\delta \vec{x})}{ds} + \frac{1}{2} \dot{\vec{x}}^\top (\delta \mathbf{G}) \dot{\vec{x}} \right] ds. \quad (2.38)$$

For the variation of the matrix \mathbf{G} we set

$$\delta \mathbf{G} = \frac{\partial \mathbf{G}}{\partial x_0} \delta x_0 + \frac{\partial \mathbf{G}}{\partial x_1} \delta x_1 + \frac{\partial \mathbf{G}}{\partial x_2} \delta x_2 + \frac{\partial \mathbf{G}}{\partial x_3} \delta x_3. \quad (2.39)$$

This is extended to a quadratic form:

$$\dot{\vec{x}}^\top \delta \mathbf{G} \dot{\vec{x}} = \dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_0} \dot{\vec{x}} \delta x_0 + \dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_1} \dot{\vec{x}} \delta x_1 + \dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_2} \dot{\vec{x}} \delta x_2 + \dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_3} \dot{\vec{x}} \delta x_3 \quad (2.40)$$

and the right-hand side is collected to a vector product, leading to the result

$$\dot{\vec{x}}^\top \delta \mathbf{G} \dot{\vec{x}} = \left[\dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_0} \dot{\vec{x}}, \dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_1} \dot{\vec{x}}, \dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_2} \dot{\vec{x}}, \dot{\vec{x}}^\top \frac{\partial \mathbf{G}}{\partial x_3} \dot{\vec{x}} \right] \delta \vec{x}. \quad (2.41)$$

The row vector on the right-hand side with

$$\left(\frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}^\top} \right) \stackrel{\text{def}}{=} \left[\frac{\partial \mathbf{G}}{\partial x_0}, \frac{\partial \mathbf{G}}{\partial x_1}, \frac{\partial \mathbf{G}}{\partial x_2}, \frac{\partial \mathbf{G}}{\partial x_3} \right]$$

and the Kronecker product can be written as

$$\left[\dot{\mathbf{x}}^\top \frac{\partial \mathbf{G}}{\partial x_0} \dot{\mathbf{x}}, \dot{\mathbf{x}}^\top \frac{\partial \mathbf{G}}{\partial x_1} \dot{\mathbf{x}}, \dot{\mathbf{x}}^\top \frac{\partial \mathbf{G}}{\partial x_2} \dot{\mathbf{x}}, \dot{\mathbf{x}}^\top \frac{\partial \mathbf{G}}{\partial x_3} \dot{\mathbf{x}} \right] = \dot{\mathbf{x}}^\top \left(\frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}^\top} \right) (\mathbf{I}_4 \otimes \dot{\mathbf{x}}). \quad (2.42)$$

With this relationship we obtain now for the variated integral (2.34)

$$\delta \int_{P_1}^{P_2} ds = \int_{P_1}^{P_2} \left[\dot{\mathbf{x}}^\top \mathbf{G} \frac{d(\delta \dot{\mathbf{x}})}{ds} + \frac{1}{2} \dot{\mathbf{x}}^\top \left(\frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}^\top} \right) (\mathbf{I}_4 \otimes \dot{\mathbf{x}}) \delta \dot{\mathbf{x}} \right] ds. \quad (2.43)$$

Taking in consideration $\delta \dot{\mathbf{x}}(P_1) = \delta \dot{\mathbf{x}}(P_2) = \mathbf{o}$, and performing integration by parts to the left summand in the integral, gives

$$\begin{aligned} \delta \int_{P_1}^{P_2} ds &= \int_{P_1}^{P_2} \left[-\frac{d}{ds} (\dot{\mathbf{x}}^\top \mathbf{G}) \delta \dot{\mathbf{x}} + \frac{1}{2} \dot{\mathbf{x}}^\top \left(\frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}^\top} \right) (\mathbf{I}_4 \otimes \dot{\mathbf{x}}) \delta \dot{\mathbf{x}} \right] ds \\ &= \int_{P_1}^{P_2} \left[-\frac{d}{ds} (\dot{\mathbf{x}}^\top \mathbf{G}) + \frac{1}{2} \dot{\mathbf{x}}^\top \left(\frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}^\top} \right) (\mathbf{I}_4 \otimes \dot{\mathbf{x}}) \right] \delta \dot{\mathbf{x}} ds = 0. \end{aligned} \quad (2.44)$$

To make the variation of the integral for every arbitrary vector function $\delta \dot{\mathbf{x}}(\cdot)$ disappear, in spite of the fundamental theorem of the calculus of variations, the vector function in the brackets must be identically zero:

$$-\frac{d}{ds} (\dot{\mathbf{x}}^\top \mathbf{G}) + \frac{1}{2} \dot{\mathbf{x}}^\top \left(\frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}^\top} \right) (\mathbf{I}_4 \otimes \dot{\mathbf{x}}) = \mathbf{o}^\top, \quad (2.45)$$

or transposed, where $(\mathbf{A} \otimes \mathbf{B})^\top = (\mathbf{A}^\top \otimes \mathbf{B}^\top)$ is used,

$$\frac{1}{2} (\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}} - \frac{d}{ds} (\mathbf{G} \dot{\mathbf{x}}) = \mathbf{o}. \quad (2.46)$$

For the second term on the left-hand side one gets

$$\frac{d}{ds} (\mathbf{G} \dot{\mathbf{x}}) = \mathbf{G} \ddot{\mathbf{x}} + \frac{d}{ds} (\mathbf{G} \dot{\mathbf{x}}). \quad (2.47)$$

And this is

$$\frac{d}{ds} (\mathbf{G}) = (\dot{\mathbf{x}}^\top \otimes \mathbf{I}_4) \frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}}.$$

Inserted into (2.47) this finally yields

$$\mathbf{G} \ddot{\mathbf{x}} = \frac{1}{2} (\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}} - (\dot{\mathbf{x}}^\top \otimes \mathbf{I}_4) \frac{\partial \mathbf{G}}{\partial \dot{\mathbf{x}}} \dot{\mathbf{x}} \quad (2.48)$$

or, putting $\ddot{\vec{x}}$ to the left-hand side,

$$\ddot{\vec{x}} = \mathbf{G}^{-1} \left[\frac{1}{2} (\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) - (\dot{\vec{x}}^\top \otimes \mathbf{I}_4) \right] \frac{\partial \mathbf{G}}{\partial \vec{x}} \dot{\vec{x}}. \quad (2.49)$$

Taking into account the lemma in the Appendix “Vectors and Matrices”:

$$\mathbf{B} \otimes \mathbf{A} = \mathbf{U}_{s \times p} (\mathbf{A} \otimes \mathbf{B}) \mathbf{U}_{q \times t}, \quad \mathbf{A} \in \mathbb{R}^{p \times q}, \mathbf{B} \in \mathbb{R}^{s \times t}, \quad (2.50)$$

(2.49) can be converted, with the help of $\dot{\vec{x}}^\top \otimes \mathbf{I}_4 = (\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \mathbf{U}_{4 \times 4}$, into

$$\ddot{\vec{x}} = \mathbf{G}^{-1} (\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \left[\frac{1}{2} \mathbf{I}_{16} - \mathbf{U}_{4 \times 4} \right] \frac{\partial \mathbf{G}}{\partial \vec{x}} \dot{\vec{x}}. \quad (2.51)$$

With

$$\mathbf{G}^{-1} (\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) = (\mathbf{G}^{-1} \otimes 1) (\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) = (\mathbf{G}^{-1} \otimes \dot{\vec{x}}^\top) = (\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) (\mathbf{G}^{-1} \otimes \mathbf{I}_4)$$

we finally obtain a form in which $\dot{\vec{x}}$ is pulled out to the left and to the right:

$$\ddot{\vec{x}} = (\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) (\mathbf{G}^{-1} \otimes \mathbf{I}_4) \left[\frac{1}{2} \mathbf{I}_{16} - \mathbf{U}_{4 \times 4} \right] \frac{\partial \mathbf{G}}{\partial \vec{x}} \dot{\vec{x}}. \quad (2.52)$$

Summarizing

$$\hat{\Gamma} \stackrel{\text{def}}{=} (\mathbf{G}^{-1} \otimes \mathbf{I}_4) \left[\mathbf{U}_{4 \times 4} - \frac{1}{2} \mathbf{I}_{16} \right] \frac{\partial \mathbf{G}}{\partial \vec{x}} \quad (2.53)$$

$$= \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{G}^{-1}) \left[\frac{1}{2} \mathbf{I}_{16} - \mathbf{U}_{4 \times 4} \right] \frac{\partial \mathbf{G}}{\partial \vec{x}}, \quad (2.54)$$

one gets the compact equation

$$\ddot{\vec{x}} = -(\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \hat{\Gamma} \dot{\vec{x}} \quad (2.55)$$

which agrees with the equation of motion (2.16), that is, in the language of the calculus of variations, this equation yields an extremal.

In (2.53) with the k th row \mathbf{g}_k^\top of the matrix \mathbf{G} is

$$\mathbf{U}_{4 \times 4} \frac{\partial \mathbf{G}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial \mathbf{g}_0^\top}{\partial \vec{x}} \\ \vdots \\ \frac{\partial \mathbf{g}_3^\top}{\partial \vec{x}} \end{pmatrix}.$$

For the four components \ddot{x}_k , $k = 0, 1, 2$ and 3 , one obtains with the k th row \mathbf{g}_k^{-T} of the matrix \mathbf{G}^{-1} ,

$$\ddot{x}_k = \dot{\vec{x}}^\top (\mathbf{g}_k^{-T} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{g}_0^\top}{\partial \vec{x}} \\ \vdots \\ \frac{\partial \mathbf{g}_3^\top}{\partial \vec{x}} \end{pmatrix} - \frac{1}{2} \frac{\partial \mathbf{G}}{\partial \vec{x}} \right] \dot{\vec{x}}. \quad (2.56)$$

With

$$\hat{\boldsymbol{\Gamma}}_k \stackrel{\text{def}}{=} (\mathbf{g}_k^{-T} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{g}_0^\top}{\partial \vec{x}} \\ \vdots \\ \frac{\partial \mathbf{g}_3^\top}{\partial \vec{x}} \end{pmatrix} - \frac{1}{2} \frac{\partial \mathbf{G}}{\partial \vec{x}} \right]$$

(2.57)

one can also write for (2.56)

$$\ddot{x}_k = -\dot{\vec{x}}^\top \hat{\boldsymbol{\Gamma}}_k \dot{\vec{x}}.$$

(2.58)

The so-obtained matrix $\hat{\boldsymbol{\Gamma}}_k$ need not be symmetric. But the value of the quadratic form (2.58) is unchanged, if the matrix $\hat{\boldsymbol{\Gamma}}_k$ in (2.58) is replaced by the symmetric 4×4 -matrix

$$\boldsymbol{\Gamma}_k \stackrel{\text{def}}{=} \frac{1}{2} (\hat{\boldsymbol{\Gamma}}_k + \hat{\boldsymbol{\Gamma}}_k^\top). \quad (2.59)$$

In this way, the matrix is said to be *symmetrised*. Expanding (2.57) yields

$$\hat{\boldsymbol{\Gamma}}_k = \sum_{i=0}^3 g_{k,i}^{[-1]} \left(\frac{\partial \mathbf{g}_i^\top}{\partial \vec{x}} - \frac{1}{2} \frac{\partial \mathbf{G}}{\partial x_i} \right),$$

and transposed

$$\hat{\boldsymbol{\Gamma}}_k^\top = \sum_{i=0}^3 g_{k,i}^{[-1]} \left(\frac{\partial \mathbf{g}_i}{\partial \vec{x}^\top} - \frac{1}{2} \frac{\partial \mathbf{G}}{\partial x_i} \right).$$

For this one can also write

$$\hat{\boldsymbol{\Gamma}}_k^\top = (\mathbf{g}_k^{-T} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{g}_0}{\partial \vec{x}^\top} \\ \vdots \\ \frac{\partial \mathbf{g}_3}{\partial \vec{x}^\top} \end{pmatrix} - \frac{1}{2} \frac{\partial \mathbf{G}}{\partial \vec{x}} \right]. \quad (2.60)$$

Inserting (2.60) into (2.59), we obtain

$$\boldsymbol{\Gamma}_k = \frac{1}{2}(\hat{\boldsymbol{\Gamma}}_k + \hat{\boldsymbol{\Gamma}}_k^T) = \frac{1}{2}(\mathbf{g}_k^{-T} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{g}_0^T}{\partial \vec{x}} \\ \vdots \\ \frac{\partial \mathbf{g}_3^T}{\partial \vec{x}} \end{pmatrix} + \begin{pmatrix} \frac{\partial \mathbf{g}_0}{\partial \vec{x}^T} \\ \vdots \\ \frac{\partial \mathbf{g}_3}{\partial \vec{x}^T} \end{pmatrix} - \frac{\partial \mathbf{G}}{\partial \vec{x}} \right]. \quad (2.61)$$

Multiplying out yields the results for the components of the Christoffel matrix $\boldsymbol{\Gamma}_k$, namely the above already derived relationship

$$\Gamma_{\alpha\beta}^k = \sum_{i=0}^3 \frac{g_{ki}^{[-1]}}{2} \left(\frac{\partial g_{\beta i}}{\partial x_\alpha} + \frac{\partial g_{\alpha i}}{\partial x_\beta} - \frac{\partial g_{\alpha\beta}}{\partial x_i} \right). \quad (2.62)$$

This can with

$$\check{\Gamma}_{\alpha\beta}^i \stackrel{\text{def}}{=} \frac{1}{2} \left(\frac{\partial g_{\beta i}}{\partial x_\alpha} + \frac{\partial g_{\alpha i}}{\partial x_\beta} - \frac{\partial g_{\alpha\beta}}{\partial x_i} \right) \quad (2.63)$$

also be written as

$$\Gamma_{\alpha\beta}^k = \sum_{i=0}^3 g_{ki}^{[-1]} \check{\Gamma}_{\alpha\beta}^i. \quad (2.64)$$

In addition, from (2.63) the interesting connection follows:

$$\frac{\partial g_{\alpha i}}{\partial x_\beta} = \check{\Gamma}_{\alpha\beta}^i + \check{\Gamma}_{i\beta}^\alpha. \quad (2.65)$$

Assembling the four components \ddot{x}_k of the vector $\ddot{\vec{x}}$ with the matrix $\boldsymbol{\Gamma}_k$ into one vector yields

$$\ddot{\vec{x}} = - \begin{pmatrix} \dot{\vec{x}}^T \boldsymbol{\Gamma}_0 \dot{\vec{x}} \\ \vdots \\ \dot{\vec{x}}^T \boldsymbol{\Gamma}_3 \dot{\vec{x}} \end{pmatrix}, \quad (2.66)$$

respectively,

$$\ddot{\vec{x}} = -(\mathbf{I}_4 \otimes \dot{\vec{x}}^T) \boldsymbol{\Gamma} \dot{\vec{x}}, \quad (2.67)$$

with

$$\boldsymbol{\Gamma} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{\Gamma}_0 \\ \vdots \\ \boldsymbol{\Gamma}_3 \end{pmatrix} = \frac{1}{2}(\mathbf{G}^{-1} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{g}_0^T}{\partial \vec{x}} \\ \vdots \\ \frac{\partial \mathbf{g}_3^T}{\partial \vec{x}} \end{pmatrix} + \begin{pmatrix} \frac{\partial \mathbf{g}_0}{\partial \vec{x}^T} \\ \vdots \\ \frac{\partial \mathbf{g}_3}{\partial \vec{x}^T} \end{pmatrix} - \frac{\partial \mathbf{G}}{\partial \vec{x}} \right]. \quad (2.68)$$

This can also be written as

$$\boldsymbol{\Gamma} = \frac{1}{2} (\mathbf{G}^{-1} \otimes \mathbf{I}_4) \left[(\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}) \frac{\partial \mathbf{G}}{\partial \vec{x}} + \begin{pmatrix} \frac{\partial \mathbf{g}_0}{\partial \vec{x}^\top} \\ \vdots \\ \frac{\partial \mathbf{g}_3}{\partial \vec{x}^\top} \end{pmatrix} \right]. \quad (2.69)$$

Introducing the matrix

$$\check{\boldsymbol{\Gamma}} \stackrel{\text{def}}{=} \frac{1}{2} \left[(\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}) \frac{\partial \mathbf{G}}{\partial \vec{x}} + \begin{pmatrix} \frac{\partial \mathbf{g}_0}{\partial \vec{x}^\top} \\ \vdots \\ \frac{\partial \mathbf{g}_3}{\partial \vec{x}^\top} \end{pmatrix} \right], \quad (2.70)$$

it can also be written as

$$\boldsymbol{\Gamma} = (\mathbf{G}^{-1} \otimes \mathbf{I}_4) \check{\boldsymbol{\Gamma}}. \quad (2.71)$$

The matrix difference $\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}$, appearing in the matrix $\check{\boldsymbol{\Gamma}}$, has the remarkable property that the first, $(4+2)$ th, $(8+3)$ th and the 16th row (resp., column) are equal to a zero-row (resp., zero-column)! For the matrix $\check{\boldsymbol{\Gamma}}$ this has the consequence that the corresponding rows consist of $\frac{\partial g_{00}}{\partial \vec{x}^\top}$, $\frac{\partial g_{11}}{\partial \vec{x}^\top}$, $\frac{\partial g_{22}}{\partial \vec{x}^\top}$ and $\frac{\partial g_{33}}{\partial \vec{x}^\top}$. Furthermore, from (2.71) it follows that

$$\check{\boldsymbol{\Gamma}} = (\mathbf{G} \otimes \mathbf{I}_4) \boldsymbol{\Gamma}, \quad (2.72)$$

that is,

$$\check{\boldsymbol{\Gamma}}_k = (\mathbf{g}_k^\top \otimes \mathbf{I}_4) \boldsymbol{\Gamma} = g_{k0} \boldsymbol{\Gamma}_0 + \cdots + g_{k3} \boldsymbol{\Gamma}_3,$$

therefore,

$$\check{\Gamma}_{\alpha\beta}^k = \sum_{i=0}^3 g_{ki} \Gamma_{\alpha\beta}^i. \quad (2.73)$$

2.3.1 Alternative Geodesic Equation of Motion

Again the equations of motion can be modified as follows: Firstly,

$$(\dot{\vec{x}}^\top \otimes \mathbf{I}_4) \frac{\partial \mathbf{G}}{\partial \vec{x}} = \frac{\partial \mathbf{G}}{\partial \vec{x}^\top} (\dot{\vec{x}} \otimes \mathbf{I}_4), \quad (2.74)$$

and secondly,

$$(\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \frac{\partial \mathbf{G}}{\partial \vec{x}} \dot{\vec{x}} = \begin{pmatrix} \dot{\vec{x}}^\top \mathbf{G}_0 \dot{\vec{x}} \\ \vdots \\ \dot{\vec{x}}^\top \mathbf{G}_3 \dot{\vec{x}} \end{pmatrix}.$$

Using the vec -operator from the Appendix (A.51) to the scalar component $\dot{\bar{x}}^\top \mathbf{G}_k \dot{\bar{x}}$ yields

$$\text{vec}(\dot{\bar{x}}^\top \mathbf{G}_k \dot{\bar{x}}) = (\dot{\bar{x}}^\top \otimes \dot{\bar{x}}^\top) \text{vec}(\mathbf{G}_k) = (\text{vec}(\mathbf{G}_k))^\top (\dot{\bar{x}} \otimes \dot{\bar{x}}),$$

so

$$(\mathbf{I}_4 \otimes \dot{\bar{x}}^\top) \frac{\partial \mathbf{G}}{\partial \bar{x}} \dot{\bar{x}} = \overline{\frac{\partial \mathbf{G}}{\partial \bar{x}^\top}} (\dot{\bar{x}} \otimes \dot{\bar{x}}), \quad (2.75)$$

with

$$\overline{\frac{\partial \mathbf{G}}{\partial \bar{x}^\top}} \stackrel{\text{def}}{=} \begin{pmatrix} (\text{vec}(\mathbf{G}_0))^\top \\ \vdots \\ (\text{vec}(\mathbf{G}_3))^\top \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{0,0}^\top & \mathbf{g}_{0,3}^\top \\ \mathbf{g}_{1,0}^\top & \cdots & \mathbf{g}_{1,3}^\top \\ \mathbf{g}_{2,0}^\top & & \mathbf{g}_{2,3}^\top \\ \mathbf{g}_{3,0}^\top & & & \mathbf{g}_{3,3}^\top \end{pmatrix} \in \mathbb{R}^{4 \times 16}, \quad (2.76)$$

where $\mathbf{g}_{i,j}^\top$ is the j th row of \mathbf{G}_i .

By the “method of careful examination”, one can write

$$\overline{\frac{\partial \mathbf{G}}{\partial \bar{x}^\top}} = \begin{pmatrix} \mathbf{g}_{0,0}^\top \\ \mathbf{g}_{1,0}^\top \\ \mathbf{g}_{2,0}^\top \\ \mathbf{g}_{3,0}^\top \\ \vdots \\ \mathbf{g}_{0,3}^\top \\ \mathbf{g}_{1,3}^\top \\ \mathbf{g}_{2,3}^\top \\ \mathbf{g}_{3,3}^\top \end{pmatrix}^B = \left(\mathbf{U}_{4 \times 4} \frac{\partial \mathbf{G}}{\partial \bar{x}} \right)^B, \quad (2.77)$$

where the superscript “ B ” means the *block-transposition* of the corresponding matrix. The block-transposition of a block matrix is defined as

$$\mathbf{A}^B \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{pmatrix}^B = (\mathbf{A}_1 \quad \dots \quad \mathbf{A}_n).$$

Equation (2.75) used in (2.48) yields

$$\ddot{\bar{x}} = -\mathbf{G}^{-1} \left[\frac{\partial \mathbf{G}}{\partial \bar{x}^\top} - \frac{1}{2} \overline{\frac{\partial \mathbf{G}}{\partial \bar{x}^\top}} \right] (\dot{\bar{x}} \otimes \dot{\bar{x}}). \quad (2.78)$$

With

$$\tilde{\boldsymbol{\Gamma}} \stackrel{\text{def}}{=} \mathbf{G}^{-1} \left[\frac{\partial \mathbf{G}}{\partial \bar{x}^\top} - \frac{1}{2} \overline{\frac{\partial \mathbf{G}}{\partial \bar{x}^\top}} \right] \in \mathbb{R}^{4 \times 16},$$

(2.79)

one finally obtains

$$\ddot{\vec{x}} = -\tilde{\Gamma}(\dot{\vec{x}} \otimes \dot{\vec{x}}). \quad (2.80)$$

One can also reach the form (2.80) as follows: It is $\ddot{x}_k = \dot{\vec{x}}^\top \boldsymbol{\Gamma}_k \dot{\vec{x}}$. Applying to this the vec -operator (see Appendix), one obtains

$$-\ddot{x}_k = \text{vec}(\dot{\vec{x}}^\top \boldsymbol{\Gamma}_k \dot{\vec{x}}) = (\dot{\vec{x}}^\top \otimes \dot{\vec{x}}^\top) \text{vec}(\boldsymbol{\Gamma}_k) = (\text{vec}(\boldsymbol{\Gamma}_k))^\top (\dot{\vec{x}} \otimes \dot{\vec{x}}).$$

With

$$\tilde{\boldsymbol{\Gamma}} \stackrel{\text{def}}{=} \begin{pmatrix} (\text{vec}(\boldsymbol{\Gamma}_0))^\top \\ \vdots \\ (\text{vec}(\boldsymbol{\Gamma}_3))^\top \end{pmatrix} \quad (2.81)$$

(2.80) is obtained again. Once again, we can write

$$\tilde{\boldsymbol{\Gamma}} = \left(U_{4 \times 4} \frac{\partial \boldsymbol{\Gamma}}{\partial \vec{x}} \right)^B. \quad (2.82)$$

Just a word regarding the derivatives with respect to s . If s^2 and ds^2 are positive, then it is a so-called time-like event. Then

$$(ds)^2 = c^2 dt^2 - dx^\top dx = c^2 dt^2 - \dot{x}^\top \dot{x} dt^2,$$

$$ds = \sqrt{c^2 - v^2} dt,$$

so

$$\gamma ds = c dt.$$

Substituting $ds = c d\tau$, one obtains

$$\gamma d\tau = dt.$$

A comparison with the results of special relativity theory provides $d\tau = dt'$, i.e. the time that elapses in the moving coordinate system \mathcal{X}' . One calls in this context τ as *proper time*.

2.4 Example: Uniformly Rotating Systems

We will consider a fixed inertial frame \mathcal{X} with the coordinates t, x, y and z and a uniformly around the z -axis *rotating* coordinate system \mathcal{K} with the coordinates τ, r, φ and z . Then the transformation equations are

$$\begin{aligned} t &= \tau, \\ x &= r \cos(\varphi + \omega t), \\ y &= r \sin(\varphi + \omega t), \\ z &= z. \end{aligned} \quad (2.83)$$

As Jacobi matrix \mathbf{J} one obtains

$$\mathbf{J} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -r\frac{\omega}{c}\sin(\varphi + \omega t) & \cos(\varphi + \omega t) & -r\sin(\varphi + \omega t) & 0 \\ r\frac{\omega}{c}\cos(\varphi + \omega t) & \sin(\varphi + \omega t) & r\cos(\varphi + \omega t) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.84)$$

and as the metric matrix

$$\mathbf{G} = \mathbf{J}^\top \mathbf{M} \mathbf{J} = \begin{pmatrix} 1 - r^2\frac{\omega^2}{c^2} & 0 & -r^2\frac{\omega}{c} & 0 \\ 0 & -1 & 0 & 0 \\ -r^2\frac{\omega}{c} & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (2.85)$$

and from this

$$\mathbf{G}^{-1} = \begin{pmatrix} 1 & 0 & -\frac{\omega}{c} & 0 \\ 0 & -1 & 0 & 0 \\ -\frac{\omega}{c} & 0 & \frac{\omega^2}{c^2} - \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (2.86)$$

In this case, ds^2 has the value

$$ds^2 = \left(1 - r^2\frac{\omega^2}{c^2}\right) d\tau^2 - dr^2 - r d\varphi^2 - 2r^2\frac{\omega}{c} d\varphi d\tau - dz^2. \quad (2.87)$$

If a clock is in the rotating system at the position (r, θ, z) and one considers two temporally directly adjacent events with $dr = d\varphi = dz = 0$, then one obtains for the proper time ds in this case the relationship (with $v = r\omega$)

$$ds = d\tau \sqrt{1 - r^2\omega^2/c^2} = d\tau \sqrt{1 - v^2/c^2} = d\tau/\gamma.$$

This is the relationship known from the theory of special relativity! For the calculation of the acceleration, the derivatives of the metric matrix are needed. In this case, $\mathbf{G}_0 = \mathbf{G}_2 = \mathbf{G}_3 = \mathbf{0}$, but

$$\mathbf{G}_1 = \frac{\partial \mathbf{G}}{\partial r} = \begin{pmatrix} -2r\frac{\omega^2}{c^2} & 0 & -2r\frac{\omega}{c} & 0 \\ 0 & 0 & 0 & 0 \\ -2r\frac{\omega}{c} & 0 & -2r & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since only the matrix $\mathbf{G}_1 \neq \mathbf{0}$, in this case (2.56) is simplified to

$$\ddot{x}_k = \left[\frac{1}{2} (g^{k1} \dot{\tilde{x}}^1) - (\dot{x}_1 g_k^{-1}) \right] \mathbf{G}_1 \dot{\tilde{x}}.$$

In particular, with $\dot{\vec{x}}^\top = [c|\dot{r}|\dot{\varphi}|\dot{z}]$ we obtain

$$\begin{aligned}\ddot{r} &= -\frac{1}{2}\dot{\vec{x}}^\top \mathbf{G}_1 \dot{\vec{x}} \\ &= -\frac{1}{2}[-2r\omega^2/c - \dot{\varphi}2r\omega/c | 0 | -2r\omega - 2r\dot{\varphi} | 0] \dot{\vec{x}} = r(\omega + \dot{\varphi})^2\end{aligned}\quad (2.88)$$

and

$$\ddot{\varphi} = -\dot{r}[-\omega/c | 0 | \omega^2/c^2 - 1/r^2 | 0] \mathbf{G}_1 \dot{\vec{x}} = -2\dot{r}\omega/r - 2\dot{r}\dot{\varphi}/r,\quad (2.89)$$

or

$$r\ddot{\varphi} = -2\dot{r}(\dot{\varphi} + \omega).\quad (2.90)$$

Equation (2.88) multiplied by the mass m represents the centrifugal force, and (2.90) multiplied by the mass m is the so-called Coriolis force! The accelerations occurring in this rotating system are determined by the elements $g_{ij} = g_{ij}(\vec{x})$ of the coordinate-dependent metric matrix $\mathbf{G}(\vec{x})$. For a local reference system, one can always specify a coordinate transformation (with $\mathbf{J}^{-1}(\vec{x})$), so that the transformed system is obviously an inertial frame. In general, for an accelerated or non-uniformly moving (e.g. rotating) system no *globally valid* transformation matrix \mathbf{J} can be specified. The given space is *curved*!

It should now be shown that the Christoffel matrices are the same for the rotating system by applying the formula (2.18). According to (2.84),

$$\mathbf{J} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -r\frac{\omega}{c} \sin(\varphi + \omega t) & \cos(\varphi + \omega t) & -r \sin(\varphi + \omega t) & 0 \\ r\frac{\omega}{c} \cos(\varphi + \omega t) & \sin(\varphi + \omega t) & r \cos(\varphi + \omega t) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},\quad (2.91)$$

i.e.

$$\mathbf{J}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\varphi + \omega t) & \sin(\varphi + \omega t) & 0 \\ -\frac{\omega}{c} & \frac{1}{r} \sin(\varphi + \omega t) & \frac{1}{r} \cos(\varphi + \omega t) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.\quad (2.92)$$

Equations (2.88) and (2.89) can be used for the Christoffel matrices. The non-zero matrix elements are:

$$\begin{aligned}\Gamma_{00}^1 &= -r\frac{\omega^2}{c^2}, & \Gamma_{02}^1 &= -r\frac{\omega}{c}, & \Gamma_{03}^1 &= -r\frac{\omega}{c}, & \Gamma_{22}^1 &= -1, \\ \Gamma_{01}^2 &= \frac{4\omega}{rc}, \quad \text{and} \quad \Gamma_{12}^2 &= \frac{4}{r}.\end{aligned}$$

Accordingly, we obtain with (2.18), e.g.

$$\begin{aligned}\Gamma_{00}^1 &= \left[0 \mid \cos(\varphi + \omega t) \mid \sin(\varphi + \omega t) \mid 0 \right] \begin{pmatrix} 0 \\ -r \frac{\omega^2}{c^2} \cos(\varphi + \omega t) \\ -r \frac{\omega^2}{c^2} \sin(\varphi + \omega t) \\ 0 \end{pmatrix} \\ &= -r \frac{\omega^2}{c^2}.\end{aligned}$$

2.5 General Coordinate Transformations

In the theory of general relativity, the invariance of the mathematical descriptions is demanded for the general laws of nature with respect to each other arbitrarily moving coordinate systems. Even more generally:

The invariance of the mathematical descriptions with respect to arbitrary coordinate transformations is required.

2.5.1 Absolute Derivatives

First, we must clarify how the derivatives eventually have to be modified so that the derived expressions are invariant under coordinate transformations.

Suppose a vector field $\mathbf{a}(\lambda)$ is defined along a curve whose parametric representation is given by $\vec{x}(\lambda)$. Going on to another coordinate system \mathcal{K}' with \mathbf{a}' , for the mathematical description of dynamic processes one is especially interested in how the derivative $d\mathbf{a}/d\lambda$ is transformed into $d\mathbf{a}'/d\lambda$. Because $\mathbf{T} = \mathbf{T}(\vec{x}(\lambda))$, it follows that

$$\frac{d\mathbf{a}'}{d\lambda} = \frac{d(\mathbf{T}\mathbf{a})}{d\lambda} = \mathbf{T} \frac{d\mathbf{a}}{d\lambda} + \left(\frac{d\vec{x}^\top}{d\lambda} \otimes \mathbf{I}_4 \right) \frac{\partial \mathbf{T}}{\partial \vec{x}} \mathbf{a}, \quad (2.93)$$

i.e. $d\mathbf{a}/d\lambda$ is not transformed into $d\mathbf{a}'/d\lambda$ as was \mathbf{a} by a simple multiplication with the transformation matrix \mathbf{T} . For this the reason, the definition of the derivative is

$$\frac{d\mathbf{a}}{d\lambda} = \lim_{\delta\lambda \rightarrow 0} \frac{\mathbf{a}(\lambda + \delta\lambda) - \mathbf{a}(\lambda)}{\delta\lambda},$$

where the difference of the vectors is formed at different places on the curve γ , to which transformation matrices $\mathbf{T}(\lambda) \neq \mathbf{T}(\lambda + \delta\lambda)$ generally belong.

Thus in order to be always able to take the same transformation matrix, the difference of two vectors must be taken at the *same* place of the curve. It is true that

$$\delta\mathbf{a} \approx \frac{d\mathbf{a}}{d\lambda} \delta\lambda \quad (2.94)$$

and, if the shift of \mathbf{a} is taken along a geodesic line,

$$\frac{d\mathbf{a}}{d\lambda} + (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \frac{d\vec{x}}{d\lambda} = \mathbf{0},$$

so

$$\frac{d\mathbf{a}}{d\lambda} = -(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \frac{d\vec{x}}{d\lambda}. \quad (2.95)$$

Here $\frac{d\vec{x}}{d\lambda}$ is the tangent vector to the geodesic curve γ , and $\vec{x}(\lambda)$ is the parametric representation of γ . Multiplying (2.95) with $\delta\lambda$, we obtain

$$\boxed{\delta\mathbf{a} = -(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \delta\vec{x}.} \quad (2.96)$$

Moving the vector $\mathbf{a}(\lambda)$ from the position $\vec{x}(\lambda)$ parallel to the position $\vec{x}(\lambda + \delta\lambda)$, the vector

$$\bar{\mathbf{a}} \stackrel{\text{def}}{=} \mathbf{a}(\lambda) + \delta\mathbf{a}$$

is obtained, or by (2.96),

$$\bar{\mathbf{a}} \approx \mathbf{a}(\lambda) - (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \delta\vec{x}. \quad (2.97)$$

On the other hand, $\mathbf{a}(\lambda + \delta\lambda) - \bar{\mathbf{a}}$ is a vector at the position $\gamma(\lambda + \delta\lambda)$, as is $(\mathbf{a}(\lambda + \delta\lambda) - \bar{\mathbf{a}})/\delta\lambda$. As $\delta\lambda \rightarrow 0$ the quotient is always a vector at the same location that varies, however.

The limit of this ratio is called *absolute derivative* $\frac{D\mathbf{a}}{d\lambda}$ of $\mathbf{a}(\lambda)$ along the curve γ :

$$\lim_{\delta\lambda \rightarrow 0} \frac{\mathbf{a}(\lambda + \delta\lambda) - \bar{\mathbf{a}}}{\delta\lambda} \approx \frac{d\mathbf{a}}{d\lambda} + \lim_{\delta\lambda \rightarrow 0} (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \frac{\delta\vec{x}}{\delta\lambda}$$

which is obtained with (2.97), and therefore one defines the absolute derivative as

$$\boxed{\frac{D\mathbf{a}}{d\lambda} \stackrel{\text{def}}{=} \frac{d\mathbf{a}}{d\lambda} + (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \frac{d\vec{x}}{d\lambda} = \dot{\mathbf{a}} + (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \dot{\vec{x}}.} \quad (2.98)$$

The differentiated vector can be decomposed into

$$\dot{\mathbf{a}} = \frac{\partial \mathbf{a}}{\partial \vec{x}^\top} \dot{\vec{x}}, \quad (2.99)$$

so that in (2.98) one can extract $\dot{\vec{x}}$ to the right:

$$\boxed{\frac{D\mathbf{a}}{d\lambda} = \left[\frac{\partial \mathbf{a}}{\partial \vec{x}^\top} + (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \right] \dot{\vec{x}}.} \quad (2.100)$$

The expression appearing in brackets is called the *covariant derivative of \mathbf{a}* and written $\mathbf{a}_{||\vec{x}^\top} (\in \mathbb{R}^{4 \times 4})$:

$$\boxed{\mathbf{a}_{||\vec{x}^\top} \stackrel{\text{def}}{=} \frac{\partial \mathbf{a}}{\partial \vec{x}^\top} + (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma}.} \quad (2.101)$$

This covariant derivative $\mathbf{a}_{||\vec{x}^\top}$ becomes the normal partial derivative $\frac{\partial}{\partial \vec{x}^\top}$, if $\boldsymbol{\Gamma} = \mathbf{0}$, that means, there is no gravitational field.

2.5.2 Transformation of the Christoffel Matrix $\tilde{\boldsymbol{\Gamma}}$

In (2.22), the Christoffel matrix is defined:

$$\tilde{\boldsymbol{\Gamma}} \stackrel{\text{def}}{=} \mathbf{J}^{-1} \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} = \frac{\partial \vec{x}}{\partial \vec{\xi}^\top} \cdot \frac{\partial^2 \vec{\xi}}{\partial \vec{x}^\top \partial \vec{x}^\top}. \quad (2.102)$$

If one changes over from the coordinate system with \vec{x} to the coordinate system with \vec{x}' , one obtains with the transformation matrices

$$\mathbf{T} \stackrel{\text{def}}{=} \frac{\partial \vec{x}'}{\partial \vec{x}^\top} \quad (2.103)$$

and

$$\bar{\mathbf{T}} \stackrel{\text{def}}{=} \frac{\partial \vec{x}}{\partial \vec{x}'^\top}, \quad (2.104)$$

the Christoffel matrix $\tilde{\boldsymbol{\Gamma}}'$ in the coordinate system with \vec{x}'

$$\begin{aligned} \tilde{\boldsymbol{\Gamma}}' &\stackrel{\text{def}}{=} \frac{\partial \vec{x}'}{\partial \vec{\xi}^\top} \cdot \frac{\partial^2 \vec{\xi}}{\partial \vec{x}'^\top \partial \vec{x}'^\top} = \underbrace{\frac{\partial \vec{x}'}{\partial \vec{x}^\top} \frac{\partial \vec{x}}{\partial \vec{\xi}^\top}}_{\mathbf{J}^{-1}} \underbrace{\frac{\partial}{\partial \vec{x}'^\top}}_{\bar{\mathbf{T}}} \left(\frac{\partial \vec{\xi}}{\partial \vec{x}'^\top} \right) \\ &= \mathbf{T} \cdot \underbrace{\frac{\partial \vec{x}}{\partial \vec{\xi}^\top}}_{\mathbf{J}} \underbrace{\frac{\partial}{\partial \vec{x}'^\top}}_{\bar{\mathbf{T}}} \left(\underbrace{\frac{\partial \vec{\xi}}{\partial \vec{x}^\top}}_{\mathbf{J}} \cdot \underbrace{\frac{\partial \vec{x}}{\partial \vec{x}'^\top}}_{\bar{\mathbf{T}}} \right). \end{aligned} \quad (2.105)$$

According to the product and chain rules, one obtains

$$\begin{aligned} \frac{\partial}{\partial \vec{x}'^\top} (\mathbf{J} \cdot \bar{\mathbf{T}}) &= \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} (\mathbf{I}_4 \otimes \bar{\mathbf{T}}) + \mathbf{J} \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'^\top} \\ &= \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} (\bar{\mathbf{T}} \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \bar{\mathbf{T}}) + \mathbf{J} \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}^\top} (\bar{\mathbf{T}} \otimes \mathbf{I}_4) \\ &= \frac{\partial \mathbf{J}}{\partial \vec{x}^\top} (\bar{\mathbf{T}} \otimes \bar{\mathbf{T}}) + \mathbf{J} \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}^\top} (\bar{\mathbf{T}} \otimes \mathbf{I}_4). \end{aligned} \quad (2.106)$$

Inserting (2.106) in (2.105) reveals

$$\tilde{\Gamma}' = \mathbf{T} \tilde{\Gamma} (\bar{\mathbf{T}} \otimes \bar{\mathbf{T}}) + \mathbf{T} \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}^T} (\bar{\mathbf{T}} \otimes \mathbf{I}_4). \quad (2.107)$$

The second term on the right-hand side shows the coordinate dependence of the transformation matrix \mathbf{T} .

Another important characteristic is obtained as follows. Differentiating $\mathbf{I}_4 = \mathbf{T} \bar{\mathbf{T}}$ with respect to \vec{x}^T yields

$$\mathbf{O} = \frac{\partial \mathbf{T}}{\partial \vec{x}^T} (\mathbf{I}_4 \otimes \bar{\mathbf{T}}) + \mathbf{T} \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}^T},$$

i.e.

$$\mathbf{T} \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}^T} = - \frac{\partial \mathbf{T}}{\partial \vec{x}^T} (\mathbf{I}_4 \otimes \bar{\mathbf{T}}). \quad (2.108)$$

Putting this in (2.107), one gets another form of the transformed Christoffel matrix, namely

$$\tilde{\Gamma}' = \mathbf{T} \tilde{\Gamma} (\bar{\mathbf{T}} \otimes \bar{\mathbf{T}}) - \frac{\partial \mathbf{T}}{\partial \vec{x}^T} (\bar{\mathbf{T}} \otimes \bar{\mathbf{T}}). \quad (2.109)$$

Moreover,

$$\frac{d^2 \vec{x}'}{d\tau^2} = \frac{d}{d\tau} \left(\underbrace{\frac{\partial \vec{x}'}{\partial \vec{x}^T}}_T \cdot \frac{d\vec{x}}{d\tau} \right) = \mathbf{T} \frac{d^2 \vec{x}}{d\tau^2} + \frac{\partial \mathbf{T}}{\partial \vec{x}^T} \underbrace{\left(\frac{d\vec{x}}{d\tau} \otimes \mathbf{I}_4 \right)}_{(\frac{d\vec{x}}{d\tau} \otimes \frac{d\vec{x}}{d\tau})} \frac{d\vec{x}}{d\tau}. \quad (2.110)$$

Multiplying (2.109) by the vector $(\frac{d\vec{x}'}{d\tau} \otimes \frac{d\vec{x}'}{d\tau})$ from the right yields

$$\begin{aligned} \tilde{\Gamma}' \left(\frac{d\vec{x}'}{d\tau} \otimes \frac{d\vec{x}'}{d\tau} \right) &= \mathbf{T} \tilde{\Gamma} (\bar{\mathbf{T}} \otimes \bar{\mathbf{T}}) \underbrace{\left(\frac{d\vec{x}'}{d\tau} \otimes \frac{d\vec{x}'}{d\tau} \right)}_{(\frac{d\vec{x}}{d\tau} \otimes \frac{d\vec{x}}{d\tau})} \\ &\quad - \underbrace{\frac{\partial \mathbf{T}}{\partial \vec{x}^T} (\bar{\mathbf{T}} \otimes \bar{\mathbf{T}})}_{(\frac{d\vec{x}}{d\tau} \otimes \frac{d\vec{x}}{d\tau})} \left(\frac{d\vec{x}'}{d\tau} \otimes \frac{d\vec{x}'}{d\tau} \right). \end{aligned} \quad (2.111)$$

Adding the two equations (2.110) and (2.111), we finally obtain

$$\frac{d^2 \vec{x}'}{d\tau^2} + \tilde{\Gamma}' \left(\frac{d\vec{x}'}{d\tau} \otimes \frac{d\vec{x}'}{d\tau} \right) = \mathbf{T} \left[\frac{d^2 \vec{x}}{d\tau^2} + \tilde{\Gamma} \left(\frac{d\vec{x}}{d\tau} \otimes \frac{d\vec{x}}{d\tau} \right) \right]. \quad (2.112)$$

The vector in the brackets in (2.112) is transformed as a vector in general! The equation of motion is invariant.

2.5.3 Transformation of the Christoffel Matrix $\hat{\Gamma}$

By (2.15),

$$\hat{\Gamma} = \mathbf{U}_{4 \times 4} \begin{pmatrix} \mathbf{J}^{-1} \mathbf{J}_0 \\ \vdots \\ \mathbf{J}^{-1} \mathbf{J}_3 \end{pmatrix} = \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}}, \quad (2.113)$$

where

$$\mathbf{J} = \frac{\partial \vec{\xi}}{\partial \vec{x}^T}.$$

Defining furthermore

$$\mathbf{J}' = \frac{\partial \vec{\xi}}{\partial \vec{x}'^T}, \quad (2.114)$$

we obtain with the transformation matrices \mathbf{T} and $\bar{\mathbf{T}}$ the relation

$$\mathbf{J} = \frac{\partial \vec{\xi}}{\partial \vec{x}^T} = \frac{\partial \vec{\xi}}{\partial \vec{x}'^T} \frac{\partial \vec{x}'}{\partial \vec{x}^T} = \frac{\partial \vec{\xi}}{\partial \vec{x}'^T} \mathbf{T}, \quad (2.115)$$

i.e. with (2.114)

$$\mathbf{J} = \mathbf{J}' \mathbf{T}, \quad (2.116)$$

or

$$\mathbf{J}' = \mathbf{J} \bar{\mathbf{T}}. \quad (2.117)$$

Hence

$$\hat{\Gamma}' = \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{J}'^{-1}) \frac{\partial \mathbf{J}'}{\partial \vec{x}'} \quad (2.118)$$

With

$$\mathbf{J}'^{-1} = \mathbf{T} \mathbf{J}^{-1}$$

and

$$\frac{\partial \mathbf{J}'}{\partial \vec{x}'} = \frac{\partial \mathbf{J} \bar{\mathbf{T}}}{\partial \vec{x}'} = \frac{\partial \mathbf{J}}{\partial \vec{x}'} \bar{\mathbf{T}} + (\mathbf{I}_4 \otimes \mathbf{J}) \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'},$$

we then obtain

$$\begin{aligned} \hat{\Gamma}' &= \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{T}) (\mathbf{I}_4 \otimes \mathbf{J}^{-1}) \left(\frac{\partial \mathbf{J}}{\partial \vec{x}'} \bar{\mathbf{T}} + (\mathbf{I}_4 \otimes \mathbf{J}) \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'} \right) \\ &= \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{T}) \left((\mathbf{I}_4 \otimes \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}'} \bar{\mathbf{T}} + \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'} \right). \end{aligned} \quad (2.119)$$

Furthermore,

$$\frac{\partial \mathbf{J}}{\partial \vec{x}'} = \left(\frac{\partial \vec{x}^\top}{\partial \vec{x}'} \otimes \mathbf{I}_4 \right) \frac{\partial \mathbf{J}}{\partial \vec{x}} = (\bar{\mathbf{T}}^\top \otimes \mathbf{I}_4) \frac{\partial \mathbf{J}}{\partial \vec{x}}, \quad (2.120)$$

so

$$\begin{aligned} \hat{\Gamma}' &= \mathbf{U}_{4 \times 4} \left[(\bar{\mathbf{T}}^\top \otimes \mathbf{T} \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}} \bar{\mathbf{T}} + (\mathbf{I}_4 \otimes \mathbf{T}) \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'} \right] \\ &= \underbrace{\mathbf{U}_{4 \times 4} (\bar{\mathbf{T}}^\top \otimes \mathbf{T}) \mathbf{U}_{4 \times 4}}_{\mathbf{T} \otimes \bar{\mathbf{T}}^\top} \underbrace{\mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{J}^{-1}) \frac{\partial \mathbf{J}}{\partial \vec{x}} \bar{\mathbf{T}}}_{\hat{\Gamma}} + \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{T}) \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'}, \end{aligned}$$

and with (2.113)

$$\boxed{\hat{\Gamma}' = (\mathbf{T} \otimes \bar{\mathbf{T}}^\top) \hat{\Gamma} \bar{\mathbf{T}} + \mathbf{U}_{4 \times 4} (\mathbf{I}_4 \otimes \mathbf{T}) \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'}.} \quad (2.121)$$

The second term on the right-hand side expresses again the dependence on the transformation matrix \mathbf{T} .

2.5.4 Coordinate Transformation and Covariant Derivative

It is true that

$$\mathbf{T} \bar{\mathbf{T}} = \mathbf{I},$$

so

$$\frac{\partial}{\partial \vec{x}'} (\mathbf{T} \bar{\mathbf{T}}) = \mathbf{0} = \frac{\partial \mathbf{T}}{\partial \vec{x}'} \bar{\mathbf{T}} + (\mathbf{I}_4 \otimes \mathbf{T}) \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'},$$

or

$$(\mathbf{I}_4 \otimes \mathbf{T}) \frac{\partial \bar{\mathbf{T}}}{\partial \vec{x}'} = - \frac{\partial \mathbf{T}}{\partial \vec{x}'} \bar{\mathbf{T}}. \quad (2.122)$$

Look now at the coordinate transformation

$$\mathbf{a}' = \mathbf{T} \mathbf{a} \quad (2.123)$$

and its partial derivative with respect to \vec{x}' :

$$\frac{\partial \mathbf{a}'}{\partial \vec{x}'} = \frac{\partial}{\partial \vec{x}'} (\mathbf{T} \mathbf{a}) = \frac{\partial \mathbf{T}}{\partial \vec{x}'} \mathbf{a} + (\mathbf{I}_4 \otimes \mathbf{T}) \frac{\partial \mathbf{a}}{\partial \vec{x}}. \quad (2.124)$$

On the other hand,

$$\frac{\partial \mathbf{a}}{\partial \vec{x}'} = \left(\frac{\partial \vec{x}^\top}{\partial \vec{x}'} \otimes \mathbf{I}_4 \right) \frac{\partial \mathbf{a}}{\partial \vec{x}} = (\bar{\mathbf{T}}^\top \otimes \mathbf{I}_4) \frac{\partial \mathbf{a}}{\partial \vec{x}}.$$

Inserting this into (2.124) yields

$$\frac{\partial \mathbf{a}'}{\partial \vec{x}'} = (\bar{T}^\top \otimes T) \frac{\partial \mathbf{a}}{\partial \vec{x}} + \frac{\partial T}{\partial \vec{x}'} \mathbf{a}. \quad (2.125)$$

How does the product of the matrix Γ and the vector \mathbf{a} transform? We have

$$\Gamma' \mathbf{a}' = (T \otimes \bar{T}^\top) \Gamma \bar{T} T \mathbf{a} + U_{4 \times 4} (I_4 \otimes T) \frac{\partial T^{-1}}{\partial \vec{x}'} T \mathbf{a}, \quad (2.126)$$

and hence obtain

$$\frac{\partial T^{-1}}{\partial \vec{x}'} = -(I_4 \otimes T^{-1}) \frac{\partial T}{\partial \vec{x}'} T^{-1}.$$

This in (2.126) yields

$$\Gamma' \mathbf{a}' = (T \otimes \bar{T}^\top) \Gamma \mathbf{a} - U_{4 \times 4} \frac{\partial T}{\partial \vec{x}'} \mathbf{a}. \quad (2.127)$$

After adding (2.125) and (2.127), multiplied from the left by $U_{4 \times 4}$, we get

$$\frac{\partial \mathbf{a}'}{\partial \vec{x}'} + U_{4 \times 4} \Gamma' \mathbf{a}' = (\bar{T}^\top \otimes T) \frac{\partial \mathbf{a}}{\partial \vec{x}} + U_{4 \times 4} (T \otimes \bar{T}^\top) U_{4 \times 4} U_{4 \times 4} \Gamma \mathbf{a},$$

so

$$\frac{\partial \mathbf{a}'}{\partial \vec{x}'} + U_{4 \times 4} \Gamma' \mathbf{a}' = (\bar{T}^\top \otimes T) \left[\frac{\partial \mathbf{a}}{\partial \vec{x}} + U_{4 \times 4} \Gamma \mathbf{a} \right]. \quad (2.128)$$

With the definition

$$\Gamma^* \stackrel{\text{def}}{=} U_{4 \times 4} \Gamma = \frac{1}{2} (I \otimes G^{-1}) \left[(I_{16} - U_{4 \times 4}) \frac{\partial G}{\partial \vec{x}} + U_{4 \times 4} \begin{pmatrix} \frac{\partial g_0}{\partial \vec{x}^\top} \\ \vdots \\ \frac{\partial g_3}{\partial \vec{x}^\top} \end{pmatrix} \right] \quad (2.129)$$

the following compact relation is obtained:

$$\frac{\partial \mathbf{a}'}{\partial \vec{x}'} + \Gamma^{*\prime} \mathbf{a}' = (\bar{T}^\top \otimes T) \left[\frac{\partial \mathbf{a}}{\partial \vec{x}} + \Gamma^* \mathbf{a} \right].$$

(2.130)

On the right-hand side of this equation, a vector from \mathbb{R}^{16} is multiplied by a 16×16 -matrix! We introduce the abbreviations

$$\mathbf{a}_{|\vec{x}} \stackrel{\text{def}}{=} \frac{\partial \mathbf{a}}{\partial \vec{x}} \quad (2.131)$$

and

$$\underline{\mathbf{a}_{||\vec{x}}} \stackrel{\text{def}}{=} \underline{\mathbf{a}_{|\vec{x}}} + \underline{\Gamma^* \mathbf{a}}, \quad (2.132)$$

and call $\mathbf{a}_{\parallel \vec{x}}$ the *covariant derivative of \mathbf{a} with respect to \vec{x}* . Now (2.130) is written as

$$\boxed{\mathbf{a}'_{\parallel \vec{x}} = (\bar{\mathbf{T}}^\top \otimes \mathbf{T}) \mathbf{a}_{\parallel \vec{x}}.} \quad (2.133)$$

The form on the right-hand side of (2.130) reminds us of the right-hand side of the expression used in the *vec*-operator (see Appendix), that is,

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{B}). \quad (2.134)$$

To use this lemma, (2.130) is first written in some detail. With the new matrix

$$\boldsymbol{\Gamma}_k^* \stackrel{\text{def}}{=} (\mathbf{i}_k^\top \otimes \mathbf{G}^{-1}) \left[\frac{1}{2} \mathbf{I}_{16} - \mathbf{U}_{4 \times 4} \right] \frac{\partial \mathbf{G}}{\partial \vec{x}}, \quad (2.135)$$

where \mathbf{i}_k^\top is the k th row of the unit matrix \mathbf{I}_4 , one obtains for (2.130)

$$\begin{aligned} \begin{pmatrix} \frac{\partial \mathbf{a}'}{\partial x'_0} \\ \vdots \\ \frac{\partial \mathbf{a}'}{\partial x'_3} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Gamma}_0^{*\prime} \mathbf{a}' \\ \vdots \\ \boldsymbol{\Gamma}_3^{*\prime} \mathbf{a}' \end{pmatrix} &= (\mathbf{I}_4^\top \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{a}'}{\partial x'_0} \\ \vdots \\ \frac{\partial \mathbf{a}'}{\partial x'_3} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Gamma}_0^{*\prime} \mathbf{a}' \\ \vdots \\ \boldsymbol{\Gamma}_3^{*\prime} \mathbf{a}' \end{pmatrix} \right] \\ &= (\bar{\mathbf{T}}^\top \otimes \mathbf{T}) \left[\begin{pmatrix} \frac{\partial \mathbf{a}}{\partial x_0} \\ \vdots \\ \frac{\partial \mathbf{a}}{\partial x_3} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Gamma}_0^* \mathbf{a} \\ \vdots \\ \boldsymbol{\Gamma}_3^* \mathbf{a} \end{pmatrix} \right]. \end{aligned} \quad (2.136)$$

Now Lemma (2.134) applied to this equation shows that

$$\begin{aligned} &\left[\frac{\partial \mathbf{a}'}{\partial x'_0} \Big| \dots \Big| \frac{\partial \mathbf{a}'}{\partial x'_3} \right] + [\boldsymbol{\Gamma}_0^{*\prime} \mathbf{a}' | \dots | \boldsymbol{\Gamma}_3^{*\prime} \mathbf{a}'] \\ &= \mathbf{T} \left\{ \left[\frac{\partial \mathbf{a}}{\partial x_0} \Big| \dots \Big| \frac{\partial \mathbf{a}}{\partial x_3} \right] + [\boldsymbol{\Gamma}_0^* \mathbf{a} | \dots | \boldsymbol{\Gamma}_3^* \mathbf{a}] \right\} \mathbf{T}^{-1}. \end{aligned} \quad (2.137)$$

With

$$\frac{\partial \mathbf{a}}{\partial \vec{x}^\top} \stackrel{\text{def}}{=} \left[\frac{\partial \mathbf{a}}{\partial x_0} \Big| \dots \Big| \frac{\partial \mathbf{a}}{\partial x_3} \right] \in \mathbb{R}^{4 \times 4} \quad \text{and} \quad \bar{\boldsymbol{\Gamma}} \stackrel{\text{def}}{=} [\boldsymbol{\Gamma}_0^* | \dots | \boldsymbol{\Gamma}_3^*] \in \mathbb{R}^{4 \times 16}$$

one obtains

$$\boxed{\frac{\partial \mathbf{a}'}{\partial \vec{x}'^\top} + \bar{\boldsymbol{\Gamma}}' (\mathbf{I}_4 \otimes \mathbf{a}') = \mathbf{T} \left[\frac{\partial \mathbf{a}}{\partial \vec{x}^\top} + \bar{\boldsymbol{\Gamma}} (\mathbf{I}_4 \otimes \mathbf{a}) \right] \mathbf{T}^{-1}.} \quad (2.138)$$

The sum of matrices

$$\frac{\partial \mathbf{a}}{\partial \vec{x}^\top} + \bar{\boldsymbol{\Gamma}} (\mathbf{I}_4 \otimes \mathbf{a})$$

is therefore transformed by a normal similarity transformation in to the following sum of matrices

$$\frac{\partial \mathbf{a}'}{\partial \vec{x}'^\top} + \bar{\Gamma}'(\mathbf{I}_4 \otimes \mathbf{a}').$$

And again some abbreviations are introduced (now the vector \vec{x} is transposed!):

$$\mathbf{a}_{|\vec{x}^\top} \stackrel{\text{def}}{=} \frac{\partial \mathbf{a}}{\partial \vec{x}^\top} \quad (2.139)$$

and

$$\mathbf{a}_{\parallel \vec{x}^\top} \stackrel{\text{def}}{=} \mathbf{a}_{|\vec{x}^\top} + \bar{\Gamma}(\mathbf{I}_4 \otimes \mathbf{a}) \in \mathbb{R}^{4 \times 4}. \quad (2.140)$$

$\mathbf{a}_{\parallel \vec{x}^\top}$ is called again the *covariant derivative* of \mathbf{a} and it is

$$\mathbf{a}'_{\parallel \vec{x}^\top} = \mathbf{T} \mathbf{a}_{\parallel \vec{x}^\top} \mathbf{T}^{-1}. \quad (2.141)$$

Important conclusion:

The formulas in the theory of general relativity are invariant with respect to coordinate transformations if in formulas of special relativity ordinary derivatives $\frac{\partial \mathbf{a}}{\partial \vec{x}^\top}$ are replaced by covariant derivatives $\mathbf{a}_{\parallel \vec{x}^\top}$!

The covariant derivative defined in (2.140) differs from the derivative defined in (2.100) by the summand $\bar{\Gamma}(\mathbf{I}_4 \otimes \mathbf{a})$; there one has instead $(\mathbf{I}_4 \otimes \mathbf{a}^\top) \Gamma$. But it is indeed true that

$$\bar{\Gamma}(\mathbf{I}_4 \otimes \mathbf{a}) = (\mathbf{I}_4 \otimes \mathbf{a}^\top) \Gamma. \quad (2.142)$$

In fact, if we rename the j th row of the sub-matrix Γ_i by γ_j^i , then the matrix Γ^* is composed as

$$\Gamma^* = U_{4 \times 4} \Gamma = \begin{pmatrix} \gamma_0^{0\top} \\ \gamma_0^{1\top} \\ \gamma_0^{2\top} \\ \gamma_0^{3\top} \\ \vdots \\ \hline \gamma_3^{0\top} \\ \gamma_3^{1\top} \\ \gamma_3^{2\top} \\ \gamma_3^{3\top} \end{pmatrix},$$

i.e.

$$\bar{\Gamma} = \begin{pmatrix} \gamma_0^{0\top} & \gamma_3^{0\top} \\ \gamma_0^{1\top} & \gamma_3^{1\top} \\ \gamma_0^{2\top} & \dots & \gamma_3^{2\top} \\ \gamma_0^{3\top} & \gamma_3^{3\top} \end{pmatrix}. \quad (2.143)$$

With $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_i^\top$ one obtains:

$$\begin{aligned} \underline{\underline{\bar{\Gamma}(\mathbf{I}_4 \otimes \mathbf{a})}} &= \begin{pmatrix} \gamma_0^{0\top} \mathbf{a} & \gamma_3^{0\top} \mathbf{a} \\ \gamma_0^{1\top} \mathbf{a} & \gamma_3^{1\top} \mathbf{a} \\ \gamma_0^{2\top} \mathbf{a} & \dots & \gamma_3^{2\top} \mathbf{a} \\ \gamma_0^{3\top} \mathbf{a} & \gamma_3^{3\top} \mathbf{a} \end{pmatrix} = \begin{pmatrix} \mathbf{a}^\top \gamma_0^0 & \mathbf{a}^\top \gamma_3^0 \\ \mathbf{a}^\top \gamma_0^1 & \mathbf{a}^\top \gamma_3^1 \\ \mathbf{a}^\top \gamma_0^2 & \dots & \mathbf{a}^\top \gamma_3^2 \\ \mathbf{a}^\top \gamma_0^3 & \mathbf{a}^\top \gamma_3^3 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{a}^\top \boldsymbol{\Gamma}_0^\top \\ \vdots \\ \mathbf{a}^\top \boldsymbol{\Gamma}_3^\top \end{pmatrix} = \begin{pmatrix} \mathbf{a}^\top \boldsymbol{\Gamma}_0 \\ \vdots \\ \mathbf{a}^\top \boldsymbol{\Gamma}_3 \end{pmatrix} = \underline{\underline{(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma}}}, \end{aligned} \quad (2.144)$$

so we can instead of (2.140) finally write the same as in (2.100)

$$\boxed{\mathbf{a}_{\parallel \vec{x}^\top} = \mathbf{a}_{|\vec{x}^\top} + (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \in \mathbb{R}^{4 \times 4}}. \quad (2.145)$$

2.6 Incidental Remark

If one starts with an equation which is valid in the presence of gravity in general relativity, then this equation must for $v^2 \ll c^2$ pass over to Newton's equation. The force of interaction of two discrete masses m and m_1 is proportional to the product of the two known masses and inversely proportional to the square of the distance of the two centres of gravity:

$$f = G \frac{mm_1}{|\mathbf{x} - \mathbf{x}_1|^2} \frac{\mathbf{x} - \mathbf{x}_1}{|\mathbf{x} - \mathbf{x}_1|},$$

or

$$f = m \cdot G \frac{m_1}{|\mathbf{x} - \mathbf{x}_1|^3} (\mathbf{x} - \mathbf{x}_1).$$

For several discrete masses m_i , this attraction is obtained as

$$f = m \cdot G \sum_i \frac{m_i}{|\mathbf{x} - \mathbf{x}_i|^3} (\mathbf{x} - \mathbf{x}_i),$$

and for a distributed mass with the mass density ρ

$$f = m \cdot G \int_V \rho(\mathbf{x}_i) \frac{\mathbf{x} - \mathbf{x}_i}{|\mathbf{x} - \mathbf{x}_i|^3} dV.$$

In an electric field of strength \mathbf{e} , which together with a charge q produces the force $\mathbf{f} = q\mathbf{e}$, we define the gravitational field strength \mathbf{e}_G , producing the force $\mathbf{f} = m\mathbf{e}_G$ acting on the mass m . For several discrete masses, the gravitational field strength is

$$\mathbf{e}_G = G \sum_i \frac{m_i}{|\mathbf{x} - \mathbf{x}_i|^3} (\mathbf{x} - \mathbf{x}_i).$$

So we can divide the analysis of the problem into two steps. In the first step, the gravitational field generated by several masses m_i at the point \mathbf{x} is determined, and in the second step, the force acting on the mass m at the point \mathbf{x} is determined.

The potential energy is the integral of force times distance, so

$$U = - \int \mathbf{f}^\top d\mathbf{s} = -m \int \mathbf{e}_G^\top d\mathbf{s} \stackrel{\text{def}}{=} m\phi.$$

If a mass m is displaced by a small distance Δx , the work done is equal to the change in potential energy

$$\Delta W = -\Delta U = f_x \Delta x.$$

Dividing this equation by Δx , this force in the x -direction is given as

$$f_x = -\frac{\Delta U}{\Delta x}.$$

Dividing by the mass m , one gets the x -component of the strength of the gravitational field

$$e_x = -\frac{\Delta \phi}{\Delta x}.$$

Whence it follows generally with $\Delta x \rightarrow \mathbf{0}$, that

$$\mathbf{e} = -\nabla \phi,$$

so

$$\mathbf{f} = -m\nabla \phi,$$

or

$$\frac{d^2\mathbf{x}}{dt^2} = -\nabla \phi(\mathbf{x}), \quad (2.146)$$

where the gravitational potential ϕ , which is a scalar function of the position \mathbf{x} , is obtained from the second-order linear partial differential equation, the Poisson equation

$$\Delta \phi(\mathbf{x}) = 4\pi G \rho(\mathbf{x}), \quad (2.147)$$

with the gravitational constant G and the mass density $\rho(\mathbf{x})$. This equation shows the relationship between gravitational potential and matter in the Newtonian Physics. The above two steps are thus:

Step 1. Find the solution $\phi(\mathbf{x})$ of the Poisson equation (2.147).

Step 2. Use the solution of (2.147) to find $\mathbf{x}(t)$.

This is the approach in the classical Newtonian Physics. How must we do or modify the two steps in the theory of general relativity, i.e. how does one generally get the g_{ik} 's and how does one establish the dynamic equations? In the case when only the element g_{00} depends on \mathbf{x} , i.e. only the 00-element of the sub-matrix $\boldsymbol{\Gamma}_i$ is different from zero, then

$$\boldsymbol{\Gamma}_i = \begin{pmatrix} \frac{\partial g_{00}}{\partial x_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

According to (2.146), the acceleration is proportional to the partial derivative of the gravitational potential ϕ with respect to the coordinates x_i . Looking at the equation

$$\ddot{\mathbf{x}} = -(\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}, \quad (2.148)$$

on the left-hand side of the equation one has acceleration and on the right-hand side there is the matrix $\boldsymbol{\Gamma}$ comprising partial derivatives of the g_{ij} 's with respect to the coordinates x_ℓ . The g_{ij} 's apparently play in General Relativity the same role as the gravitational potential ϕ does in classical Physics. There the gravitational potential ϕ was determined by the Poisson equation whose form is, for the most part, given by the Laplace operator Δ :

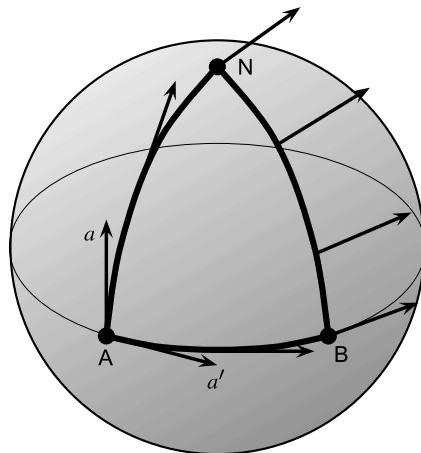
$$\Delta\phi = \frac{\partial^2\phi}{\partial x_1^2} + \frac{\partial^2\phi}{\partial x_2^2} + \frac{\partial^2\phi}{\partial x_3^2} = 4\pi G\rho(\mathbf{x}).$$

So one is now looking for a mathematical expression involving the second derivatives of the g_{ij} 's with respect to the four spacetime coordinates x_i . Such a term appears, in fact, in the differential geometry of Gauss and Riemann, to be precise, in the investigation of the curvature of surfaces or hyper-surfaces in three- or n -dimensional spaces, where the surfaces are described by quadratic forms with the g_{ij} 's as the corresponding matrix elements. Therefore, the appendix “Some Differential Geometry” deals with the theory of curvature of surfaces in three- and n -dimensional spaces.

2.7 Parallel Transport

For further considerations, we need the definition of *parallel transport*:

1. The parallel displacement of a vector \mathbf{a} , which is tangential to the curved surface and runs along a geodesic of this surface, is defined as follows: The origin point of the vector moves along the geodesic and the vector itself moves continuously so that its angle with the geodesic and its length remains constant. It then changes by a parallel transport along $\delta\mathbf{x}$ according to (2.96) with $\delta\mathbf{a} = -(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \delta\mathbf{x}$.

Fig. 2.1 Parallel transport

2. The parallel displacement of a vector on a surface along a broken line, consisting of some geodesic pieces, is done so that from the first corner to the second corner the vector is moved along the first geodesic arc, then along the second arc, etc.
3. Finally, the parallel translation of a vector along a smooth curve is described by the limiting process, in which the curve is approximated by broken lines consisting of geodesic pieces.

If a vector \mathbf{a} in a *flat* space, where $\boldsymbol{\Gamma} = \mathbf{0}$, moves in parallel along a closed loop, it comes back to the starting point with the same length and direction. But if a modified vector comes back, so it must be true that $\boldsymbol{\Gamma} \neq \mathbf{0}$ —there exists curvature.

Example If you move a vector \mathbf{a} on a sphere (Fig. 2.1), beginning at the equator at a point A, along a meridian to the north pole N, then along another meridian back to the equator at a point B, and finally back along the equator to the starting point A, then the incoming vector \mathbf{a}' has a direction other than the initial vector \mathbf{a} . Calling the difference between the starting and end vector

$$\Delta\mathbf{a} \stackrel{\text{def}}{=} \mathbf{a}' - \mathbf{a},$$

the question is: What happens to $\Delta\mathbf{a}$ when the circulated area is smaller and smaller? Of course, $\Delta\mathbf{a}$ approaches the zero vector, but not the ratio $\Delta\mathbf{a}/(\text{circulated area})$.

We now replace the spherical triangle of the example by a differentially small square and consider not the difference vector $\Delta\mathbf{a}$ of a complete circulation. Instead, the initial vector \mathbf{a} is moved the half way around the square in one direction. Then the same vector \mathbf{a} is shifted half way in the other direction. At the meeting point the difference $\Delta\mathbf{a}$ arises. This difference vector is deduced and considered more precisely in the following section.

2.8 Riemannian Curvature Matrix

If one moves a vector $\mathbf{a}(p_0) \in \mathbb{R}^4$ from a point p_0 by $\delta\mathbf{x} \in \mathbb{R}^4$ to the point p_1 , then it changes according to (2.96) by

$$\delta\mathbf{a} = -(\mathbf{I}_4 \otimes \mathbf{a}(p_0)^\top) \boldsymbol{\Gamma}(p_0) \delta\mathbf{x}. \quad (2.149)$$

Therefore,

$$\mathbf{a}(p_1) = \mathbf{a}(p_0) + \delta\mathbf{a} = \mathbf{a}(p_0) - (\mathbf{I}_4 \otimes \mathbf{a}(p_0)^\top) \boldsymbol{\Gamma}(p_0) \delta\mathbf{x}. \quad (2.150)$$

A further displacement of p_1 in the direction $\delta\bar{\mathbf{x}}$ to p_2 gives the change

$$\delta\bar{\mathbf{a}} = -(\mathbf{I}_4 \otimes \mathbf{a}(p_1)^\top) \boldsymbol{\Gamma}(p_1) \delta\bar{\mathbf{x}}. \quad (2.151)$$

For $\boldsymbol{\Gamma}(p_1)$ as a first approximation one can write

$$\boldsymbol{\Gamma}(p_1) \approx \boldsymbol{\Gamma}(p_0) + \sum_{v=0}^3 \frac{\partial \boldsymbol{\Gamma}}{\partial x_v} \delta x_v = \boldsymbol{\Gamma}(p_0) + \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\delta\mathbf{x} \otimes \mathbf{I}_4). \quad (2.152)$$

Equations (2.150) and (2.152) used in (2.151) yield

$$\begin{aligned} \delta\bar{\mathbf{a}} &= -(\mathbf{I}_4 \otimes [\mathbf{a}(p_0) - (\mathbf{I}_4 \otimes \mathbf{a}(p_0)^\top) \boldsymbol{\Gamma}(p_0) \delta\mathbf{x}]^\top) \left(\boldsymbol{\Gamma}(p_0) + \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\delta\mathbf{x} \otimes \mathbf{I}_4) \right) \delta\bar{\mathbf{x}} \\ &= -(\mathbf{I}_4 \otimes \mathbf{a}(p_0)^\top) \boldsymbol{\Gamma}(p_0) \delta\bar{\mathbf{x}} - (\mathbf{I}_4 \otimes \mathbf{a}(p_0)^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\delta\mathbf{x} \otimes \mathbf{I}_4) \delta\bar{\mathbf{x}} \\ &\quad + (\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \mathbf{a}(p_0)^\top) \boldsymbol{\Gamma} \delta\mathbf{x}]^\top) \boldsymbol{\Gamma} \delta\bar{\mathbf{x}} + \mathcal{O}(\delta\bar{\mathbf{x}} \cdot (d\mathbf{x}^2)) \\ &= \left[-(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} - (\mathbf{I}_4 \otimes \mathbf{a}^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\delta\mathbf{x} \otimes \mathbf{I}_4) \right. \\ &\quad \left. + (\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \mathbf{a}(p_0)^\top) \boldsymbol{\Gamma} \delta\mathbf{x}]^\top) \boldsymbol{\Gamma} \right] \delta\bar{\mathbf{x}}. \end{aligned} \quad (2.153)$$

The third term in the brackets can be transformed into the following form:

$$\begin{aligned} (\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \delta\mathbf{x}]^\top) \boldsymbol{\Gamma} &= \overline{\boldsymbol{\Gamma}} (\mathbf{I}_4 \otimes (\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \delta\mathbf{x}) \\ &= \overline{\boldsymbol{\Gamma}} (\mathbf{I}_{16} \otimes \mathbf{a}^\top) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma} \delta\mathbf{x}) \\ &= (\overline{\boldsymbol{\Gamma}} \otimes \mathbf{a}^\top) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) (\mathbf{I}_4 \otimes \delta\mathbf{x}) \\ &= (\mathbf{I}_4 \otimes \mathbf{a}^\top) (\overline{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) (\mathbf{I}_4 \otimes \delta\mathbf{x}). \end{aligned} \quad (2.154)$$

Using this in (2.153) results in

$$\delta\bar{\mathbf{a}} = -(\mathbf{I}_4 \otimes \mathbf{a}^\top) \left[\boldsymbol{\Gamma} + \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\delta\mathbf{x} \otimes \mathbf{I}_4) - (\overline{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) (\mathbf{I}_4 \otimes \delta\mathbf{x}) \right] \delta\bar{\mathbf{x}}. \quad (2.155)$$

If we now first go in the direction $\delta\bar{\mathbf{x}}$ and then in the direction $\delta\mathbf{x}$, we obtain accordingly

$$\delta\bar{\mathbf{a}} = -(\mathbf{I}_4 \otimes \mathbf{a}^\top) \left[\boldsymbol{\Gamma} + \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\delta\bar{\mathbf{x}} \otimes \mathbf{I}_4) - (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma})(\mathbf{I}_4 \otimes \delta\bar{\mathbf{x}}) \right] \delta\mathbf{x}. \quad (2.156)$$

For the final product in the third summand of (2.155) we can also write

$$(\mathbf{I}_4 \otimes \delta\mathbf{x})\delta\bar{\mathbf{x}} = (\mathbf{I}_4 \otimes \delta\mathbf{x})(\delta\bar{\mathbf{x}} \otimes 1) = (\delta\bar{\mathbf{x}} \otimes \delta\mathbf{x}) = \mathbf{U}_{4 \times 4}(\delta\mathbf{x} \otimes \delta\bar{\mathbf{x}}). \quad (2.157)$$

It is true that

$$\Delta\mathbf{a} = (\mathbf{a} + \delta\bar{\mathbf{a}}) - (\mathbf{a} + \delta\bar{\mathbf{a}}) = \delta\bar{\mathbf{a}} - \delta\bar{\mathbf{a}}, \quad (2.158)$$

so with (2.155), (2.156) and (2.157),

$$\begin{aligned} \Delta\mathbf{a} &= (\mathbf{I}_4 \otimes \mathbf{a}^\top) \left(\boldsymbol{\Gamma}(\delta\mathbf{x} - \delta\bar{\mathbf{x}}) \right. \\ &\quad \left. + \left[\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} + (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) \right] (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16})(\delta\mathbf{x} \otimes \delta\bar{\mathbf{x}}) \right). \end{aligned} \quad (2.159)$$

With the 16×16 Riemannian curvature matrix

$$\boxed{\mathbf{R} \stackrel{\text{def}}{=} \left[\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} + (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) \right] (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}) \in \mathbb{R}^{16 \times 16},} \quad (2.160)$$

we have

$$\boxed{\Delta\mathbf{a} = (\mathbf{I}_4 \otimes \mathbf{a}^\top) [\boldsymbol{\Gamma}(\delta\mathbf{x} - \delta\bar{\mathbf{x}}) + \mathbf{R}(\delta\mathbf{x} \otimes \delta\bar{\mathbf{x}})] \in \mathbb{R}^4.} \quad (2.161)$$

We also define a slightly modified curvature matrix:

$$\boxed{\check{\mathbf{R}} \stackrel{\text{def}}{=} (\mathbf{G} \otimes \mathbf{I}_4)\mathbf{R}.} \quad (2.162)$$

2.9 Properties of the Riemannian Curvature Matrix

2.9.1 Composition of R and \check{R}

Which form do the elements of the Riemannian curvature matrix

$$\mathbf{R} = \left[\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} + (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) \right] (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}) \in \mathbb{R}^{16 \times 16}$$

have? The property, already mentioned earlier, that in the occurring matrix difference $\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}$ the first, (4 + 2)th, (8 + 3)th and the 16-th row/column are equal

to the zero row/column, has for the Riemannian curvature matrix the consequence that its corresponding columns are zero columns!

Next we obtain with (2.143)

$$\bar{\Gamma} = \begin{pmatrix} \boldsymbol{\gamma}_0^{0\top} & \boldsymbol{\gamma}_3^{0\top} \\ \boldsymbol{\gamma}_0^{1\top} & \boldsymbol{\gamma}_3^{1\top} \\ \boldsymbol{\gamma}_0^{2\top} & \dots & \boldsymbol{\gamma}_3^{2\top} \\ \boldsymbol{\gamma}_0^{3\top} & & \boldsymbol{\gamma}_3^{3\top} \end{pmatrix} = [\bar{\Gamma}_0, \dots, \bar{\Gamma}_3]$$

for

$$\begin{aligned} (\bar{\Gamma} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) &= [(\bar{\Gamma}_0 \otimes \mathbf{I}_4), \dots, (\bar{\Gamma}_3 \otimes \mathbf{I}_4)] \begin{pmatrix} \boldsymbol{\Gamma} & 0 & 0 & 0 \\ 0 & \boldsymbol{\Gamma} & 0 & 0 \\ 0 & 0 & \boldsymbol{\Gamma} & 0 \\ 0 & 0 & 0 & \boldsymbol{\Gamma} \end{pmatrix} \\ &= [(\bar{\Gamma}_0 \otimes \mathbf{I}_4)\boldsymbol{\Gamma}, \dots, (\bar{\Gamma}_3 \otimes \mathbf{I}_4)\boldsymbol{\Gamma}]. \end{aligned}$$

This matrix product contributes to the matrix element $R_{\alpha\beta}^{\gamma\delta}$ the sum

$$[(\boldsymbol{\gamma}_{\delta}^{\gamma\top} \otimes \mathbf{I}_4)\boldsymbol{\Gamma}]_{\alpha\beta} = [\Gamma_{\delta 0}^{\gamma}\boldsymbol{\Gamma}_0 + \dots + \Gamma_{\delta 3}^{\gamma}\boldsymbol{\Gamma}_3]_{\alpha\beta} = \sum_v \underline{\underline{\Gamma}}_{\delta v}^{\gamma} \Gamma_{\alpha\beta}^v, \quad (2.163)$$

where $\boldsymbol{\gamma}_{\delta}^{\gamma\top} \in \mathbb{R}^4$ is the γ th row of the sub-matrix $\bar{\Gamma}_{\delta}$, i.e. the δ th row of the sub-matrix $\boldsymbol{\Gamma}_{\gamma}$.

Furthermore,

$$\begin{aligned} &(\bar{\Gamma} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma})\mathbf{U}_{4 \times 4} \\ &= (\bar{\Gamma} \otimes \mathbf{I}_4) \begin{pmatrix} \boldsymbol{\Gamma} & 0 & 0 & 0 \\ 0 & \boldsymbol{\Gamma} & 0 & 0 \\ 0 & 0 & \boldsymbol{\Gamma} & 0 \\ 0 & 0 & 0 & \boldsymbol{\Gamma} \end{pmatrix} \mathbf{U}_{4 \times 4} \\ &= (\bar{\Gamma} \otimes \mathbf{I}_4) \begin{pmatrix} \boldsymbol{\gamma}_0 & | & \boldsymbol{\gamma}_1 & | & \boldsymbol{\gamma}_2 & | & \boldsymbol{\gamma}_3 \\ \ddots & | & \ddots & | & \ddots & | & \ddots \\ \boldsymbol{\gamma}_0 & | & \boldsymbol{\gamma}_1 & | & \boldsymbol{\gamma}_2 & | & \boldsymbol{\gamma}_3 \end{pmatrix}. \end{aligned}$$

This matrix product ($\boldsymbol{\gamma}_{\delta} \in \mathbb{R}^{16}$ is the δ th column of $\boldsymbol{\Gamma}$) contributes to the matrix element $R_{\alpha\beta}^{\gamma\delta}$ the sum

$$\left[(\boldsymbol{\gamma}^{\gamma\top} \otimes \mathbf{I}_4) \begin{pmatrix} \boldsymbol{\gamma}_{\delta} & & \\ & \ddots & \\ & & \boldsymbol{\gamma}_{\delta} \end{pmatrix} \right]_{\alpha\beta}$$

$$\begin{aligned}
&= \left[[\boldsymbol{\gamma}_0^{\gamma\top} \otimes \mathbf{I}_4, \dots, \boldsymbol{\gamma}_0^{\gamma\top} \otimes \mathbf{I}_4] \begin{pmatrix} \boldsymbol{\gamma}_\delta & & \\ & \ddots & \\ & & \boldsymbol{\gamma}_\delta \end{pmatrix}_{\alpha\beta} \right] \\
&= [(\boldsymbol{\gamma}_0^{\gamma\top} \otimes \mathbf{I}_4) \boldsymbol{\gamma}_\delta, \dots, (\boldsymbol{\gamma}_3^{\gamma\top} \otimes \mathbf{I}_4) \boldsymbol{\gamma}_\delta]_{\alpha\beta} = [(\boldsymbol{\gamma}_\beta^{\gamma\top} \otimes \mathbf{I}_4) \boldsymbol{\gamma}_\delta]_\alpha \\
&= [\underline{\Gamma_{\beta 0}^\gamma \boldsymbol{\gamma}_\delta^0 + \dots + \Gamma_{\beta 3}^\gamma \boldsymbol{\gamma}_\delta^3}]_\alpha = \sum_v \underline{\Gamma_{\beta v}^\gamma \Gamma_{\delta\alpha}^v}. \tag{2.164}
\end{aligned}$$

In accordance with (2.160), one gets finally with (2.163) and (2.164)

$$R_{\alpha\beta}^{\gamma\delta} = \frac{\partial}{\partial x_\beta} \Gamma_{\alpha\delta}^\gamma - \frac{\partial}{\partial x_\delta} \Gamma_{\alpha\beta}^\gamma + \sum_v \Gamma_{\beta v}^\gamma \Gamma_{\delta\alpha}^v - \sum_v \Gamma_{\delta v}^\gamma \Gamma_{\alpha\beta}^v. \tag{2.165}$$

From this form, one can immediately read off the property

$$\underline{\underline{R_{\alpha\beta}^{\gamma\delta}} = -R_{\alpha\delta}^{\gamma\beta}}. \tag{2.166}$$

With the help of (2.165) one can also verify the so-called *cyclic identity*:

$$\underline{\underline{R_{\alpha\beta}^{\gamma\delta} + R_{\beta\delta}^{\gamma\alpha} + R_{\delta\alpha}^{\gamma\beta} = 0}}. \tag{2.167}$$

From (2.165), using (2.63), (2.73) and (2.65), one can also derive a closed form for $\check{R}_{\alpha\beta}^{\gamma\delta}$ in (2.160) as follows:

$$\begin{aligned}
\check{R}_{\alpha\beta}^{\gamma\delta} &= \sum_i g_{\gamma i} R_{\alpha\beta}^{i\delta} = \sum_i g_{\gamma i} \left(\frac{\partial}{\partial x_\beta} \Gamma_{\alpha\delta}^i - \frac{\partial}{\partial x_\delta} \Gamma_{\alpha\beta}^i + \sum_v \Gamma_{\beta v}^i \Gamma_{\delta\alpha}^v - \sum_v \Gamma_{\delta v}^i \Gamma_{\alpha\beta}^v \right) \\
&= \left(\frac{\partial}{\partial x_\beta} \check{R}_{\alpha\delta}^\gamma - \sum_i \Gamma_{\alpha\delta}^i \frac{\partial g_{\gamma i}}{\partial x_\beta} \right) - \left(\frac{\partial}{\partial x_\delta} \check{R}_{\alpha\beta}^\gamma - \sum_i \Gamma_{\alpha\beta}^i \frac{\partial g_{\gamma i}}{\partial x_\delta} \right) \\
&\quad + \sum_v \check{R}_{\beta v}^\gamma \Gamma_{\delta\alpha}^v - \sum_v \check{R}_{\delta v}^\gamma \Gamma_{\alpha\beta}^v \\
&= \left(\frac{\partial}{\partial x_\beta} \check{R}_{\alpha\delta}^\gamma - \sum_i \Gamma_{\alpha\delta}^i (\check{R}_{\gamma\beta}^i + \check{R}_{i\beta}^\gamma) \right) - \left(\frac{\partial}{\partial x_\delta} \check{R}_{\alpha\beta}^\gamma - \sum_i \Gamma_{\alpha\beta}^i (\check{R}_{\gamma\delta}^i + \check{R}_{i\delta}^\gamma) \right) \\
&\quad + \sum_v \check{R}_{\beta v}^\gamma \Gamma_{\delta\alpha}^v - \sum_v \check{R}_{\delta v}^\gamma \Gamma_{\alpha\beta}^v \\
&= \frac{1}{2} \left(\frac{\partial}{\partial x_\beta} \left(\frac{\partial g_{\delta\gamma}}{\partial x_\alpha} + \frac{\partial g_{\alpha\gamma}}{\partial x_\delta} - \frac{\partial g_{\alpha\delta}}{\partial x_\gamma} \right) - \frac{\partial}{\partial x_\delta} \left(\frac{\partial g_{\beta\gamma}}{\partial x_\alpha} + \frac{\partial g_{\alpha\gamma}}{\partial x_\beta} - \frac{\partial g_{\alpha\beta}}{\partial x_\gamma} \right) \right)
\end{aligned}$$

$$\begin{aligned}
& - \sum_i \Gamma_{\alpha\delta}^i \check{\Gamma}_{\gamma\beta}^i - \sum_i \Gamma_{\alpha\delta}^i \check{\Gamma}_{i\beta}^\gamma + \sum_i \Gamma_{\alpha\beta}^i \check{\Gamma}_{\gamma\delta}^i + \sum_i \Gamma_{\alpha\beta}^i \check{\Gamma}_{i\delta}^\gamma \\
& + \sum_\nu \check{\Gamma}_{\beta\nu}^\gamma \Gamma_{\delta\alpha}^\nu - \sum_\nu \check{\Gamma}_{\delta\nu}^\gamma \Gamma_{\alpha\beta}^\nu.
\end{aligned}$$

After cancelling out some of the terms, we finally get the closed form

$$\boxed{\check{R}_{\alpha\beta}^{\gamma\delta} = \frac{\partial}{\partial x_\beta} \check{\Gamma}_{\alpha\delta}^\gamma - \frac{\partial}{\partial x_\delta} \check{\Gamma}_{\alpha\beta}^\gamma + \sum_i \Gamma_{\alpha\beta}^i \check{\Gamma}_{\gamma\delta}^i - \sum_i \Gamma_{\alpha\delta}^i \check{\Gamma}_{\gamma\beta}^i} \quad (2.168)$$

$$\begin{aligned}
& = \frac{1}{2} \left(\frac{\partial^2 g_{\delta\gamma}}{\partial x_\alpha \partial x_\beta} - \frac{\partial^2 g_{\alpha\delta}}{\partial x_\gamma \partial x_\beta} - \frac{\partial^2 g_{\beta\gamma}}{\partial x_\alpha \partial x_\delta} + \frac{\partial^2 g_{\alpha\beta}}{\partial x_\gamma \partial x_\delta} \right) \\
& + \sum_i \Gamma_{\alpha\beta}^i \check{\Gamma}_{\gamma\delta}^i - \sum_i \Gamma_{\alpha\delta}^i \check{\Gamma}_{\gamma\beta}^i,
\end{aligned} \quad (2.169)$$

or

$$\boxed{\check{R}_{\alpha\beta}^{\gamma\delta} = \frac{1}{2} \left(\frac{\partial^2 g_{\delta\gamma}}{\partial x_\alpha \partial x_\beta} - \frac{\partial^2 g_{\alpha\delta}}{\partial x_\gamma \partial x_\beta} - \frac{\partial^2 g_{\beta\gamma}}{\partial x_\alpha \partial x_\delta} + \frac{\partial^2 g_{\alpha\beta}}{\partial x_\gamma \partial x_\delta} \right) + \sum_i \check{\Gamma}_{\gamma\delta}^i \sum_\nu g_{i\nu}^{(-1)} \check{\Gamma}_{\alpha\beta}^\nu - \sum_i \check{\Gamma}_{\gamma\beta}^i \sum_\nu g_{i\nu}^{(-1)} \check{\Gamma}_{\alpha\delta}^\nu.} \quad (2.170)$$

From (2.170) the following identities are obtained directly by comparing the corresponding forms:

$$\underline{\check{R}_{\alpha\beta}^{\gamma\delta}} = -\underline{\check{R}_{\gamma\beta}^{\alpha\delta}}, \quad (2.171)$$

$$\underline{\check{R}_{\alpha\beta}^{\gamma\delta}} = -\underline{\check{R}_{\alpha\delta}^{\gamma\beta}}, \quad (2.172)$$

$$\underline{\check{R}_{\alpha\beta}^{\gamma\delta}} = \underline{\check{R}_{\gamma\delta}^{\alpha\beta}}. \quad (2.173)$$

Also in this case, a *cyclic identity* is valid:

$$\boxed{\check{R}_{\alpha\beta}^{\gamma\delta} + \check{R}_{\beta\delta}^{\gamma\alpha} + \check{R}_{\delta\alpha}^{\gamma\beta} = 0.} \quad (2.174)$$

If the two vectors $d\mathbf{x}$ and $d\bar{\mathbf{x}}$ are perpendicular to each other, then the area of the formed rectangle is equal to $|d\mathbf{x}| \cdot |d\bar{\mathbf{x}}|$. In the *differential geometry*, the limit of the ratio of Δa to the area is now called the *curvature* κ :

$$\kappa \stackrel{\text{def}}{=} \lim_{|d\mathbf{x}|, |d\bar{\mathbf{x}}| \rightarrow 0} \frac{|\Delta a(d\mathbf{x}, d\bar{\mathbf{x}})|}{|d\mathbf{x}| \cdot |d\bar{\mathbf{x}}|}$$

or with $\mathbf{n} \stackrel{\text{def}}{=} \frac{dx}{|dx|}$ and $\bar{\mathbf{n}} \stackrel{\text{def}}{=} \frac{d\vec{x}}{|d\vec{x}|}$,

$$\kappa \stackrel{\text{def}}{=} \lim_{\epsilon \rightarrow 0} \frac{|\Delta \mathbf{a}(\epsilon \mathbf{n}, \epsilon \bar{\mathbf{n}})|}{\epsilon^2}.$$

With the help of (2.161) we then obtain

$$\kappa = |(\mathbf{I}_4 \otimes \mathbf{a}^\top) \mathbf{R}(\mathbf{n} \otimes \bar{\mathbf{n}})|. \quad (2.175)$$

Riemannian Coordinate System For the study of the properties of the Riemannian curvature matrix, it is advantageous to first perform a coordinate transformation, so that in the new coordinate system, the Christoffel matrices have the property $\boldsymbol{\Gamma} = \mathbf{0}$. Such a coordinate transformation is in the case of a curved spacetime, where \mathbf{G} depends on \vec{x} , only locally possible, but then for each \vec{x} ! By the inverse transformation of the obtained statements, they are again globally valid. So, we look for a local coordinate transformation such that in the new coordinate system $\boldsymbol{\Gamma} = \mathbf{0}$ is valid. For geodesic lines the following applies to the four coordinates:

$$\frac{d^2 x_k}{ds^2} + \left(\frac{d\vec{x}}{ds} \right)^\top \boldsymbol{\Gamma}_k \frac{d\vec{x}}{ds} = 0. \quad (2.176)$$

The x_k 's belong to an arbitrary coordinate system in which the geodesic is described by $x_k = x_k(s)$ and s is the arc length along the curve. At a fixed point \mathcal{P} with the coordinate $\vec{x}^{(0)}$, any coordinate can be developed in a power series:

$$x_k = x_k^{(0)} + \zeta_k s + \frac{1}{2} \left(\frac{d^2 x_k}{ds^2} \right)_{\mathcal{P}} s^2 + \frac{1}{3!} \left(\frac{d^3 x_k}{ds^3} \right)_{\mathcal{P}} s^3 + \dots, \quad (2.177)$$

where ζ_k is the k th component of the tangent vector

$$\zeta \stackrel{\text{def}}{=} \left(\frac{d\vec{x}}{ds} \right)_{\mathcal{P}}$$

to the geodesic at the point \mathcal{P} . Then, however, according to (2.176),

$$\frac{d^2 x_k}{ds^2} = -\zeta^\top (\boldsymbol{\Gamma}_k)_{\mathcal{P}} \zeta. \quad (2.178)$$

Plugging this into (2.177), considering a small neighbourhood of \mathcal{P} , a small $x_k - x_k^{(0)}$, and neglecting powers higher than two gives

$$x_k = x_k^{(0)} + \zeta_k s - \frac{1}{2} \zeta^\top (\boldsymbol{\Gamma}_k)_{\mathcal{P}} \zeta s^2. \quad (2.179)$$

Now, calling $\zeta s = \vec{x}'$, the following is obtained from (2.179):

$$x_k = x_k^{(0)} + x'_k - \frac{1}{2} \vec{x}'^\top (\boldsymbol{\Gamma}_k)_{\mathcal{P}} \vec{x}'.$$

This relation suggests the following coordinate transformation from \vec{x} to \vec{x}' :

$$x'_k = x_k - x_k^{(0)} + \frac{1}{2}(\vec{x} - \vec{x}^{(0)})^\top (\boldsymbol{\Gamma}_k)_{\mathcal{P}}(\vec{x} - \vec{x}^{(0)}). \quad (2.180)$$

What is the corresponding metric matrix \mathbf{G}' in

$$ds^2 = \vec{x}'^\top \mathbf{G}' \vec{x}'? \quad (2.181)$$

In this coordinate system, the geodesic has the equation

$$\frac{d^2 \vec{x}'}{ds^2} + \left(\mathbf{I}_4 \otimes \left(\frac{d \vec{x}'}{ds} \right)^\top \right) \boldsymbol{\Gamma}'(\vec{x}') \left(\frac{d \vec{x}'}{ds} \right) = \mathbf{0}. \quad (2.182)$$

But since in the new coordinate system the geodesics are straight lines of the form $\vec{x}' = \zeta s$, in (2.182) the expression $(\mathbf{I}_4 \otimes \zeta^\top) \boldsymbol{\Gamma}'(\zeta s) \zeta$ must be equal to the zero vector. Since ζ are arbitrary vectors, the following must be valid for $s = 0$:

$$\underline{\underline{\boldsymbol{\Gamma}'(0) = \mathbf{0}}}. \quad (2.183)$$

Implications for the Riemannian Curvature Matrix If at a point \mathcal{P} for a special coordinate system one has $\boldsymbol{\Gamma}_{\mathcal{P}} = \mathbf{0}$, then naturally any $\check{\Gamma}_{\alpha\beta}^{\gamma}$ defined in (2.63) is equal to zero. But since

$$\underline{\underline{\check{\Gamma}_{k\ell}^i + \check{\Gamma}_{ki}^\ell = \frac{1}{2} \left(\frac{\partial g_{\ell i}}{\partial x_k} + \frac{\partial g_{ki}}{\partial x_\ell} - \frac{\partial g_{k\ell}}{\partial x_i} \right) + \frac{1}{2} \left(\frac{\partial g_{i\ell}}{\partial x_k} + \frac{\partial g_{k\ell}}{\partial x_i} - \frac{\partial g_{ki}}{\partial x_\ell} \right) = \frac{\partial g_{i\ell}}{\partial x_k}}}. \quad (2.184)$$

also all first partial derivatives of the elements of the metric matrix are zero, i.e.

$$\underline{\underline{\frac{\partial \mathbf{G}}{\partial \vec{x}} \Big|_{\mathcal{P}} = \mathbf{0}}}. \quad (2.185)$$

In the local coordinate system with $\boldsymbol{\Gamma}_{\mathcal{P}} = \mathbf{0}$, the Riemannian curvature matrix has the form

$$\mathbf{R}_{\mathcal{P}} = \frac{\partial \boldsymbol{\Gamma}}{\partial \vec{x}^\top} \Big|_{\mathcal{P}} (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}). \quad (2.186)$$

The structure is

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}^{00} & \mathbf{R}^{01} & \mathbf{R}^{02} & \mathbf{R}^{03} \\ \mathbf{R}^{10} & \mathbf{R}^{11} & \mathbf{R}^{12} & \mathbf{R}^{13} \\ \mathbf{R}^{20} & \mathbf{R}^{21} & \mathbf{R}^{22} & \mathbf{R}^{23} \\ \mathbf{R}^{30} & \mathbf{R}^{31} & \mathbf{R}^{32} & \mathbf{R}^{33} \end{pmatrix}. \quad (2.187)$$

Each sub-matrix $\mathbf{R}^{\gamma\delta}$ has the form

$$\mathbf{R}^{\gamma\delta} = \begin{pmatrix} R_{00}^{\gamma\delta} & R_{01}^{\gamma\delta} & R_{02}^{\gamma\delta} & R_{03}^{\gamma\delta} \\ R_{10}^{\gamma\delta} & R_{11}^{\gamma\delta} & R_{12}^{\gamma\delta} & R_{13}^{\gamma\delta} \\ R_{20}^{\gamma\delta} & R_{21}^{\gamma\delta} & R_{22}^{\gamma\delta} & R_{23}^{\gamma\delta} \\ R_{30}^{\gamma\delta} & R_{31}^{\gamma\delta} & R_{32}^{\gamma\delta} & R_{33}^{\gamma\delta} \end{pmatrix}. \quad (2.188)$$

$\mathbf{R}^{\gamma\delta}$ is thus the sub-matrix of \mathbf{R} at the intersection of the γ th row and δ th column. Furthermore, $R_{\alpha\beta}^{\gamma\delta}$ is the matrix element of the sub-matrix $\mathbf{R}^{\gamma\delta}$ at the intersection of the α th row and β th column.

With (2.68) and (2.185) one has

$$\begin{aligned} \frac{\partial \boldsymbol{\Gamma}}{\partial \vec{x}^\top} \Big|_{\mathcal{P}} &= \frac{1}{2} \frac{\partial}{\partial \vec{x}^\top} \left((\mathbf{G}^{-1} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{g}_0^\top}{\partial \vec{x}} \\ \vdots \\ \frac{\partial \mathbf{g}_3^\top}{\partial \vec{x}} \end{pmatrix} + \begin{pmatrix} \frac{\partial \mathbf{g}_0}{\partial \vec{x}^\top} \\ \vdots \\ \frac{\partial \mathbf{g}_3}{\partial \vec{x}^\top} \end{pmatrix} - \frac{\partial \mathbf{G}}{\partial \vec{x}} \right] \right) \\ &= \frac{1}{2} \frac{\partial}{\partial \vec{x}^\top} (\mathbf{G}^{-1} \otimes \mathbf{I}_4) [\mathbf{0} + \mathbf{0} - \mathbf{0}] \\ &\quad + \frac{1}{2} (\mathbf{G}^{-1} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial^2 \mathbf{g}_0^\top}{\partial \vec{x}^\top \partial \vec{x}} \\ \vdots \\ \frac{\partial^2 \mathbf{g}_3^\top}{\partial \vec{x}^\top \partial \vec{x}} \end{pmatrix} + \begin{pmatrix} \frac{\partial^2 \mathbf{g}_0}{\partial \vec{x}^\top \partial \vec{x}^\top} \\ \vdots \\ \frac{\partial^2 \mathbf{g}_3}{\partial \vec{x}^\top \partial \vec{x}^\top} \end{pmatrix} - \frac{\partial^2 \mathbf{G}}{\partial \vec{x}^\top \partial \vec{x}} \right]. \end{aligned} \quad (2.189)$$

For

$$\check{\mathbf{R}} = (\mathbf{G} \otimes \mathbf{I}_4) \mathbf{R} \quad (2.190)$$

with

$$\check{\boldsymbol{\Gamma}} \stackrel{\text{def}}{=} (\mathbf{G} \otimes \mathbf{I}_4) \boldsymbol{\Gamma} \quad (2.191)$$

one finally obtains

$$\check{\mathbf{R}}_{\mathcal{P}} \stackrel{\text{def}}{=} \frac{\partial \check{\boldsymbol{\Gamma}}}{\partial \vec{x}^\top} (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}). \quad (2.192)$$

Naturally, because of the in \mathbf{R} occurring matrix difference $\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}$, the first, $(4+2)$ th, $(8+3)$ th and the 16th column of $\check{\mathbf{R}}$ and of $\check{\mathbf{R}}_{\mathcal{P}}$ are equal to the zero column. In

$$\check{\boldsymbol{\Gamma}} = \begin{pmatrix} \check{\boldsymbol{\Gamma}}_0 \\ \check{\boldsymbol{\Gamma}}_1 \\ \check{\boldsymbol{\Gamma}}_2 \\ \check{\boldsymbol{\Gamma}}_3 \end{pmatrix},$$

$\check{\Gamma}_v$ has the form

$$\check{\Gamma}_v = \frac{1}{2} \left(\frac{\partial \mathbf{g}_v^\top}{\partial \mathbf{x}} + \frac{\partial \mathbf{g}_v}{\partial \mathbf{x}^\top} - \frac{\partial \mathbf{G}}{\partial x_v} \right),$$

i.e. the elements of $\check{\Gamma}_v$ are (see also (2.63))

$$\check{\Gamma}_{\alpha\beta}^v = \frac{1}{2} \left(\frac{\partial g_{\beta v}}{\partial x_\alpha} + \frac{\partial g_{\alpha v}}{\partial x_\beta} - \frac{\partial g_{\alpha\beta}}{\partial x_v} \right). \quad (2.193)$$

For $(\check{R}_{\alpha\beta}^{\gamma\delta})_{\mathcal{P}}$ from (2.165) one obtains

$$\underline{\underline{(\check{R}_{\alpha\beta}^{\gamma\delta})_{\mathcal{P}}} = \frac{1}{2} \left(\frac{\partial^2 g_{\gamma\delta}}{\partial x_\alpha \partial x_\beta} - \frac{\partial^2 g_{\alpha\delta}}{\partial x_\gamma \partial x_\beta} - \frac{\partial^2 g_{\beta\gamma}}{\partial x_\alpha \partial x_\delta} + \frac{\partial^2 g_{\alpha\beta}}{\partial x_\gamma \partial x_\delta} \right)}.$$
(2.194)

From (2.194), by comparing the corresponding forms, follows

$$\underline{\underline{(\check{R}_{\alpha\beta}^{\gamma\delta})_{\mathcal{P}}} = (\check{R}_{\alpha\beta}^{\delta\gamma})_{\mathcal{P}}}, \quad (2.195)$$

and the so-called *cyclic identity*

$$\underline{\underline{(\check{R}_{\alpha\beta}^{\gamma\delta})_{\mathcal{P}} + (\check{R}_{\alpha\gamma}^{\delta\beta})_{\mathcal{P}} + (\check{R}_{\alpha\delta}^{\beta\gamma})_{\mathcal{P}} = 0}}. \quad (2.196)$$

From (2.165), at a point \mathcal{P} , we get

$$\underline{\underline{(R_{\alpha\beta}^{\gamma\delta})_{\mathcal{P}} = \left(\frac{\partial}{\partial x_\beta} \Gamma_{\alpha\delta}^\gamma - \frac{\partial}{\partial x_\delta} \Gamma_{\alpha\beta}^\gamma \right)_\mathcal{P}}}.$$
(2.197)

Partially differentiating (2.197) yields

$$\underline{\underline{\left(\frac{\partial}{\partial x_\kappa} R_{\alpha\beta}^{\gamma\delta} \right)_\mathcal{P} = \left(\frac{\partial^2}{\partial x_\kappa \partial x_\beta} \Gamma_{\alpha\delta}^\gamma - \frac{\partial^2}{\partial x_\kappa \partial x_\delta} \Gamma_{\alpha\beta}^\gamma \right)_\mathcal{P}}}.$$
(2.198)

With the help of (2.198), by substituting the corresponding terms, the so-called *Bianchi identity* is obtained:

$$\underline{\underline{\frac{\partial}{\partial x_\kappa} R_{\alpha\beta}^{\gamma\delta} + \frac{\partial}{\partial x_\beta} R_{\alpha\delta}^{\gamma\kappa} + \frac{\partial}{\partial x_\delta} R_{\alpha\kappa}^{\gamma\beta} = 0}}. \quad (2.199)$$

Since this is valid in *any* event \mathcal{P} , it is valid *anywhere*.

2.10 The Ricci Matrix and Its Properties

The reason for considering the Riemannian curvature theory is to find with its help a way to determine the components of the Christoffel matrix Γ , which are needed

to calculate the solution of the equations

$$\ddot{\vec{x}} = -(\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \boldsymbol{\Gamma} \dot{\vec{x}},$$

describing the dynamic behaviour of particles in a gravitational field. But the Riemannian curvature matrix has, being a 16×16 -matrix, 256 components, so, it would provide 256 equations in extreme cases. Considering, however, that there are 4 rows and 4 columns of \mathbf{R} with all zeros, only $12 \cdot 12 = 144$ equations remain—still too many. The number of equations can be reduced significantly, however, by reducing, using clever addition of matrix elements, the number of components of the newly formed matrix. Setting up in this way a *symmetric* 4×4 -matrix, we would get exactly 10 independent equations for the determination of the 10 independent components of \mathbf{G} . Such a matrix is called the Ricci matrix, which can be obtained in two ways. One way is through the sum of sub-matrices on the diagonal of \mathbf{R} ; it is described in the Appendix. The second way is as follows:

The Ricci matrix \mathbf{R}_{Ric} consists of the traces of sub-matrices $\mathbf{R}^{\gamma\delta}$ of \mathbf{R}

$$R_{\text{Ric},\gamma\delta} \stackrel{\text{def}}{=} \text{trace}(\mathbf{R}^{\gamma\delta}) = \sum_{v=0}^3 R_{vv}^{\gamma\delta}. \quad (2.200)$$

Accordingly, the components of the new matrix $\check{\mathbf{R}}_{\text{Ric}}$ are defined as

$$\check{R}_{\text{Ric},\gamma\delta} \stackrel{\text{def}}{=} \sum_{v=0}^3 \check{R}_{vv}^{\gamma\delta}. \quad (2.201)$$

From (2.194) one may read immediately that the Ricci matrix $\check{\mathbf{R}}_{\text{Ric}}$ is *symmetric* because $\check{R}_{vv}^{\gamma\delta} = \check{R}_{vv}^{\delta\gamma}$.

Moreover, because of (2.162),

$$\mathbf{R} = (\mathbf{G}^{-1} \otimes \mathbf{I}_4) \check{\mathbf{R}}, \quad (2.202)$$

so

$$\mathbf{R}^{\gamma\delta} = (\mathbf{g}_\gamma^{-T} \otimes \mathbf{I}_4) \check{\mathbf{R}}^\delta = \sum_{\mu=0}^3 g_{\gamma\mu}^{[-1]} \check{R}^{\mu\delta}, \quad (2.203)$$

where \mathbf{g}_γ^{-T} is the γ th row of \mathbf{G}^{-1} and $\check{\mathbf{R}}^\delta$ the 16×4 -matrix consisting of the four sub-matrices in the δ th block column of $\check{\mathbf{R}}$, i.e. for the matrix elements one has

$$R_{\alpha\beta}^{\gamma\delta} = \sum_{\mu=0}^3 g_{\gamma\mu}^{[-1]} \check{R}_{\alpha\beta}^{\mu\delta}. \quad (2.204)$$

With the help of (2.200), for the Ricci matrix components one gets

$$R_{\text{Ric},\gamma\delta} = \sum_v R_{vv}^{\gamma\delta} = \sum_v \sum_\mu g_{\gamma\mu}^{[-1]} \check{R}_{vv}^{\mu\delta}, \quad (2.205)$$

or, due to (2.173),

$$R_{\text{Ric},\gamma\delta} = \sum_{\nu} \sum_{\mu} g_{\gamma\mu}^{[-1]} \check{R}_{\mu\delta}^{\nu\nu}. \quad (2.206)$$

The *curvature scalar* R is obtained by taking the trace of the Ricci matrix:

$$\begin{aligned} R &\stackrel{\text{def}}{=} \text{trace}(\mathbf{R}_{\text{Ric}}) = \sum_{\alpha} \sum_{\nu} R_{\nu\nu}^{\alpha\alpha} \\ &= \sum_{\alpha} \sum_{\nu} \sum_{\mu} g_{\alpha\mu}^{[-1]} \check{R}_{\nu\nu}^{\mu\alpha} = \sum_{\alpha} \sum_{\mu} g_{\alpha\mu}^{[-1]} \check{R}_{\text{Ric},\mu\alpha}. \end{aligned} \quad (2.207)$$

Conversely, one obtains

$$\check{R}_{\alpha\beta}^{\gamma\delta} = \sum_{\mu=0}^3 g_{\gamma\mu} R_{\alpha\beta}^{\mu\delta}. \quad (2.208)$$

Because of (2.165) it follows directly that

$$R_{\text{Ric},\gamma\delta} = \sum_{\nu=0}^3 \left(\frac{\partial}{\partial x_{\delta}} \Gamma_{\gamma\nu}^{\nu} - \frac{\partial}{\partial x_{\nu}} \Gamma_{\gamma\delta}^{\nu} + \sum_{\mu=0}^3 \Gamma_{\delta\mu}^{\nu} \Gamma_{\nu\gamma}^{\mu} - \sum_{\mu=0}^3 \Gamma_{\nu\mu}^{\nu} \Gamma_{\gamma\delta}^{\mu} \right) \quad (2.209)$$

and from (2.168)

$$\check{R}_{\text{Ric},\gamma\delta} = \sum_{\nu=0}^3 \left(\frac{\partial}{\partial x_{\delta}} \check{\Gamma}_{\gamma\nu}^{\nu} - \frac{\partial}{\partial x_{\nu}} \check{\Gamma}_{\gamma\delta}^{\nu} + \sum_{\mu=0}^3 \Gamma_{\gamma\delta}^{\mu} \check{\Gamma}_{\nu\nu}^{\mu} - \sum_{\mu=0}^3 \Gamma_{\gamma\nu}^{\mu} \check{\Gamma}_{\nu\delta}^{\mu} \right). \quad (2.210)$$

2.10.1 Symmetry of the Ricci Matrix \mathbf{R}_{Ric}

Even if the Riemannian matrix \mathbf{R} itself is not symmetric, the derived Ricci matrix \mathbf{R}_{Ric} is nevertheless symmetric; and that will be shown below. The symmetry will follow from the component equations (2.209) of the Ricci matrix. One sees immediately that the second and fourth summand are symmetric in γ and δ . Looking at $\sum_{\nu=0}^3 \frac{\partial}{\partial x_{\delta}} \Gamma_{\gamma\nu}^{\nu}$, it is not directly seen that this term is symmetric in γ and δ . This can be seen using the Laplace expansion theorem for determinants.¹ Developing the determinant of \mathbf{G} along the ν th row, one gets

$$g \stackrel{\text{def}}{=} \det(\mathbf{G}) = g_{\nu 1} A_{\nu 1} + \cdots + g_{\nu \delta} A_{\nu \delta} + \cdots + g_{\nu n} A_{\nu n},$$

¹The sum of the products of all elements of a row (or column) with its adjuncts is equal to the determinant's value: $\det A = \sum_{j=1}^n (-1)^{i+j} \cdot a_{ij} \cdot \det A_{ij}$ (development along the i th row)

where $A_{\nu\delta}$ is the element in the ν th row and δ th column of the adjoint of \mathbf{G} . If $g_{\delta\nu}^{[-1]}$ is the $(\delta\nu)$ -element of the inverse matrix \mathbf{G}^{-1} , then $g_{\delta\nu}^{[-1]} = \frac{1}{g} A_{\nu\delta}$, so $A_{\nu\delta} = g g_{\delta\nu}^{[-1]}$. Thus we obtain

$$\frac{\partial g}{\partial g_{\nu\delta}} = A_{\nu\delta} = g g_{\delta\nu}^{[-1]},$$

or

$$\delta g = g g_{\delta\nu}^{[-1]} \delta g_{\nu\delta},$$

and

$$\frac{\partial g}{\partial x_\gamma} = g g_{\delta\nu}^{[-1]} \frac{\partial g_{\nu\delta}}{\partial x_\gamma},$$

i.e.

$$\frac{1}{g} \frac{\partial g}{\partial x_\gamma} = g_{\delta\nu}^{[-1]} \frac{\partial g_{\nu\delta}}{\partial x_\gamma}. \quad (2.211)$$

Due to (2.62), on the other hand, one has

$$\sum_{\nu=0}^3 \Gamma_{\gamma\nu}^\nu = \sum_{\nu=0}^3 \sum_{\delta=0}^3 \frac{g_{\delta\nu}^{[-1]}}{2} \left(\frac{\partial g_{\nu\delta}}{\partial x_\gamma} + \frac{\partial g_{\gamma\delta}}{\partial x_\nu} - \frac{\partial g_{\gamma\nu}}{\partial x_\delta} \right),$$

i.e. the last two summands cancel out, and the following remains:

$$\sum_{\nu=0}^3 \Gamma_{\gamma\nu}^\nu = \sum_{\nu=0}^3 \sum_{\delta=0}^3 \frac{g_{\delta\nu}^{[-1]}}{2} \frac{\partial g_{\nu\delta}}{\partial x_\gamma}.$$

Then it follows by (2.211) that

$$\sum_{\nu=0}^3 \underline{\frac{\partial}{\partial x_\delta} \Gamma_{\gamma\nu}^\nu} = \sum_{\nu=0}^3 \sum_{\delta=0}^3 \frac{1}{\sqrt{|g|}} \frac{\partial^2 \sqrt{|g|}}{\partial x_\gamma \partial x_\delta}. \quad (2.212)$$

From this form the symmetry in γ and δ can immediately be observed.

Now remains to show that the third term

$$\sum_{\nu=0}^3 \sum_{\mu=0}^3 \Gamma_{\delta\mu}^\nu \Gamma_{\nu\gamma}^\mu$$

in (2.209) is symmetric. But this can be spotted from the following equalities:

$$\sum_{\nu,\mu=0}^3 \Gamma_{\delta\mu}^\nu \Gamma_{\nu\gamma}^\mu = \sum_{\nu,\mu=0}^3 \Gamma_{\mu\delta}^\nu \Gamma_{\gamma\nu}^\mu = \sum_{\nu,\mu=0}^3 \Gamma_{\nu\delta}^\mu \Gamma_{\gamma\mu}^\nu.$$

Combining everything together demonstrates that the Ricci matrix \mathbf{R}_{Ric} is, in fact, symmetric!

2.10.2 The Divergence of the Ricci Matrix

Multiplying the Bianchi identities (2.199) in the form

$$\frac{\partial}{\partial x_\kappa} R_{\alpha\beta}^{\nu\delta} + \frac{\partial}{\partial x_\beta} R_{\alpha\delta}^{\nu\kappa} + \frac{\partial}{\partial x_\delta} R_{\alpha\kappa}^{\nu\beta} = 0$$

with $g_{\gamma\nu}$ and summing over ν , we obtain at a point \mathcal{P} , where $\frac{\partial G}{\partial x} = \mathbf{0}$,

$$\frac{\partial}{\partial x_\kappa} \sum_{\nu=0}^3 g_{\gamma\nu} R_{\alpha\beta}^{\nu\delta} + \frac{\partial}{\partial x_\beta} \sum_{\nu=0}^3 g_{\gamma\nu} R_{\alpha\delta}^{\nu\kappa} + \frac{\partial}{\partial x_\delta} \sum_{\nu=0}^3 g_{\gamma\nu} R_{\alpha\kappa}^{\nu\beta} = 0.$$

With (2.208) we obtain the following modified Bianchi identity:

$$\underline{\underline{\frac{\partial}{\partial x_\kappa} \check{R}_{\alpha\beta}^{\gamma\delta} + \frac{\partial}{\partial x_\beta} \check{R}_{\alpha\delta}^{\gamma\kappa} + \frac{\partial}{\partial x_\delta} \check{R}_{\alpha\kappa}^{\gamma\beta} = 0.}}$$

For the third term we can, with respect to (2.172), also write

$$-\frac{\partial}{\partial x_\delta} \check{R}_{\alpha\beta}^{\gamma\kappa}.$$

Substituting now $\alpha = \beta$ and summing over α , we obtain

$$\frac{\partial}{\partial x_\kappa} \check{R}_{\text{Ric},\gamma\delta} + \sum_{\alpha=0}^3 \frac{\partial}{\partial x_\alpha} \check{R}_{\alpha\delta}^{\gamma\kappa} - \frac{\partial}{\partial x_\delta} \check{R}_{\text{Ric},\gamma\kappa} = 0. \quad (2.214)$$

In the second term, one can, based on (2.171), replace $\check{R}_{\alpha\delta}^{\gamma\kappa}$ by $-\check{R}_{\gamma\delta}^{\alpha\kappa}$. If we set $\gamma = \delta$ and sum over γ , we obtain from (2.214) with the “curvature scalar” $\check{R} \stackrel{\text{def}}{=} \sum_{\gamma=0}^3 \check{R}_{\text{Ric},\gamma\gamma}$, the trace of the Ricci matrix \check{R}_{Ric} :

$$\frac{\partial}{\partial x_\kappa} \check{R} - \sum_{\alpha=0}^3 \frac{\partial}{\partial x_\alpha} \check{R}_{\text{Ric},\alpha\kappa} - \sum_{\gamma=0}^3 \frac{\partial}{\partial x_\gamma} \check{R}_{\text{Ric},\gamma\kappa} = 0. \quad (2.215)$$

If in the last sum the summation index γ is replaced by α , we can finally summarize

$$\underline{\underline{\frac{\partial}{\partial x_\kappa} \check{R} - 2 \sum_{\alpha=0}^3 \frac{\partial}{\partial x_\alpha} \check{R}_{\text{Ric},\alpha\kappa} = 0.}}$$

We would get the same result, when this equation was assumed:

$$\frac{\partial}{\partial x_\kappa} \check{R}_{\gamma\delta}^{\alpha\beta} - 2 \frac{\partial}{\partial x_\beta} \check{R}_{\gamma\delta}^{\alpha\kappa} = 0. \quad (2.217)$$

Indeed, setting $\delta = \gamma$ and summing over γ , we receive first:

$$\frac{\partial}{\partial x_\kappa} \check{R}_{\text{Ric},\alpha\beta} - 2 \frac{\partial}{\partial x_\beta} \check{R}_{\text{Ric},\alpha\kappa} = 0.$$

Substituting now $\alpha = \beta$ and summing over α , we receive again (2.216).

A different result is obtained when starting at first from (2.217) (with v instead of α), multiplying this equation by $g_{\alpha v}^{[-1]}$,

$$\frac{\partial}{\partial x_\kappa} g_{\alpha v}^{[-1]} \check{R}_{\gamma\delta}^{\nu\beta} - 2 \frac{\partial}{\partial x_\beta} g_{\alpha v}^{[-1]} \check{R}_{\gamma\delta}^{\nu\kappa} = 0,$$

then setting $\gamma = \delta$, summing over γ and v , and using (2.214):

$$\begin{aligned} & \sum_\gamma \sum_v \frac{\partial}{\partial x_\kappa} g_{\alpha v}^{[-1]} \check{R}_{\gamma\delta}^{\nu\beta} - 2 \sum_\gamma \sum_v \frac{\partial}{\partial x_\beta} g_{\alpha v}^{[-1]} \check{R}_{\gamma\delta}^{\nu\kappa} \\ &= \frac{\partial}{\partial x_\kappa} R_{\text{Ric},\alpha\beta} - 2 \frac{\partial}{\partial x_\beta} R_{\text{Ric},\alpha\kappa} = 0. \end{aligned}$$

If we now set $\alpha = \beta$ and sum over α , we finally obtain the important relationship

$$\underline{\underline{\frac{\partial}{\partial x_\kappa} R - 2 \sum_\alpha \frac{\partial}{\partial x_\alpha} R_{\text{Ric},\alpha\kappa} = 0.}} \quad (2.218)$$

These are the four equations for the four spacetime coordinates x_0, \dots, x_3 . Finally, the overall result can be represented as

$$\vec{\nabla}^\top \left(\mathbf{R}_{\text{Ric}} - \frac{1}{2} R \mathbf{I}_4 \right) = \mathbf{0}^\top. \quad (2.219)$$

2.11 General Theory of Gravitation

2.11.1 The Einstein's Matrix \mathfrak{G}

With the Einstein's matrix

$$\mathfrak{G} \stackrel{\text{def}}{=} \mathbf{R}_{\text{Ric}} - \frac{1}{2} R \mathbf{I}_4, \quad (2.220)$$

taking into account that the matrix \mathfrak{G} is symmetric, (2.219) can be restated as

$$\vec{\nabla}^\top \mathfrak{G} = \mathbf{0}^\top. \quad (2.221)$$

The following is a very important property of the Einstein's matrix:

The divergence of the Einstein's matrix \mathfrak{G} vanishes!

2.11.2 Newton's Theory of Gravity

According to Newton, for the acceleration the following is valid:

$$\frac{d^2\mathbf{x}}{dt^2} = -\nabla\phi(\mathbf{x}), \quad (2.222)$$

where $\phi(\mathbf{x})$ is the gravitational potential and $\mathbf{x} \in \mathbb{R}^3$. One can also write

$$\frac{d^2\mathbf{x}}{dt^2} + \nabla\phi(\mathbf{x}) = \mathbf{0}. \quad (2.223)$$

The Newtonian universal time is a parameter which has two degrees of freedom, namely the time origin t_0 and a , the unit of time, both can be chosen arbitrarily: $t = t_0 + a\tau$. Thus we obtain

$$\frac{d^2t}{d\tau^2} = 0, \quad \frac{d^2\mathbf{x}}{d\tau^2} + \frac{\partial\phi}{\partial\mathbf{x}} \left(\frac{dt}{d\tau} \right)^2 = \mathbf{0}. \quad (2.224)$$

This can also be written with the spacetime vector $\vec{\mathbf{x}} = \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix} \in \mathbb{R}^4$:

$$\frac{d^2\vec{\mathbf{x}}}{d\tau^2} + (\mathbf{I}_4 \otimes \dot{\vec{\mathbf{x}}}^\top) \boldsymbol{\Gamma} \dot{\vec{\mathbf{x}}} = \mathbf{0}. \quad (2.225)$$

Here $\boldsymbol{\Gamma} \in \mathbb{R}^{16 \times 4}$ has the form

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{\partial\phi}{\partial x_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{\partial\phi}{\partial x_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{\partial\phi}{\partial x_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now, how can one get the expression of the Poisson equation

$$\Delta\phi(\mathbf{x}) = \frac{\partial^2\phi}{\partial x_1^2} + \frac{\partial^2\phi}{\partial x_2^2} + \frac{\partial^2\phi}{\partial x_3^2} = 4\pi G\rho(\mathbf{x})? \quad (2.226)$$

For the curvature matrix R one needs the matrix

and the matrix

$$\boldsymbol{\Gamma}^* = \mathbf{U}_{4 \times 4} \boldsymbol{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \frac{\partial \phi}{\partial x_1} & 0 & 0 & 0 \\ \frac{\partial \phi}{\partial x_2} & 0 & 0 & 0 \\ \frac{\partial \phi}{\partial x_3} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

in order to determine the following matrix

However, for the product $(\bar{\Gamma} \otimes I_4)(I_4 \otimes \Gamma)$ we get the zero matrix, so that the curvature matrix is only composed as

$$R = \frac{\partial \Gamma}{\partial x^\top} (U_{4 \times 4} - I_{16}).$$

It is true that

$$\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} U_{4 \times 4} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \frac{\partial^2 \phi}{\partial x_1^2} & \frac{\partial^2 \phi}{\partial x_1 \partial x_2} & \frac{\partial^2 \phi}{\partial x_1 \partial x_3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \frac{\partial^2 \phi}{\partial x_2 \partial x_1} & \frac{\partial^2 \phi}{\partial x_2^2} & \frac{\partial^2 \phi}{\partial x_2 \partial x_3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \frac{\partial^2 \phi}{\partial x_3 \partial x_1} & \frac{\partial^2 \phi}{\partial x_3 \partial x_2} & \frac{\partial^2 \phi}{\partial x_3^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

finally resulting in

The 4×4 -sub-matrices on the main diagonal of the curvature matrix \mathbf{R} contain exactly the components of the left-hand side of the Poisson equation. This subsequently provides the motivation for the introduction of the Ricci matrix. Here it is specifically equal to

$$\mathbf{R}_{\text{Ric}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{\partial^2 \phi}{\partial x_1^2} & -\frac{\partial^2 \phi}{\partial x_1 \partial x_2} & -\frac{\partial^2 \phi}{\partial x_1 \partial x_3} \\ 0 & -\frac{\partial^2 \phi}{\partial x_2 \partial x_1} & -\frac{\partial^2 \phi}{\partial x_2^2} & -\frac{\partial^2 \phi}{\partial x_2 \partial x_3} \\ 0 & -\frac{\partial^2 \phi}{\partial x_3 \partial x_1} & -\frac{\partial^2 \phi}{\partial x_3 \partial x_2} & -\frac{\partial^2 \phi}{\partial x_3^2} \end{pmatrix}. \quad (2.228)$$

If we form the trace of the Ricci matrix, then

$$R = \text{trace}(\mathbf{R}_{\text{Ric}}) = -\left(\frac{\partial^2 \phi}{\partial x_1^2} + \frac{\partial^2 \phi}{\partial x_2^2} + \frac{\partial^2 \phi}{\partial x_3^2}\right), \quad (2.229)$$

so that finally the Poisson equation can also be succinctly written as

$$-R = 4\pi G\rho. \quad (2.230)$$

The relationship in Newtonian mechanics between the gravitational potential ϕ and matter is the Poisson equation

$$\Delta\phi = 4\pi G\rho,$$

where ρ is the mass density and G the Newtonian Gravitational Constant. In the theory of General Relativity, it is now required to set up general invariant equations of gravitation between the g_{ik} 's and matter. It is useful to characterise the matter by the energy–momentum matrix $\mathbf{T} = \mathbf{T}_{\text{total}}$.

For the rotating system in Sect. 2.5, we had $\ddot{r} = r(\omega + \dot{\phi})^2$ so that there the potential was $\phi = -\frac{\omega^2 r^2}{2}$ and $g_{00} = 1 - \frac{r^2 \omega^2}{c^2} = 1 + \frac{2\phi}{c^2}$. On the other hand, in Newtonian non-relativistic limit $\frac{v^2}{c^2} \leq 1$, we obtained for the energy–momentum matrix (see Sect. 1.9.2) $T_{00} = c^2 \rho_0$ and for the remaining $T_{ij} \approx 0$.

With $\phi = \frac{c^2}{2}(g_{00} - 1)$ we obtain $\Delta\phi = \frac{c^2}{2} \Delta g_{00}$. So one can write for the above Poisson equation

$$\Delta g_{00} = \frac{8\pi G}{c^4} T_{00}. \quad (2.231)$$

2.11.3 The Einstein's Equation with \mathfrak{G}

If one assumes that in the general case, i.e. in the presence of gravitational fields on the right-hand side of (2.231), the symmetric energy–momentum matrix is \mathbf{T} , a matrix containing the second partial derivatives of the elements of the metric matrix \mathbf{G}

must appear on the left-hand side of the equation. If therefore the Einstein's matrix \mathfrak{G} is taken, one obtains the Einstein's Field Equation

$$\mathfrak{G} = -\frac{8\pi G}{c^4} \mathbf{T}. \quad (2.232)$$

Because the matrix $\mathbf{T} \in \mathbb{R}^{4 \times 4}$ is symmetric, also the Einstein's matrix

$$\mathfrak{G} = \mathbf{R}_{\text{Ric}} - \frac{1}{2} R \mathbf{I}_4 \quad (2.233)$$

must be symmetric. This is indeed the case because both the Ricci matrix \mathbf{R}_{Ric} and the diagonal matrix $R \mathbf{I}_4$ are symmetric. \mathbf{R}_{Ric} is extracted, according to (2.200), from the Riemannian curvature matrix

$$\mathbf{R} = \left(\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} + (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) \right) (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}). \quad (2.234)$$

Final Form of the Einstein's Equation The energy–momentum matrix \mathbf{T} on the right-hand side of the Einstein's field equation (2.232) has the property that $\mathbf{T} \vec{\nabla} = \mathbf{0}$, if one considers a closed system, i.e. if there are no external forces acting. The same must apply on the left: $\mathfrak{G} \vec{\nabla} = \mathbf{0}$. But that is, due to (2.221), true for the symmetric matrix \mathfrak{G} !

Overall, the final Einstein's Field Equation is fixed as an axiom as

$$\mathfrak{G} = \mathbf{R}_{\text{Ric}} - \frac{R}{2} \mathbf{I}_4 = -\frac{8\pi G}{c^4} \mathbf{T}. \quad (2.235)$$

This is a matrix differential equation for determining the metric matrix \mathbf{G} . This is no longer an “action-at-a-distance-equation” as Newton's, but describes the relationships at a spacetime point \bar{x} ! \mathbf{R}_{Ric} depends linearly on the second order derivatives of g_{ik} , and depends nonlinearly on the g_{ik} directly. So the Einstein's equation is a coupled system of second-order nonlinear partial differential equations for the determination of the components g_{ik} of the metric matrix \mathbf{G} as a function of the given distribution of matter, i.e. the matrix \mathbf{T} , the source of the gravitational field.

By taking the trace, from (2.235) it follows that

$$R - \frac{R}{2} \cdot 4 = -\frac{8\pi G}{c^4} T,$$

so

$$R = \frac{8\pi G}{c^4} T, \quad (2.236)$$

where

$$T \stackrel{\text{def}}{=} \text{trace}(\mathbf{T}).$$

Substituting (2.236) into (2.235), one obtains the following form of the Einstein's Field Equation:

$$\boxed{\mathbf{R}_{\text{Ric}} = \frac{8\pi G}{c^4} \left(\frac{T}{2} \mathbf{I}_4 - \mathbf{T} \right)}. \quad (2.237)$$

By the way, the constant factor $\frac{8\pi G}{c^4}$ has the numerical value

$$\underline{\underline{\frac{8\pi G}{c^4} = 1.86 \cdot 10^{-27} \text{ cm/g}}} \quad (2.238)$$

2.12 Summary

2.12.1 Covariance Principle

Einstein postulated the *Equivalence Principle*:

Gravitational forces are equivalent to inertial forces.

Gravitational fields can be eliminated by transitioning to an accelerated coordinate system. In this new local inertial frame, the laws of the theory of special relativity are valid.

Directly from the equivalence principle follows the *Covariance Principle*:

The equations of physics must be invariant under general coordinate transformations.

In particular, this means that even in a local inertial frame the equations must be valid, so the transition from a metric matrix \mathbf{G} to a Minkowski matrix \mathbf{M} results in the laws of special relativity.

The connection is $\mathbf{G} = \mathbf{J}^\top \mathbf{M} \mathbf{J}$ and $\mathbf{M} = \mathbf{J}^{-1\top} \mathbf{G} \mathbf{J}^{-1}$, i.e. using the special transformation matrix $\mathbf{T}(\vec{x}) = \mathbf{J}^{-1}(\vec{x})$ leads for a particular event \vec{x} to a local inertial frame in which the laws of special relativity apply. Conversely, one passes from the special local inertial frame with the event $\vec{\xi}$ to the same event in the general coordinate system \vec{x} by the transformation $\vec{x} = \mathbf{J}^{-1} \vec{\xi}$.

The physical equations in the theory of general relativity must be formulated so that they are invariant (covariant) with respect to general coordinate transformations.

Above we derived: The formulas in the theory of general relativity are invariant under coordinate transformations if one replaces in the formulas of special relativity the ordinary derivatives $\frac{\partial a}{\partial x^\tau}$ by the covariant derivatives $a_{|\alpha\tau}$. Then we're done! The law applies generally.

For a particle on which no force acts, for example, in an inertial system with $\dot{\xi} \stackrel{\text{def}}{=} \frac{d\xi}{d\tau}$, applies

$$\frac{d\dot{\xi}}{d\tau} = \mathbf{0}. \quad (2.239)$$

Herein the ordinary differential $d\xi$ must be replaced by the covariant differential $D\vec{u}$, with $\vec{u} = \dot{\vec{x}} \stackrel{\text{def}}{=} J^{-1}\dot{\xi}$. According to (2.145), we first obtain

$$Du = \vec{u}_{||\vec{x}^\top} d\vec{x} = \frac{\partial \vec{u}}{\partial \vec{x}^\top} d\vec{x} + (\mathbf{I}_4 \otimes \vec{u}^\top) \Gamma d\vec{x}, \quad (2.240)$$

so

$$\frac{D\vec{u}}{d\tau} = \vec{u}_{||\vec{x}^\top} \frac{d\vec{x}}{d\tau} = \frac{d\vec{u}}{d\tau} + (\mathbf{I}_4 \otimes \vec{u}^\top) \Gamma \frac{d\vec{x}}{d\tau}.$$

This used in (2.239) instead of $\frac{d\vec{u}}{d\tau}$ generally yields

$$\frac{D\vec{u}}{d\tau} = \mathbf{0}, \quad (2.241)$$

or

$$\frac{d\vec{u}}{d\tau} = -(\mathbf{I}_4 \otimes \vec{u}^\top) \Gamma \vec{u}. \quad (2.242)$$

Through the Christoffel matrix Γ , the effect of the gravitational field is expressed. If no gravitational field is present, then $\Gamma = \mathbf{0}$ and (2.239) is again obtained. A comparison of (2.242) with the equation of a geodesic shows that a material particle moves on a geodesic, so to speak, “on the shortest path in curved space”.

If, in addition to gravitational forces, other forces appear, e.g. exerted by electric fields, then for an inertial frame one has the equation

$$m \frac{d\dot{\xi}}{d\tau} = \vec{f}. \quad (2.243)$$

This equation multiplied from the left with J^{-1} and again used with the covariant derivative leads with

$$\vec{f}_x \stackrel{\text{def}}{=} J^{-1} \vec{f}$$

to

$$m \frac{D\vec{u}}{d\tau} = \vec{f}_x, \quad (2.244)$$

or

$$m \frac{d\vec{u}}{d\tau} = \vec{f}_x - m(\mathbf{I}_4 \otimes \vec{u}^\top) \Gamma \vec{u}, \quad (2.245)$$

where on the right-hand side, next to the other forces \vec{f}_x , the gravitational forces are written.

2.12.2 Einstein's Field Equation and Momentum

The Einstein's field equation expresses that each form of matter and energy is a source of the gravitational field. The gravitational field is described by the metric matrix G , the components of which must be determined by using Einstein's Field Equation

$$\mathbf{R}_{\text{Ric}} - \frac{R}{2} \mathbf{I}_4 = -\frac{8\pi G}{c^4} \mathbf{T}. \quad (2.246)$$

So one basically has to perform the same procedure as in Newton's dynamics:

1. Solve the Poisson equation $\Delta\phi(\mathbf{x}) = 4\pi G\rho(\mathbf{x})$, to determine the potential function ϕ .
2. Establish the solution of the equation $\frac{d^2\mathbf{x}}{dt^2} = -\nabla\phi(\mathbf{x})$.

In the theory of general relativity, we now arrive at the following procedure:

1. Solve Einstein's Field Equation (2.246) $\mathbf{R}_{\text{Ric}} - \frac{R}{2} \mathbf{I}_4 = -\frac{8\pi G}{c^4} \mathbf{T}$, to determine the metric matrix G .
2. Establish the solution of (2.245)

$$m \frac{d\vec{u}}{d\tau} = \vec{f}_x - m(\mathbf{I}_4 \otimes \vec{u}^\top) \boldsymbol{\Gamma} \vec{u}.$$

2.13 Hilbert's Action Functional

Above, the Einstein's equation was postulated as an axiom. Einstein has found this equation through years of hard work. As a supplement, we will now show that one can obtain the Einstein's equation with the help of the calculus of variations. Following Hilbert, this is derived from a variational principle, but initially only for the free gravitational field, i.e. when $\mathbf{T} = \mathbf{0}$.

The gravitational field is determined primarily by the metric matrix G , i.e. the curvature of space caused by this field. All curvature parameters are concentrated in the curvature scalar R which is formed through the trace of the Ricci matrix \mathbf{R}_{Ric} . Now a variational functional, the Hilbert's action functional, is formed so that the space curvature is minimal:

$$W_{\text{Grav}} = \int R dV. \quad (2.247)$$

However, this integral is not invariant under coordinate transformations Θ . Therefore, we need the following considerations: Let

$$\Theta^\top(\vec{x}) \mathbf{G}(\vec{x}) \Theta(\vec{x}) = \mathbf{M}, \quad (2.248)$$

i.e. $\Theta(\vec{x})$ is the matrix which at the point \vec{x} transforms the metric matrix \mathbf{G} into the Minkowski matrix \mathbf{M} . Computing the determinants on both sides of (2.248), we obtain

$$\underbrace{\det(\mathbf{G})}_g \underbrace{\det(\Theta)^2}_{\Theta^2} = \det(\mathbf{M}) = -1, \quad (2.249)$$

i.e.

$$\sqrt{-g} = \frac{1}{\Theta}.$$

In a Cartesian coordinate system, the integral of a scalar with respect to the scalar $dV = dx_0 \cdot dx_1 \cdot dx_2 \cdot dx_3$ is also a scalar. When passing to curvilinear coordinates \vec{x}' , the integration element dV is

$$\frac{1}{\Theta} dV' = \sqrt{-g'} dV'.$$

In curvilinear coordinates, $\sqrt{-g} dV$ behaves as an invariant when integrated over any area of the four-dimensional space. If f is a scalar, then $f \sqrt{-g}$, which is invariant when integrated over dV , is named the *scalar density*. This quantity provides a scalar for its multiplication with the four-dimensional volume element dV .

For this reason, we now consider only the effect

$$W_{\text{Grav}} = \int R \left(\boldsymbol{\Gamma}(\boldsymbol{x}), \frac{\partial \boldsymbol{\Gamma}}{\partial \boldsymbol{x}^\top} \right) \sqrt{-g} d^4 \vec{x}. \quad (2.250)$$

Under $d^4 \vec{x}$ the four-dimensional volume element $dx_0 \cdot dx_1 \cdot dx_2 \cdot dx_3$ is understood. The Einstein's equation will follow from (2.250) and the condition $\delta W_{\text{Grav}} = 0$ for any variation δg_{ik} . $R \sqrt{-g}$ is a Lagrange density which is integrated over the volume. The elements of $\Gamma_{\alpha\beta}^k$ of the Christoffel matrix $\boldsymbol{\Gamma}$ are (2.62)

$$\Gamma_{\alpha\beta}^k = \sum_{i=0}^3 \frac{g_{ki}^{[-1]}}{2} \left(\frac{\partial g_{\beta i}}{\partial x_\alpha} + \frac{\partial g_{\alpha i}}{\partial x_\beta} - \frac{\partial g_{\alpha\beta}}{\partial x_i} \right) \quad (2.251)$$

and the elements of the Riemannian curvature matrix \mathbf{R} are, according to (2.165),

$$R_{\alpha\beta}^{\gamma\delta} = \frac{\partial}{\partial x_\beta} \Gamma_{\alpha\delta}^\gamma - \frac{\partial}{\partial x_\delta} \Gamma_{\alpha\beta}^\gamma + \sum_v \Gamma_{\beta v}^\gamma \Gamma_{\delta\alpha}^v - \sum_v \Gamma_{\delta v}^\gamma \Gamma_{\alpha\beta}^v. \quad (2.252)$$

The curvature scalar R , due to (2.207), is obtained from the Ricci matrix by

$$\begin{aligned} R = \text{trace}(\mathbf{R}_{\text{Ric}}) &= \sum_{\alpha} \sum_{\nu} R_{\nu\nu}^{\alpha\alpha} \\ &= \sum_{\alpha} \sum_{\nu} \sum_{\mu} g_{\alpha\mu}^{[-1]} \check{R}_{\nu\nu}^{\mu\alpha} = \sum_{\alpha} \sum_{\mu} g_{\alpha\mu}^{[-1]} \check{R}_{\text{Ric},\mu\alpha}, \end{aligned} \quad (2.253)$$

which, in accordance with (2.210), is

$$\check{R}_{\text{Ric},\gamma\delta} = \sum_{\nu=0}^3 \left(\frac{\partial}{\partial x_{\delta}} \check{\Gamma}_{\gamma\nu}^{\nu} - \frac{\partial}{\partial x_{\nu}} \check{\Gamma}_{\gamma\delta}^{\nu} + \sum_{\mu=0}^3 \Gamma_{\gamma\delta}^{\mu} \check{\Gamma}_{\nu\nu}^{\mu} - \sum_{\mu=0}^3 \Gamma_{\gamma\nu}^{\mu} \check{\Gamma}_{\nu\delta}^{\mu} \right). \quad (2.254)$$

The Lagrange/Hamilton theory applied to the action integral (2.250) provides the Euler/Lagrange equations for the variational problem associated to (2.250). We consider the elements of $g_{ki}^{[-1]}$ and $\Gamma_{\alpha\beta}^k$ as independent functions $f_i(\vec{x})$. The integral therefore is a functional of the form

$$\int L \left(f_i(\vec{x}), \frac{\partial f_i}{\partial x_k}(\vec{x}) \right) d^4 \vec{x}. \quad (2.255)$$

This yields the Euler/Lagrange equations

$$\frac{\partial L}{\partial f_i} = \sum_{k=0}^3 \frac{\partial}{\partial x_k} \frac{\partial L}{\partial (\frac{\partial f_i}{\partial x_k})}. \quad (2.256)$$

The Euler/Lagrange equations for (2.250) are

$$\frac{\partial}{\partial g_{\alpha\beta}^{[-1]}} \left(\sqrt{-g} \sum_{\delta} \sum_{\mu} g_{\mu\delta}^{[-1]} \check{R}_{\text{Ric},\mu\delta} \right) = 0, \quad (2.257)$$

$$\frac{\partial}{\partial \Gamma_{\alpha\beta}^{\gamma}} \left(\sqrt{-g} \sum_{\delta} \sum_{\mu} g_{\mu\delta}^{[-1]} \check{R}_{\text{Ric},\mu\delta} \right) = \sum_{\delta=0}^3 \frac{\partial}{\partial x_{\delta}} \left(\sqrt{-g} \sum_{\delta} \sum_{\mu} g_{\mu\delta}^{[-1]} \frac{\partial \check{R}_{\text{Ric},\mu\delta}}{\partial (\frac{\partial \Gamma_{\alpha\beta}^{\gamma}}{\partial x_{\delta}})} \right). \quad (2.258)$$

The first equation (2.257) provides the Einstein's field equation. Indeed, one can first write this compactly:

$$\frac{\partial L}{\partial \mathbf{G}^{-1}} = \frac{\partial (\sqrt{-g} R)}{\partial \mathbf{G}^{-1}} = \frac{\partial \sqrt{-g}}{\partial \mathbf{G}^{-1}} R + \sqrt{-g} \frac{\partial R}{\partial \mathbf{G}^{-1}} = 0. \quad (2.259)$$

For $\frac{\partial \sqrt{-g}}{\partial \mathbf{G}^{-1}}$, one then obtains

$$\frac{\partial \sqrt{-g}}{\partial \mathbf{G}^{-1}} = \frac{-1}{2\sqrt{-g}} \cdot \frac{\partial g}{\partial \mathbf{G}^{-1}}. \quad (2.260)$$

In addition,

$$\frac{\partial}{\partial \mathbf{G}^{-1}} \left(\frac{1}{g} \cdot g \right) = \mathbf{0} = \frac{1}{g} \frac{\partial g}{\partial \mathbf{G}^{-1}} + g \frac{\partial(1/g)}{\partial \mathbf{G}^{-1}}. \quad (2.261)$$

According to the Laplace expansion theorem for determinants (“The sum of the products of all elements of a row with their cofactors is equal to the value of the determinant”), we obtain by developing the determinant of \mathbf{G}^{-1} along the γ th row:

$$\det(\mathbf{G}^{-1}) = \frac{1}{g} = g_{\gamma 0}^{[-1]} A_{\gamma 0}^{[-1]} + \cdots + g_{\gamma \beta}^{[-1]} A_{\gamma \beta}^{[-1]} + \cdots + g_{\gamma 3}^{[-1]} A_{\gamma 3}^{[-1]}, \quad (2.262)$$

where $A_{\gamma \beta}^{[-1]}$ is the element in the γ th row and β th column of the adjoint of \mathbf{G}^{-1} ; $g_{\gamma \beta}^{[-1]}$ is the $(\gamma \beta)$ -element of \mathbf{G}^{-1} . It is, of course, true that

$$\mathbf{G} = \frac{\text{adj}(\mathbf{G}^{-1})}{\det(\mathbf{G}^{-1})}$$

so element-wise $g_{\beta \gamma} = g \cdot A_{\gamma \beta}^{[-1]}$, or $A_{\gamma \beta}^{[-1]} = \frac{1}{g} g_{\beta \gamma}$. Thus, we obtain by partial differentiation of (2.262) with respect to $g_{\gamma \beta}^{[-1]}$

$$\frac{\partial(1/g)}{\partial g_{\gamma \beta}^{[-1]}} = A_{\gamma \beta}^{[-1]} = \frac{1}{g} g_{\beta \gamma},$$

so

$$\frac{\partial(1/g)}{\partial \mathbf{G}^{-1}} = \frac{1}{g} \mathbf{G}. \quad (2.263)$$

Using this in (2.261) provides

$$\frac{\partial g}{\partial \mathbf{G}^{-1}} = -g \mathbf{G}, \quad (2.264)$$

and, together with (2.260), finally yields

$$\underline{\underline{\frac{\partial \sqrt{-g}}{\partial \mathbf{G}^{-1}}} = \frac{-1}{2} \sqrt{-g} \cdot \mathbf{G}}. \quad (2.265)$$

All what is missing in (2.259) is $\frac{\partial R}{\partial \mathbf{G}^{-1}}$. Using $R = \sum_{\alpha} \sum_{\mu} g_{\alpha \mu}^{[-1]} \check{R}_{\text{Ric}, \mu \alpha}$,

$$\underline{\underline{\frac{\partial R}{\partial \mathbf{G}^{-1}}} = \begin{pmatrix} \frac{\partial R}{\partial g_{00}^{[-1]}} & \cdots & \frac{\partial R}{\partial g_{03}^{[-1]}} \\ \vdots & \ddots & \vdots \\ \frac{\partial R}{\partial g_{30}^{[-1]}} & \cdots & \frac{\partial R}{\partial g_{33}^{[-1]}} \end{pmatrix} = \begin{pmatrix} \check{R}_{\text{Ric}, 00} & \cdots & \check{R}_{\text{Ric}, 03} \\ \vdots & \ddots & \vdots \\ \check{R}_{\text{Ric}, 30} & \cdots & \check{R}_{\text{Ric}, 33} \end{pmatrix} = \underline{\underline{\check{R}_{\text{Ric}}}}. \quad (2.266)}$$

Multiplying this matrix with \mathbf{G}^{-1} , by the way, yields

$$\mathbf{R}_{\text{Ric}} = \mathbf{G}^{-1} \check{\mathbf{R}}_{\text{Ric}}. \quad (2.267)$$

Inserting (2.265) and (2.266) into (2.259) yields first

$$\sqrt{-g} \left(\frac{-1}{2} \cdot \mathbf{G} R + \check{\mathbf{R}}_{\text{Ric}} \right) = \mathbf{0}, \quad (2.268)$$

i.e. the following particular form of the Einstein's field equation

$$\check{\mathbf{R}}_{\text{Ric}} - \frac{R}{2} \cdot \mathbf{G} = \mathbf{0}. \quad (2.269)$$

Multiplying this equation from the left with the inverse metric matrix \mathbf{G}^{-1} , we finally obtain the Einstein's field equation for a source-free gravitational field ($\mathbf{T} = \mathbf{0}$) as in (2.235)

$$\mathbf{R}_{\text{Ric}} - \frac{R}{2} \mathbf{I}_4 = \mathbf{0}.$$

(2.270)

2.13.1 Effects of Matter

So far, only the gravitational field in vacuum has been treated. To account for the sources of the gravitational field, e.g. matter, the action functional must include an additive term W_M that describes the source:

$$W \stackrel{\text{def}}{=} W_{\text{Grav}} + W_M = \int (kR + \mathcal{L}_M) \sqrt{-g} d^4\vec{x}. \quad (2.271)$$

The Lagrange equation with respect to \mathbf{G}^{-1} is

$$\begin{aligned} \frac{\partial(\sqrt{-g}(kR + \mathcal{L}_M))}{\partial \mathbf{G}^{-1}} &= \mathbf{0} \\ &= k \left(\frac{\partial \sqrt{-g}}{\partial \mathbf{G}^{-1}} R + \sqrt{-g} \frac{\partial R}{\partial \mathbf{G}^{-1}} \right) + \frac{\partial(\sqrt{-g}\mathcal{L}_M)}{\partial \mathbf{G}^{-1}}. \end{aligned} \quad (2.272)$$

The term inside the large parentheses gives the left-hand side of (2.268), i.e. together we first get for (2.272)

$$\mathbf{0} = k\sqrt{-g} \left(\frac{-1}{2} \cdot \mathbf{G} R + \check{\mathbf{R}}_{\text{Ric}} \right) + \sqrt{-g} \left(\frac{\partial \mathcal{L}_M}{\partial \mathbf{G}^{-1}} + \underbrace{\frac{\mathcal{L}_M}{\sqrt{-g}}}_{-\frac{1}{2}\sqrt{-g}\mathbf{G}} \frac{\partial \sqrt{-g}}{\partial \mathbf{G}^{-1}} \right).$$

Now, if we define the energy–momentum matrix as

$$-\frac{1}{2}\check{\mathbf{T}} \stackrel{\text{def}}{=} \frac{\partial \mathcal{L}_M}{\partial \mathbf{G}^{-1}} - \frac{1}{2}\mathcal{L}_M \mathbf{G}, \quad (2.273)$$

then we finally obtain with $k = c^4/(16\pi G)$ by left multiplication of (2.273) by \mathbf{G}^{-1} and with $\mathbf{T} \stackrel{\text{def}}{=} \mathbf{G}^{-1}\check{\mathbf{T}}$ the Einstein's field equation

$$\boxed{\mathbf{R}_{\text{Ric}} - \frac{R}{2}\mathbf{I}_4 = \frac{8\pi G}{c^4}\mathbf{T}.} \quad (2.274)$$

2.14 Most Important Definitions and Formulas

Christoffel Matrix (2.68):

$$\boxed{\boldsymbol{\Gamma} \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{\Gamma}_0 \\ \vdots \\ \boldsymbol{\Gamma}_3 \end{pmatrix} = \frac{1}{2}(\mathbf{G}^{-1} \otimes \mathbf{I}_4) \left[\begin{pmatrix} \frac{\partial \mathbf{g}_0^\top}{\partial \vec{x}} \\ \vdots \\ \frac{\partial \mathbf{g}_3^\top}{\partial \vec{x}} \end{pmatrix} + \begin{pmatrix} \frac{\partial \mathbf{g}_0}{\partial \vec{x}^\top} \\ \vdots \\ \frac{\partial \mathbf{g}_3}{\partial \vec{x}^\top} \end{pmatrix} - \frac{\partial \mathbf{G}}{\partial \vec{x}} \right] \in \mathbb{R}^{16 \times 4}}$$

with the components given, in accordance with (2.63) and (2.64), as

$$\Gamma_{\alpha\beta}^k = \sum_{i=0}^3 g_{ki}^{[-1]} \frac{1}{2} \left(\frac{\partial g_{\beta i}}{\partial x_\alpha} + \frac{\partial g_{\alpha i}}{\partial x_\beta} - \frac{\partial g_{\alpha\beta}}{\partial x_i} \right).$$

Motion in a Gravitational Field (2.67):

$$\ddot{\vec{x}} = -(\mathbf{I}_4 \otimes \dot{\vec{x}}^\top) \boldsymbol{\Gamma} \dot{\vec{x}}.$$

Riemannian Curvature Matrix (2.160):

$$\boxed{\mathbf{R} \stackrel{\text{def}}{=} \left[\frac{\partial \boldsymbol{\Gamma}}{\partial \vec{x}^\top} + (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) \right] (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16}) \in \mathbb{R}^{16 \times 16},}$$

with the components (2.165)

$$R_{\alpha\beta}^{\gamma\delta} = \frac{\partial}{\partial x_\beta} \Gamma_{\alpha\delta}^\gamma - \frac{\partial}{\partial x_\delta} \Gamma_{\alpha\beta}^\gamma + \sum_v \Gamma_{\beta v}^\gamma \Gamma_{\delta\alpha}^v - \sum_v \Gamma_{\delta v}^\gamma \Gamma_{\alpha\beta}^v.$$

Ricci Matrix $\mathbf{R}_{\text{Ric}} \in \mathbb{R}^{4 \times 4}$ (2.200):

$$R_{\text{Ric},\gamma\delta} \stackrel{\text{def}}{=} \text{trace}(\mathbf{R}^{\gamma\delta}) = \sum_{\nu=0}^3 R_{\nu\nu}^{\gamma\delta},$$

consisting of the traces of the sub-matrices $\mathbf{R}^{\gamma\delta}$ of \mathbf{R} with the components in accordance with (2.209):

$$R_{\text{Ric},\gamma\delta} = \sum_{\nu=0}^3 \left(\frac{\partial}{\partial x_\delta} \Gamma_{\gamma\nu}^\nu - \frac{\partial}{\partial x_\nu} \Gamma_{\gamma\delta}^\nu + \sum_{\mu=0}^3 \Gamma_{\delta\mu}^\nu \Gamma_{\nu\gamma}^\mu - \sum_{\mu=0}^3 \Gamma_{\nu\mu}^\nu \Gamma_{\gamma\delta}^\mu \right).$$

Curvature Scalar R (2.200):

$$R \stackrel{\text{def}}{=} \text{trace}(\mathbf{R}_{\text{Ric}}) = \sum_{\alpha} \sum_{\nu} R_{\nu\nu}^{\alpha\alpha},$$

which is obtained by taking the trace of the Ricci matrix.

Einstein's Field Equation (2.235) (fixed as an axiom):

$$\mathbf{R}_{\text{Ric}} - \frac{R}{2} \mathbf{I}_4 = -\frac{8\pi G}{c^4} \mathbf{T},$$

or with $T \stackrel{\text{def}}{=} \text{trace } \mathbf{T}$, see (2.237):

$$\mathbf{R}_{\text{Ric}} = \frac{8\pi G}{c^4} \left(\frac{T}{2} \mathbf{I}_4 - \mathbf{T} \right).$$

Chapter 3

Gravitation of a Spherical Mass

The chapter begins with the first solution of Einstein's equations obtained in 1916 by Schwarzschild for a spherical mass. From these results, the influence of a mass in time and space and the redshift of the spectral lines are given. This gives also a first indication of the existence of "black holes". Finally, a brief look at the effect of rotating masses and on the Lense–Thirring effect is taken.

3.1 Schwarzschild's Solution

Because of its simplicity, the solution of Einstein's field equation will now be determined for the *outside* of a spherically symmetric, uniform, time-invariant mass distribution. This first exact solution of Einstein's field equation was given in 1916 by Schwarzschild.

As $r \rightarrow \infty$, the desired metric should be the Minkowski metric, i.e.

$$ds^2 = c^2 dt^2 - dr^2 - r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (3.1)$$

where r, θ and φ are the spherical coordinates. For a gravitational field, we now start with

$$ds^2 = A(r) dt^2 - B(r) dr^2 - r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (3.2)$$

Since the field must be spherically symmetric, the factors A and B may depend only on r and not on θ or φ . Due to (3.1), the factor $A(r)$ must tend to c^2 as $r \rightarrow \infty$, and the factor $B(r) \rightarrow 1$. The metric matrix thus has the form

$$\mathbf{G} = \begin{pmatrix} A(r) & 0 & 0 & 0 \\ 0 & -B(r) & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2 \sin^2 \theta \end{pmatrix}. \quad (3.3)$$

3.1.1 Christoffel Matrix Γ

Due to (2.57)

$$\hat{\Gamma}_k \stackrel{\text{def}}{=} (\mathbf{I}_4 \otimes \mathbf{g}_k^{-T}) \left[\mathbf{I}_{16} - \frac{1}{2} \mathbf{U}_{4 \times 4} \right] \frac{\partial \mathbf{G}}{\partial \vec{x}}$$

and

$$\mathbf{G}^{-1} = \begin{pmatrix} \frac{1}{A(r)} & 0 & 0 & 0 \\ 0 & \frac{-1}{B(r)} & 0 & 0 \\ 0 & 0 & \frac{-1}{r^2} & 0 \\ 0 & 0 & 0 & \frac{-1}{r^2 \sin^2 \theta} \end{pmatrix}, \quad (3.4)$$

with

$$\frac{\partial \mathbf{G}}{\partial \vec{x}} = \left(\begin{array}{cccc} & \mathbf{0}_4 & & \\ \hline A'(r) & 0 & 0 & 0 \\ 0 & -B'(r) & 0 & 0 \\ 0 & 0 & -2r & 0 \\ 0 & 0 & 0 & -2r \sin^2 \theta \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2r^2 \sin \theta \cos \theta \\ \hline & \mathbf{0}_4 & & \end{array} \right),$$

in which $A' \stackrel{\text{def}}{=} \frac{\partial A}{\partial r}$ and $\mathbf{0}_4$ is a 4×4 -zero-matrix, and

$$\left[\mathbf{I}_{16} - \frac{1}{2} \mathbf{U}_{4 \times 4} \right] \frac{\partial \mathbf{G}}{\partial \vec{x}} = \left(\begin{array}{cccc} 0 & 0 & 0 & 0 \\ -\frac{A'}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline A' & 0 & 0 & 0 \\ 0 & -\frac{B'}{2} & 0 & 0 \\ 0 & 0 & -2r & 0 \\ 0 & 0 & 0 & -2r \sin^2 \theta \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & r & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2r^2 \sin \theta \cos \theta \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r \sin^2 \theta \\ 0 & 0 & 0 & r^2 \sin \theta \cos \theta \\ 0 & 0 & 0 & 0 \end{array} \right),$$

one obtains

$$\begin{aligned}\hat{\boldsymbol{\Gamma}}_0 &= (\mathbf{I}_4 \otimes [1/A(r) | 0 | 0 | 0]) \left[\mathbf{I}_{16} - \frac{1}{2} \mathbf{U}_{4 \times 4} \right] \begin{pmatrix} \frac{\partial G}{\partial t} \\ \frac{\partial G}{\partial r} \\ \frac{\partial G}{\partial \theta} \\ \frac{\partial G}{\partial \varphi} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ \frac{A'}{A} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (3.5)$$

Accordingly, we obtain

$$\begin{aligned}\hat{\boldsymbol{\Gamma}}_1 &= \begin{pmatrix} \frac{A'}{2B} & 0 & 0 & 0 \\ 0 & \frac{B'}{2B} & 0 & 0 \\ 0 & 0 & -\frac{r}{B} & 0 \\ 0 & 0 & 0 & -\frac{r}{B} \sin^2 \theta \end{pmatrix}, \\ \hat{\boldsymbol{\Gamma}}_2 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{r} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sin \theta \cos \theta \end{pmatrix} \end{aligned}$$

and

$$\hat{\boldsymbol{\Gamma}}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{r} \\ 0 & 0 & 0 & 2 \cot \theta \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Symmetrizing according to

$$\boldsymbol{\Gamma}_k = \frac{1}{2} (\hat{\boldsymbol{\Gamma}}_k + \hat{\boldsymbol{\Gamma}}_k^\top)$$

yields the symmetric matrices

$$\boldsymbol{\Gamma}_0 = \begin{pmatrix} 0 & \frac{A'}{(2A)} & 0 & 0 \\ \frac{A'}{(2A)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\boldsymbol{\Gamma}_1 = \hat{\boldsymbol{\Gamma}}_1,$$

$$\boldsymbol{\Gamma}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{r} & 0 \\ 0 & \frac{1}{r} & 0 & 0 \\ 0 & 0 & 0 & -\sin \theta \cos \theta \end{pmatrix}$$

and

$$\boldsymbol{\Gamma}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{r} \\ 0 & 0 & 0 & \cot\theta \\ 0 & \frac{1}{r} & \cot\theta & 0 \end{pmatrix},$$

which put altogether give

$$\boldsymbol{\Gamma} = \left(\begin{array}{cccc} 0 & \frac{A'}{2A} & 0 & 0 \\ \frac{A'}{2A} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & \frac{B'}{2B} & 0 & 0 \\ 0 & 0 & -\frac{r}{B} & 0 \\ 0 & 0 & 0 & -\frac{r}{B} \sin^2\theta \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{r} & 0 \\ 0 & \frac{1}{r} & 0 & 0 \\ 0 & 0 & 0 & -\sin\theta \cos\theta \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{r} \\ 0 & 0 & 0 & \cot\theta \\ 0 & \frac{1}{r} & \cot\theta & 0 \end{array} \right). \quad (3.6)$$

The nonzero Christoffel elements can be read from the matrices as:

$$\begin{aligned} \Gamma_{10}^0 = \Gamma_{01}^0 &= \frac{A'}{2A}, & \Gamma_{00}^1 &= \frac{A'}{2B}, & \Gamma_{11}^1 &= \frac{B'}{2B}, & \Gamma_{22}^1 &= -\frac{r}{B}, \\ \Gamma_{33}^1 &= -\frac{r}{B} \sin^2\theta, & \Gamma_{12}^2 &= \Gamma_{21}^2 = \frac{1}{r}, & \Gamma_{33}^2 &= -\sin\theta \cos\theta, \\ \Gamma_{13}^3 &= \Gamma_{31}^3 = \frac{1}{r} \quad \text{and} \quad \Gamma_{23}^3 &= \Gamma_{32}^3 = \cot\theta. \end{aligned}$$

3.1.2 Ricci Matrix \mathbf{R}_{Ric}

The Ricci matrix \mathbf{R}_{Ric} is obtained from the Riemannian curvature matrix

$$\mathbf{R} = \left[\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\tau} + (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) \right] (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16})$$

by summing the 4×4 -sub-matrices on the main diagonal. For the calculation of \mathbf{R} , in addition to $\boldsymbol{\Gamma}$, we need

$$\frac{\partial \boldsymbol{\Gamma}}{\partial \vec{x}^T}$$

$$= \left(\begin{array}{c|ccc|c|c} \mathbf{0}_4 & 0 & \frac{A''A-A'^2}{2A^2} & 0 & 0 & \mathbf{0}_4 & \mathbf{0}_4 \\ \hline \mathbf{0}_4 & \frac{A''A-A'^2}{2A^2} & 0 & 0 & 0 & & \\ & 0 & 0 & 0 & 0 & & \\ & 0 & 0 & 0 & 0 & & \\ \hline \mathbf{0}_4 & \frac{A''B-A'B'}{2B^2} & 0 & 0 & 0 & & \\ \mathbf{0}_4 & 0 & \frac{B''B-B'^2}{2B^2} & 0 & 0 & \mathbf{0}_{3 \times 4} & \mathbf{0}_4 \\ & 0 & 0 & -\frac{B-rB'}{B^2} & 0 & & \\ & 0 & 0 & 0 & -\sin^2 \theta \frac{B-rB'}{B^2} & 0 & \\ \hline \mathbf{0}_4 & 0 & 0 & 0 & 0 & & \\ \mathbf{0}_4 & 0 & 0 & -\frac{1}{r^2} & 0 & \mathbf{0}_{3 \times 4} & \mathbf{0}_4 \\ & 0 & -\frac{1}{r^2} & 0 & 0 & & \\ & 0 & 0 & 0 & 0 & 0 & \\ \hline \mathbf{0}_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{0}_4 & 0 & 0 & 0 & \frac{-1}{r^2} & 0 & \\ & 0 & 0 & 0 & 0 & 0 & \\ & 0 & \frac{-1}{r^2} & 0 & 0 & 0 & \frac{-1}{\sin^2 \theta} \\ \end{array} \right)$$

and

$$\bar{\boldsymbol{\Gamma}} = \left(\begin{array}{c|cc|c|cc|c|cc|c|cc|c} 0 & \frac{A'}{2A} & 0 & 0 & \frac{A'}{2A} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{A'}{2B} & 0 & 0 & 0 & 0 & \frac{B'}{2B} & 0 & 0 & 0 & 0 & -\frac{r}{B} & 0 & 0 & -\frac{r}{B} \sin^2 \theta \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{r} & 0 & 0 & \frac{1}{r} & 0 & 0 & 0 & -\sin \theta \cos \theta \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{r} & 0 & 0 & 0 & \cot \theta & 0 & \\ \end{array} \right).$$

For the Ricci matrix \mathbf{R}_{Ric} the 16×16 -matrix $\frac{\partial \boldsymbol{\Gamma}}{\partial \vec{x}^T} (\mathbf{U}_{4 \times 4} - \mathbf{I}_{16})$ contributes the term

$$\left(\begin{array}{cccc} \frac{A'B'-A''B}{2B^2} & 0 & 0 & 0 \\ 0 & \frac{A''A-A'^2}{2A^2} - \frac{2}{r^2} & 0 & 0 \\ 0 & 0 & \frac{B-rB'}{B^2} - \frac{1}{\sin^2 \theta} & 0 \\ 0 & 0 & 0 & \sin^2 \theta \frac{B-rB'}{B^2} + \cos^2 \theta - \sin^2 \theta \end{array} \right)$$

and the matrix $(\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma})(\mathbf{U}_{4 \times 4} - \mathbf{I}_{16})$ the term

$$\left(\begin{array}{cccc} \frac{A'^2}{4AB} - \frac{A'B'}{4B^2} - \frac{A'}{rB} & 0 & 0 & 0 \\ 0 & \frac{A'^2}{4A^2} - \frac{A'B'}{4AB} - \frac{B'}{rB} + \frac{2}{r^2} & 0 & 0 \\ 0 & 0 & \frac{rB'}{B^2} + \frac{rA'}{2AB} + \cot^2 \theta & 0 \\ 0 & 0 & 0 & \frac{rB'}{B^2} \sin^2 \theta + \frac{rA'}{2AB} \sin^2 \theta - \cos^2 \theta \end{array} \right).$$

Finally, we obtain for the four elements on the main diagonal of the Ricci matrix \mathbf{R}_{Ric} :

$$\underline{\underline{R_{\text{Ric},00} = -\frac{A''}{2B} + \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{A'}{rB}}}, \quad (3.7)$$

$$\underline{\underline{R_{\text{Ric},11} = \frac{A''}{2A} - \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{B'}{rB}}}, \quad (3.8)$$

$$\underline{\underline{R_{\text{Ric},22} = \frac{1}{B} + \frac{r}{2B} \left(\frac{A'}{A} - \frac{B'}{B} \right) - 1}} \quad (3.9)$$

and

$$\underline{\underline{R_{\text{Ric},33} = \sin^2 \theta R_{\text{Ric},22}}} \quad (3.10)$$

The other matrix elements are equal to zero: $R_{\text{Ric},v\mu} = 0$ for $v \neq \mu$.

3.1.3 The Factors $A(r)$ and $B(r)$

To finally write down the metric matrix \mathbf{G} , the two factors $A(r)$ and $B(r)$ are required. Since we search for the solution of Einstein's field equation outside the sphere containing the mass and having the energy-momentum matrix \mathbf{T} equal to the zero matrix, the Ansatz $\mathbf{R}_{\text{Ric}} = \mathbf{0}$ thus yields

$$-\frac{A''}{2B} + \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{A'}{rB} = 0, \quad (3.11)$$

$$\frac{A''}{2A} - \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{B'}{rB} = 0, \quad (3.12)$$

and

$$\frac{1}{B} + \frac{r}{2B} \left(\frac{A'}{A} - \frac{B'}{B} \right) - 1 = 0. \quad (3.13)$$

Adding to (3.11) divided by A (3.12) divided by B , we obtain the condition

$$A'B + AB' = 0, \quad (3.14)$$

which means nothing else than that

$$AB = \text{constant}. \quad (3.15)$$

However, since $A(\infty) = c^2$ and $B(\infty) = 1$ as $r \rightarrow \infty$, the constant must be equal to c^2 , so

$$A(r)B(r) = c^2 \quad \text{and} \quad B(r) = \frac{c^2}{A(r)}.$$

Substituting this into (3.13), we obtain $A + rA' = c^2$, which we also can write

$$\frac{d(rA)}{dr} = c^2. \quad (3.16)$$

Integrating this equation yields

$$rA = c^2(r + K), \quad (3.17)$$

so

$$A(r) = c^2 \left(1 + \frac{K}{r} \right) \quad \text{and} \quad B(r) = \left(1 + \frac{K}{r} \right)^{-1}. \quad (3.18)$$

K still remains to be determined. In Sect. 2.8, a uniform rotating system was treated. In (2.85), we obtained the element g_{00} of the metric matrix \mathbf{G} as

$$g_{00} = 1 - \frac{r^2\omega^2}{c^2} \quad (3.19)$$

and in (2.88) the centrifugal force $mr\omega^2$. According to Sect. 2.8, the centrifugal force f depends, on the other hand, on the centrifugal potential φ , so together

$$f = -m\nabla\varphi. \quad (3.20)$$

Taking

$$\varphi = -\frac{r^2\omega^2}{2} \quad (3.21)$$

(3.20) provides just the above centrifugal force. Equation (3.21) inserted into (3.19) provides the general connection

$$g_{00} = 1 + \frac{2\varphi}{c^2}. \quad (3.22)$$

Outside of our spherical mass M , as Newton's approximation the gravitational potential

$$\varphi = -\frac{GM}{r} \quad (3.23)$$

is obtained. This used in (3.22) and compared with (3.18) yields

$$K = -\frac{2GM}{c^2}, \quad (3.24)$$

so that we finally obtain the Schwarzschild's metric in matrix form:

$$ds^2 = d\vec{x}^\top \mathbf{G} d\vec{x} = d\vec{x}^\top \begin{pmatrix} 1 - \frac{2GM}{c^2r} & 0 & 0 & 0 \\ 0 & -(1 - \frac{2GM}{c^2r})^{-1} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2 \sin^2 \theta \end{pmatrix} d\vec{x}, \quad (3.25)$$

with

$$d\vec{x} \stackrel{\text{def}}{=} \begin{pmatrix} c dt \\ dr \\ d\theta \\ d\varphi \end{pmatrix},$$

and the calculated Schwarzschild's metric is then

$$ds^2 = \left(1 - \frac{2GM}{c^2r}\right) c^2 dt^2 - \left(1 - \frac{2GM}{c^2r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (3.26)$$

For the so-called Schwarzschild's radius

$$r_S \stackrel{\text{def}}{=} \frac{2GM}{c^2}, \quad (3.27)$$

the Schwarzschild's metric shows a singularity as for $r = 0$. Of these, however, only the singularity at $r = 0$ is a real singularity. This is suggested even by the fact that, when $r = r_S$, the determinant of \mathbf{G} , namely $g = -r^4 \sin^2 \theta$, has no singularity. The singularity at $r = r_S$ is not a *physical singularity* but a *coordinate singularity*, which depends entirely on the choice of the coordinate system, i.e. there would exist no singularity at this place if a different coordinate system were chosen! More about this in the next section on black holes.

3.2 Influence of a Massive Body on the Environment

3.2.1 Introduction

The Schwarzschild's solution (3.26) is valid only outside the solid sphere of radius r_M generating the gravity, so only for $r_M < r < \infty$. Because for the Schwarzschild's radius r_S the element $g_{11} = (1 - \frac{r_S}{r})^{-1}$ goes to infinity, r_S is also a limit. In general, $r_S \ll r_M$, but for the so-called *Black Holes* one has $r_M < r_S$, and in this case the solution is limited to $r_S < r < \infty$.

3.2.2 Changes to Length and Time

How does a length change in the environment of a massive body? For a constant time, i.e. for $dt = 0$ one receives, due to (3.26),

$$d\ell^2 = \left(1 - \frac{2GM}{c^2r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2). \quad (3.28)$$

On the surface of a sphere of radius $r > r_M$ with the centre being the centre of mass, as also for $dr = 0$, the tangential infinitely small distances are given by

$$dL = r(d\theta^2 + \sin^2\theta d\varphi^2)^{1/2}. \quad (3.29)$$

This is a result that is valid on every sphere, whether with or without gravity. But what happens with a distance in the radial direction? In this case, $d\theta$ and $d\varphi$ are equal to zero, thus for infinitely small distances in the radial direction, according to (3.28), one obtains

$$dR = \left(1 - \frac{2GM}{c^2r}\right)^{-1/2} dr, \quad (3.30)$$

so $dR > dr$, and so the length is much longer, the larger the mass M and the smaller the distance r from the mass! dr is through the mass ‘elongated’, caused by a curvature of space.

Let us now turn to the time. For a clock at the point, where r, θ , and φ are constant, we obtain from the Schwarzschild’s metric (3.26)

$$ds^2 = c^2 d\tau^2 = c^2 \left(1 - \frac{2GM}{c^2r}\right) dt^2,$$

so

$$d\tau = \left(1 - \frac{2GM}{c^2r}\right)^{1/2} dt. \quad (3.31)$$

If $d\tau < dt$, the closer one is located to the mass, the shorter are the time intervals, i.e. the slower the time passes! For an observer the time goes slower the closer he is to the mass. Particularly large is the time dilation near a *Black Hole* in which the mass is so concentrated that the body radius is smaller than its own Schwarzschild’s radius. For a black hole with a mass of ten solar masses, the Schwarzschild’s radius is $r_S = 30$ km. At a distance of 1 cm from the so-called horizon, which is the spherical shell around the centre of mass with radius r_S , $\gamma = (1 - \frac{2GM}{c^2r})^{1/2} = 1.826 \cdot 10^{-5}$, so the time goes by about 55000 times slower than far away: after a year near the horizon of the black hole, in a place far away, e.g. on the home planet of an astronaut, 55000 years would have gone by! All is summed up into a popular statement:

If you climb a mountain, you are smaller and age faster!

The relations (3.30) and (3.31) are very similar to the relationships of space and time contractions for the relatively to each other moving reference frames in special relativity; see Chap. 1. There with

$$\gamma \stackrel{\text{def}}{=} \left(1 - \frac{v^2}{c^2}\right)^{-1/2}$$

for the length contraction we had

$$d\ell = \frac{1}{\gamma} d\ell_0$$

and for the time dilation

$$dt = \gamma d\tau.$$

If we introduce the pseudo-velocity

$$v_G^2(r) \stackrel{\text{def}}{=} \frac{2GM}{r}$$

then with

$$\gamma_G(r) \stackrel{\text{def}}{=} \left(1 - \frac{v_G^2(r)}{c^2}\right)^{-1/2}$$

the above gravitational relationships can be written as follows:

$$dr = \frac{1}{\gamma_G(r)} dR \quad \text{and} \quad dt = \gamma_G(r) d\tau.$$

By the way, $v_G(r)$ is the escape velocity of a planet with the diameter $2r$ and mass M .

3.2.3 Redshift of Spectral Lines

Suppose in the gravitational field of a mass a light signal is sent from a transmitter at a fixed point $\mathbf{x}_T = [r_T, \theta_T, \varphi_T]^\top$ and it drifts along a geodesic line to a fixed receiver at the point $\mathbf{x}_R = [r_R, \theta_R, \varphi_R]^\top$. Let, furthermore, λ be any parameter along the geodesic line, with $\lambda = \lambda_T$ for the sent event and $\lambda = \lambda_R$ for the received event. Then for a photon, by (2.11), the following is valid:

$$\frac{d\vec{x}^\top}{d\lambda} G \frac{d\vec{x}}{d\lambda} = 0,$$

so here

$$c^2 \left(1 - \frac{2GM}{c^2 r}\right) \left(\frac{dt}{d\lambda}\right)^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} \left(\frac{dr}{d\lambda}\right)^2 - r^2 \left(\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\left(\frac{d\varphi}{d\lambda}\right)^2\right) = 0,$$

i.e.

$$\frac{dt}{d\lambda} = \frac{1}{c} \left[\left(1 - \frac{2GM}{c^2 r}\right)^{-2} \left(\frac{dr}{d\lambda}\right)^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} r^2 \left(\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\left(\frac{d\varphi}{d\lambda}\right)^2\right) \right]^{1/2}.$$

From this we obtain the signal transmission time

$$t_R - t_T = \frac{1}{c} \int_{\lambda_T}^{\lambda_R} \left[\left(1 - \frac{2GM}{c^2 r}\right)^{-2} \left(\frac{dr}{d\lambda}\right)^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} r^2 \left(\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\left(\frac{d\varphi}{d\lambda}\right)^2\right) \right]^{1/2} d\lambda.$$

This time depends only on the path that the light takes between the spatially fixed transmitter and the spatially fixed receiver. So for two consecutive transmitted signals 1 and 2 the duration is equal:

$$t_{E,1} - t_{S,1} = t_{E,2} - t_{S,2},$$

and also

$$\Delta t_R \stackrel{\text{def}}{=} t_{E,2} - t_{E,1} = t_{S,2} - t_{S,1} \stackrel{\text{def}}{=} \Delta t_T,$$

so the Schwarzschild's time difference at the transmitter is equal to the Schwarzschild's time difference at the receiver, though the clock of an observer at the transmitter location would show the proper time difference to

$$\Delta \tau_T = \left(1 - \frac{2GM}{c^2 r_T}\right)^{1/2} \Delta t_T$$

and accordingly at the receiving location the proper time difference would be

$$\Delta \tau_R = \left(1 - \frac{2GM}{c^2 r_R}\right)^{1/2} \Delta t_R.$$

Since $\Delta t_R = \Delta t_T$, one obtains the ratio

$$\frac{\Delta \tau_R}{\Delta \tau_T} = \left(\frac{1 - \frac{2GM}{c^2 r_R}}{1 - \frac{2GM}{c^2 r_T}} \right)^{1/2}.$$

If the transmitter is a pulsating atom, emitting N pulses in the proper time interval $\Delta \tau_T$, an observer at the transmitter would assign to the signal a proper frequency

$v_T \stackrel{\text{def}}{=} \frac{N}{\Delta\tau_T}$. An observer at the receiver will see the N pulses arriving in the proper time $\Delta\tau_R$, thus at the frequency $v_R = \frac{N}{\Delta\tau_R}$. Then the frequency ratio is obtained as

$$\boxed{\frac{v_R}{v_T} = \left(\frac{1 - \frac{2GM}{c^2 r_T}}{1 - \frac{2GM}{c^2 r_R}} \right)^{1/2}.} \quad (3.32)$$

If the transmitter is closer to the mass compared to the receiver ($r_T < r_R$), then a shift to the red in the “colour” of the signal takes place. Conversely, if the transmitter is further away than the receiver, one gets a blueshift. For $r_S, r_E \gg 2GM$ one obtains the approximation

$$\frac{v_R}{v_T} \approx 1 + \frac{GM}{c^2} \left(\frac{1}{r_R} - \frac{1}{r_T} \right)$$

and the relative frequency change

$$\frac{\Delta\nu}{v_T} \stackrel{\text{def}}{=} \frac{v_R - v_T}{v_T} \approx \frac{GM}{c^2} \left(\frac{1}{r_R} - \frac{1}{r_T} \right).$$

If the transmitter (e.g. a radiating atom) is on the solar surface and the observing receiver on the surface of the Earth, then

$$\frac{\Delta\nu}{v_T} \approx 2 \cdot 10^{-6}.$$

This effect is indeed poorly measurable due to a variety of disturbances by the atmosphere. However, this redshift is measurable using the Mössbauer's effect (see the relevant Physics literature).

3.2.4 Deflection of Light

According to (2.11), for light one has

$$\vec{x}^\top \mathbf{G} \vec{x} = 0. \quad (3.33)$$

Differentiating (3.33) with respect to the orbital parameter λ , one obtains

$$\frac{\partial \vec{x}^\top}{\partial \lambda} \mathbf{G} \frac{\partial \vec{x}}{\partial \lambda} = 0. \quad (3.34)$$

Setting for \mathbf{G} the Schwarzschild's metric, we obtain

$$c^2 \left(1 - \frac{r_S}{r} \right) \left(\frac{\partial t}{\partial \lambda} \right)^2 - \left(1 - \frac{r_S}{r} \right)^{-1} \left(\frac{\partial r}{\partial \lambda} \right)^2 - r^2 \left(\frac{\partial \theta}{\partial \lambda}^2 + \sin^2 \theta \left(\frac{\partial \varphi}{\partial \lambda} \right)^2 \right) = 0. \quad (3.35)$$

Without loss of generality, especially in the present centrally symmetric solution, we can assume that $\theta = \pi/2$ and $\frac{\partial\theta}{\partial\lambda} = 0$, and that the solution is in a plane through the centre of mass, then (3.35) simplifies to

$$c^2 \left(1 - \frac{r_s}{r}\right) \left(\frac{\partial t}{\partial\lambda}\right)^2 - \left(1 - \frac{r_s}{r}\right)^{-1} \left(\frac{\partial r}{\partial\lambda}\right)^2 - r^2 \left(\frac{\partial\varphi}{\partial\lambda}\right)^2 = 0. \quad (3.36)$$

For a light ray in the gravitational field the following holds:

$$\frac{\partial^2 \vec{x}}{\partial\lambda^2} = - \left(I_4 \otimes \frac{\partial \vec{x}^\top}{\partial\lambda} \right) \boldsymbol{\Gamma} \frac{\partial \vec{x}}{\partial\lambda}. \quad (3.37)$$

With the Christoffel elements from Sect. 3.1.1 and $\vec{x}^\top = [x_0 | r | \theta | \varphi]$ expanding (3.37) yields

$$\frac{d^2 x_0}{d\lambda^2} = - \frac{A'}{A} \frac{dx_0}{d\lambda} \frac{dr}{d\lambda}, \quad (3.38)$$

$$\frac{d^2 r}{d\lambda^2} = - \frac{A'}{B} \left(\frac{dx_0}{d\lambda} \right)^2 - \frac{B'}{2B} \left(\frac{dr}{d\lambda} \right)^2 + \frac{r}{B} \left(\frac{d\theta}{d\lambda} \right)^2 + \frac{r \sin^2 \theta}{B} \left(\frac{d\varphi}{d\lambda} \right)^2, \quad (3.39)$$

$$\frac{d^2 \theta}{d\lambda^2} = - \frac{2}{r} \frac{d\theta}{d\lambda} \frac{dr}{d\lambda} + \sin \theta \cos \theta \left(\frac{d\varphi}{d\lambda} \right)^2, \quad (3.40)$$

$$\frac{d^2 \varphi}{d\lambda^2} = - \frac{2}{r} \frac{d\varphi}{d\lambda} \frac{dr}{d\lambda} - \cot \theta \frac{d\theta}{d\lambda} \frac{d\varphi}{d\lambda}. \quad (3.41)$$

If the coordinate system is selected so that for the start time λ_0 one has

$$\theta = \pi/2 \quad \text{and} \quad \frac{\partial\theta}{\partial\lambda} = 0, \quad (3.42)$$

then from (3.40) $\frac{d^2\theta}{d\lambda^2} = 0$, so $\theta(\lambda) \equiv \pi/2$, i.e. the entire path remains in the plane through the centre of mass. With this θ value from (3.41) one obtains

$$\frac{d^2 \varphi}{d\lambda^2} + \frac{2}{r} \frac{d\varphi}{d\lambda} \frac{dr}{d\lambda} = 0,$$

which can be summarized as follows:

$$\frac{1}{r^2} \frac{d}{d\lambda} \left(r^2 \frac{d\varphi}{d\lambda} \right) = 0. \quad (3.43)$$

It is therefore true that

$$\underline{\underline{r^2 \frac{d\varphi}{d\lambda} = \text{const.} = h.}} \quad (3.44)$$

Equation (3.38) can be converted to

$$\frac{d}{d\lambda} \left(\ln \frac{dx_0}{d\lambda} + \ln A \right) = 0 \quad (3.45)$$

and integrated to yield

$$\underline{\underline{A \frac{dx_0}{d\lambda} = \text{const.} = k}} \quad (3.46)$$

Equations (3.42), (3.44) and (3.46) used in (3.39) provide

$$\frac{d^2r}{d\lambda^2} + \frac{k^2 A'}{2A^2 B} + \frac{B'}{2B} \left(\frac{dr}{d\lambda} \right)^2 - \frac{h^2}{Br^3} = 0. \quad (3.47)$$

Multiplying this equation by $2B \frac{dr}{d\lambda}$, we obtain first

$$\frac{d}{d\lambda} \left(B \frac{dr}{d\lambda} + \frac{h^2}{r^2} - \frac{k^2}{A} \right) = 0 \quad (3.48)$$

and after integration

$$B \left(\frac{dr}{d\lambda} \right)^2 + \frac{h^2}{r^2} - \frac{k^2}{A} = \text{const.} = 0, \quad (3.49)$$

or

$$\frac{dr}{d\lambda} = \sqrt{\frac{\frac{k^2}{A} - \frac{h^2}{r^2}}{B}}. \quad (3.50)$$

But we don't look for $r(\lambda)$, we want $\varphi(r)$. Since

$$\frac{d\varphi}{dr} = \frac{d\varphi}{d\lambda} \frac{d\lambda}{dr},$$

it follows by (3.44) and (3.50) that

$$\frac{d\varphi}{dr} = \frac{h}{r^2} \sqrt{\frac{B}{\frac{k^2}{A} - \frac{h^2}{r^2}}} \quad (3.51)$$

and integrated

$$\underline{\underline{\varphi(r) = \varphi(r_0) + \int_{r_0}^r \frac{\sqrt{B(\psi)}}{\psi^2 \sqrt{\frac{k^2}{A(\psi)h^2} - \frac{1}{\psi^2}}} d\psi.}} \quad (3.52)$$

Let r_0 be the smallest distance from the centre of mass, having a passing flying photon. It moves in the plane through the centre of mass with $\theta = \pi/2$. The coordinate system is still placed such that $\varphi(r_0) = 0$. If gravity is not present, the

photon will continue to fly straight and as $r \rightarrow \infty$ we will have $\varphi(\infty) = \pi/2$. It comes from $\varphi(-\infty) = -\pi/2$. Overall, therefore, the photon has covered an angle of $\Delta\varphi = \varphi(\infty) - \varphi(-\infty) = \pi$. Considering the gravity, the photon is “bent” to the mass. It passes from $(r_0, \varphi(r_0))$ to $\varphi(\infty) = \pi/2 + \alpha/2$ as $r \rightarrow \infty$. For reasons of symmetry, it is then $\varphi(-\infty) = \pi/2 - \alpha/2$ as $r \rightarrow -\infty$, so that the photon has covered an overall angle of $\Delta\varphi = \varphi(\infty) - \varphi(-\infty) = \pi + \alpha$. This angle α is now calculated for the solar mass $M_\odot = 1.9891 \times 10^{33}$ g.

Since r_0 should be the minimum distance, it is true that

$$\left. \frac{dr}{d\varphi} \right|_{r_0} = 0. \quad (3.53)$$

On the other hand, $\frac{d\varphi}{dr}$ is precisely the integrand of (3.52), whence it follows

$$\frac{k^2}{h^2} = \frac{A(r_0)}{r_0^2}. \quad (3.54)$$

The integral used in (3.52) yields

$$\varphi(\infty) = \int_{r_0}^{\infty} \frac{\sqrt{B(r)}}{r \sqrt{\frac{r^2 A(r_0)}{r_0^2 A(r)} - 1}} dr. \quad (3.55)$$

For this integral no simple antiderivative exists so that it should be calculated by approximations. One has

$$A(r) = 1 - \frac{2GM_\odot}{c^2 r} \quad (3.56)$$

and

$$B(r) = \left(1 - \frac{2GM_\odot}{c^2 r}\right)^{-1} \approx 1 + \frac{2GM_\odot}{c^2 r}. \quad (3.57)$$

Furthermore,

$$\begin{aligned} \frac{A(r_0)}{A(r)} &= \left(1 - \frac{2GM_\odot}{c^2 r_0}\right) \left(1 - \frac{2GM_\odot}{c^2 r}\right)^{-1} \\ &\approx \left(1 - \frac{2GM_\odot}{c^2 r_0}\right) \left(1 + \frac{2GM_\odot}{c^2 r}\right) = 1 + \frac{2GM_\odot}{c^2} \left(\frac{1}{r} - \frac{1}{r_0}\right) \end{aligned}$$

and

$$\frac{r^2 A(r_0)}{r_0^2 A(r)} - 1 \approx \left(\frac{r^2}{r_0^2} - 1\right) \left(1 - \frac{2GM_\odot r}{c^2 r_0(r+r_0)}\right). \quad (3.58)$$

Therefore, the integral (3.55) becomes

$$\begin{aligned}\varphi(\infty) &= \int_{r_0}^{\infty} \frac{r_0}{\sqrt{r^2 - r_0^2}} \left(\frac{1}{r} + \frac{GM_{\odot}}{c^2 r^2} + \frac{GM_{\odot}}{c^2(r+r_0)} \right) dr \\ &= \left[\arccos \frac{r_0}{r} + \frac{GM_{\odot}}{c^2 r_0} \frac{\sqrt{r^2 - r_0^2}}{r} + \frac{GM_{\odot}}{c^2 r_0} \sqrt{\frac{r - r_0}{r + r_0}} \right]_{r_0}^{\infty},\end{aligned}\quad (3.59)$$

so

$$\varphi(\infty) = \underline{\underline{\frac{\pi}{2} + \frac{2GM_{\odot}}{c^2 r_0}}}, \quad (3.60)$$

from which one can read off

$$\alpha = \underline{\underline{\frac{4GM_{\odot}}{c^2 r_0}}}.\quad (3.61)$$

If r_0 is precisely the solar radius $R_{\odot} = 6.957 \times 10^8$ m, we obtain

$$\underline{\underline{\alpha = 1.75''}}.\quad (3.62)$$

This prediction was one of the first confirmations of the Theory of General Relativity in 1919 at a solar eclipse measured by Eddington.

3.3 Schwarzschild's Inner Solution

For the determination of the metric inside a symmetric sphere, the right-hand side of Einstein's field equation with the energy-momentum matrix is needed. The hydro-mechanical energy-momentum matrix was deduced in Chap. 1 in (1.183) as

$$\boldsymbol{T}_{\text{mech}} = \left(\rho_0 + \frac{p}{c^2} \right) \bar{\boldsymbol{u}} \bar{\boldsymbol{u}}^{\top} - p \boldsymbol{M}.$$

This was done in the Lorentz basis of an inertial frame. It can thus also be the local inertial system of a general coordinate system. If for the relationship between the local coordinates $d\vec{x}$ and the global coordinates $d\vec{x}'$ the transformation equation is $d\vec{x} = \boldsymbol{J} d\vec{x}'$, then for the velocities one has $\bar{\boldsymbol{u}} = \boldsymbol{J} \bar{\boldsymbol{u}}'$. If used above this yields

$$\boldsymbol{T}_{\text{mech}} = \left(\rho_0 + \frac{p}{c^2} \right) \boldsymbol{J} \bar{\boldsymbol{u}}' \bar{\boldsymbol{u}}'^{\top} \boldsymbol{J}^{\top} - p \boldsymbol{M}.\quad (3.63)$$

Now multiplying this equation from the left with the matrix product $\boldsymbol{J}^{\top} \boldsymbol{M}$ and from the right with the matrix product $\boldsymbol{M} \boldsymbol{J}$, we finally obtain with $\boldsymbol{M} \boldsymbol{M} = \boldsymbol{I}$ and $\boldsymbol{J}^{\top} \boldsymbol{M} \boldsymbol{J} = \boldsymbol{G}$

$$\boldsymbol{T}_{\text{mech,Riemann}} = \left(\rho_0 + \frac{p}{c^2} \right) \boldsymbol{G} \bar{\boldsymbol{u}}' \bar{\boldsymbol{u}}'^{\top} \boldsymbol{G} - p \boldsymbol{G}.\quad (3.64)$$

Now the metric matrix \mathbf{G} appears globally.

The Einstein's field equation is used in the form (2.237)

$$\mathbf{R}_{\text{Ric}} = \frac{8\pi G}{c^4} \left(\frac{T}{2} \mathbf{I}_4 - \mathbf{T} \right). \quad (3.65)$$

It is assumed that the masses inside the sphere do not move, the velocity components are zero: $u'_i = 0$, for $i = 1, 2, 3$. Therefore, from the condition

$$c^2 = \vec{\mathbf{u}}'^\top \mathbf{G} \vec{\mathbf{u}}'$$

there remains

$$c^2 = g_{00} u_0^2.$$

If the metric matrix is chosen again as in (3.2), namely

$$ds^2 = A(r) dt^2 - B(r) dr^2 - r^2(d\theta^2 + \sin^2 \theta d\varphi^2), \quad (3.66)$$

then

$$u_0 = \frac{c}{\sqrt{A(r)}}.$$

Thus we get for \mathbf{T} the diagonal matrix

$$\mathbf{T} = \text{diag}(\varrho c^2 A(r), p B(r), pr^2, pr^2 \sin^2 \theta). \quad (3.67)$$

With the trace $T = c^2 \varrho - 3p$ of the local energy-momentum matrix, we obtain for the right-hand side of the Einstein's field equation

$$\begin{aligned} & \frac{8\pi G}{c^4} \left(\frac{T}{2} \mathbf{I}_4 - \mathbf{T} \right) \\ &= \frac{4\pi G}{c^4} \text{diag}((\varrho c^2 + 3p)A, (\varrho c^2 - p)B, (\varrho c^2 - p)r^2, (\varrho c^2 - p)r^2 \sin^2 \theta). \end{aligned} \quad (3.68)$$

This with the elements (3.7), (3.8), (3.9) and (3.10) on the main diagonal of the Ricci matrix \mathbf{R}_{Ric} gives

$$\begin{aligned} R_{\text{Ric},00} &= -\frac{A''}{2B} + \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{A'}{rB}, \\ R_{\text{Ric},11} &= \frac{A''}{2A} - \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{B'}{rB}, \\ R_{\text{Ric},22} &= \frac{1}{B} + \frac{r}{2B} \left(\frac{A'}{A} - \frac{B'}{B} \right) - 1 \end{aligned}$$

and

$$R_{\text{Ric},33} = \sin^2 \theta R_{\text{Ric},22}$$

(the remaining matrix elements are equal to zero: $R_{\text{Ric},v\mu} = 0$ for $v \neq \mu$), combined with Einstein's field equation yielding the three determining equations for the factors $A(r)$ and $B(r)$:

$$\frac{A''}{2B} - \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) + \frac{A'}{rB} = \frac{4\pi G}{c^4} (\varrho c^2 + 3p) A, \quad (3.69)$$

$$-\frac{A''}{2A} + \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) + \frac{B'}{rB} = \frac{4\pi G}{c^4} (\varrho c^2 - p) B \quad (3.70)$$

and

$$-\frac{1}{B} - \frac{r}{2B} \left(\frac{A'}{A} - \frac{B'}{B} \right) + 1 = \frac{4\pi G}{c^4} (\varrho c^2 - p) r^2. \quad (3.71)$$

Adding together (3.69) multiplied by $r^2/(2A)$, (3.70) multiplied by $r^2/(2B)$ and (3.71), we obtain

$$\frac{B'r}{B^2} + 1 - \frac{1}{B} = \frac{8\pi G}{c^2} \varrho r^2. \quad (3.72)$$

This can also be rewritten as

$$\frac{d}{dr} \frac{r}{B(r)} = 1 - \frac{8\pi G}{c^2} \varrho r^2. \quad (3.73)$$

Integrating this equation provides

$$\frac{r}{B(r)} = \int_0^r \left(1 - \frac{8\pi G}{c^2} \varrho(\alpha) \alpha^2 \right) d\alpha = r - \frac{2G}{c^2} \mathcal{M}(r), \quad (3.74)$$

with

$$\mathcal{M}(r) \stackrel{\text{def}}{=} 4\pi \int_0^r \varrho(\alpha) \alpha^2 d\alpha. \quad (3.75)$$

Solving (3.74) for $B(r)$ finally yields

$$\underline{\underline{B(r) = \left(1 - \frac{2G\mathcal{M}(r)}{c^2 r} \right)^{-1}}} \quad (3.76)$$

The density function $\varrho(r)$ is zero for $r > R$, i.e. outside the spherical mass, so $\mathcal{M}(r) = \mathcal{M}(R) = M$ for $r > R$. We thus obtain the same coefficient $B(r)$ for $r > R$ as for the Schwarzschild's outer solution.

Now $A(r)$ is still to be determined. For that, $B(r)$ and

$$B'(r) = B^2(r) \left(\frac{8\pi G}{c^2} \varrho(r) r - \frac{2G\mathcal{M}(r)}{c^2 r^2} \right)$$

is used in (3.71), which gives

$$-\frac{A'(r)r}{2A(r)} \left(1 - \frac{2G\mathcal{M}(r)}{c^2r}\right) + \frac{4\pi G}{c^2} \varrho(r)r^2 + \frac{G\mathcal{M}(r)}{c^2r} = \frac{4\pi G}{c^4} (\varrho c^2 - p)r^2,$$

i.e.

$$\frac{A'(r)}{A(r)} = \frac{d}{dr}(\ln A(r) + \text{const.}) = \left(\frac{8\pi G}{c^4} p(r)r + \frac{2G\mathcal{M}(r)}{c^2r^2}\right) \left(1 - \frac{2G\mathcal{M}(r)}{c^2r}\right)^{-1}$$

$$\stackrel{\text{def}}{=} f(r).$$

The sum $\ln A(r) + \text{const.}$ is thus the antiderivative for the right-hand side $f(r)$ of this equation. Therefore, $\ln A(r)$ equals the integral of $f(r)$. If we use as integration limits r and ∞ and $A(\infty) = 1$, we finally obtain

$$\underline{\underline{A(r) = \exp \left[-\frac{2G}{c^2} \int_r^\infty \frac{1}{\alpha^2} (\mathcal{M}(\alpha) + 4\pi\alpha^3 p(\alpha)/c^2) \left(1 - \frac{2G\mathcal{M}}{c^2\alpha}\right)^{-1} d\alpha \right]}}. \quad (3.77)$$

For $r > R$, this solution is equal to the Schwarzschild's outer solution because for $r > R$ one has $\varrho(r) = p(r) = 0$ and $\mathcal{M}(r) = \mathcal{M}(R) = M$, so

$$A(r) = \exp \left[-\frac{2G}{c^2} \int_r^\infty \frac{M}{\alpha^2} \left(1 - \frac{r_s}{\alpha}\right)^{-1} d\alpha \right]. \quad (3.78)$$

With the new integration variable $\xi = 1 - \frac{r_s}{r}$ we obtain for $r > R$

$$A(r) = \exp \left[\int_1^{1-\xi} \frac{1}{\xi'} d\xi' \right] = 1 - \frac{r_s}{r}, \quad (3.79)$$

i.e. the same solution as the Schwarzschild's outer solution. The calculated factors $A(r)$ and $B(r)$ therefore apply both inside and outside the spherical mass of radius R .

3.4 Black Holes

3.4.1 Astrophysics

The three most interesting celestial objects are *white dwarfs*, *neutron stars* and *black holes*. Stars like the Sun are built up by compression of interstellar clouds, caused by gravity. The contraction process comes to a standstill when the temperature inside becomes so great that nuclear fusion starts. By the nuclear fusion, hydrogen is converted into helium. The further fate of a star depends on its initial mass: The initial mass equal to

- ≈ 0.05 solar mass ends up as a *brown dwarf*;
- ≈ 1 solar mass and of volume comparable to that of the Earth grows into a red giant and ends after an explosion (supernova) as a *white dwarf*;
- ≈ 3 solar masses, with a corresponding radius of about 12 km, grows into a supergiant and ends after explosion (supernova) as a *neutron star*;
- ≈ 30 solar masses grows into a supergiant and ends after explosion (supernova) as a *black hole*.

A white dwarf is composed mostly of carbon and oxygen, produced by nuclear fusion, that are left over when the nuclear solar fuel is consumed. A white dwarf has a mass about that of our Sun and the diameter of the Earth, so a fairly compact structure. However, a neutron star is much more compact, i.e. it has a factor of 10^9 higher density. It has roughly the mass of a white dwarf but a diameter of only 24 km! A black hole, on the other hand, has different masses, sizes, and densities, as is further explained below.

The Schwarzschild's solution (3.26)

$$ds^2 = \left(1 - \frac{r_S}{r}\right)c^2 dt^2 - \left(1 - \frac{r_S}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta d\varphi^2)$$

applies only outside the considered spherical mass. This formula gives the tiny spacetime interval ds which mass elements cover, moving from an event A to a closely adjacent event B. The Schwarzschild's radius is therefore only interesting for very large masses or masses with sufficient density, where $r_S > R$ is the radius of the spherical mass.

However, for the solar mass $M_\odot \approx 2 \cdot 10^{30}$ kg the Schwarzschild's radius is

$$r_{S\odot} \approx 3 \text{ km},$$

i.e. it is much smaller than the solar radius $r_\odot \approx 7 \cdot 10^5$ km, so our solution is valid only outside the mass of the Sun! For the Earth with a mass of $M_\oplus \approx 6 \cdot 10^{24}$ kg we even obtain a Schwarzschild's radius of

$$r_{S\oplus} \approx 9 \text{ mm!}$$

It seems that the Schwarzschild's radius only plays a role for highly concentrated matter. But that is not the case. Indeed, consider just the radius of a spherical mass $R = r_S$, that is, $R = \frac{2GM}{c^2}$. The so-called Schwarzschild's *density* ρ_S would in this case be

$$\begin{aligned} \underline{\underline{\rho_S}} &\stackrel{\text{def}}{=} \frac{M}{\frac{4}{3}\pi r_S^3} = \frac{M}{\frac{4}{3}\pi(\frac{2GM}{c^2})^3} \\ &= \frac{3c^6}{32\pi G^3 M^2} = 2.33 \cdot 10^{71} \cdot M_{[\text{kg}]}^{-2} \left[\frac{\text{kg}}{\text{cm}^3} \right]. \end{aligned}$$

The required density decreases inversely proportionally with the square of the mass! If one wants to do a black hole from Earth's mass, i.e. shrink the Earth to a radius

of 9 mm, then the resulting body would have a density of $8.58 \cdot 10^{22} \text{ kg/cm}^3$. For the Sun a density of approximately $8.6 \cdot 10^{10} \text{ kg/cm}^3$ would be necessary; that is, 86 million tons per cubic centimetre! So still the enormous density of the neutron liquid¹ is in a neutron star. In the centre of the Milky Way, there is a black hole with a mass of $5.2 \cdot 10^{36} \text{ kg}$, or 2.6 million solar masses, and in the centre of the Virgo galaxy cluster, a black hole with the mass of $6 \cdot 10^{39} \text{ kg}$ (3 billion solar masses).

3.4.2 Further Details about “Black Holes”

A spherical mass of radius less than the Schwarzschild’s radius, $R < r_S$, is called a *black hole*. The name is explained as follows. The Schwarzschild’s solution (3.26) for constant angles θ and φ , i.e. for $d\theta = d\varphi = 0$, with $r_S = \frac{2GM}{c^2}$, is

$$ds^2 = \left(1 - \frac{r_S}{r}\right) c^2 dt^2 - \left(1 - \frac{r_S}{r}\right)^{-1} dr^2. \quad (3.80)$$

For light, i.e. photons, one has $ds^2 = 0$, so from (3.80) follows

$$c^2 dt^2 = \left(1 - \frac{r_S}{r}\right)^{-2} dr^2,$$

giving

$$c dt = \pm \left(1 - \frac{r_S}{r}\right)^{-1} dr = \pm \frac{r}{r - r_S} dr,$$

which integrated yields

$$c \int_{t_0}^t dt = ct - ct_0 = \int_{r_0}^r \frac{r}{r - r_S} dr = \pm \left[r + r_S \ln \frac{r - r_S}{r_0 - r_S} - r_0 \right]$$

and with $c_0 \stackrel{\text{def}}{=} ct_0 - r_0 - r_S \ln(r_0 - r_S)$ finally

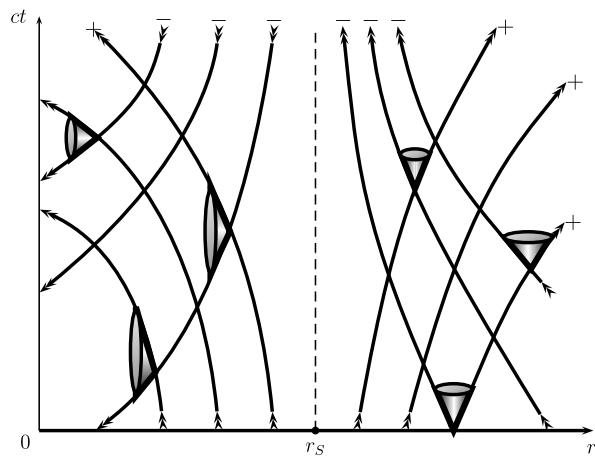
$$\underline{\underline{ct = \pm(r + r_S \ln(r - r_S) + c_0)}}, \quad (3.81)$$

where the constant c_0 depends on the start time t_0 and on the starting location R_0 , and ct has the dimension of a length. The substance of this formula can be pictured in the $(ct - r)$ -half-plane, as shown in Fig. 3.1.

The light cones are to the right of the dividing line at $r = r_S$ upward open towards the time axis, i.e. particles (including photons) in the picture fly only upwards. In particular, even photons on the minus-sign trajectories seemingly not exceed the

¹Neutron liquid consists predominantly of neutrons, and their average density is approximately equal to that of atomic nuclei.

Fig. 3.1 Schwarzschild's solution



dividing line $r = r_S$, so can never arrive at the field with $r < r_S$. This is particularly recognised at the speed of the particles or photons, resulting from the above equations in

$$\frac{dr}{dt} = \pm \left(1 - \frac{r_S}{r}\right)c.$$

This velocity goes to zero as r goes to r_S . For $r < r_S$ the light cone shows the future in the direction $r = 0$. No particle or photon can escape from the field $r < r_S$! No light can overstep the limit of r_S ; the “black hole” is, in fact, “black”! For a particle at r_S as is in Fig. 3.1, the time t decreases, so runs backwards! It is obvious that for this region, the time t is not very suitable. On the other hand, it is a fallacy that the boundary $r = r_S$ cannot be exceeded from the outside because the formulas report only the behaviour which would be seen by an observer. An observer, flying with the particle, would normally exceed the limit $r = r_S$ because for him the speed in the vicinity $r = r_S$ would not tend to zero, he would fly continuously through this sphere with finite speed. This can be shown as follows:

Due to Newton's theory, we arrive at the following context. Integrating the equation

$$m_0 a(t) = m_0 \frac{dv}{dt} = G \frac{M m_0}{r^2}$$

with respect to

$$a = \frac{dv}{dt} = \frac{dv}{dr} \frac{dr}{dt} = v \frac{dv}{dr},$$

one obtains

$$\int_0^v \tilde{v} d\tilde{v} = \frac{v^2}{2} = \int_{\infty}^r a d\tilde{r} = \int_{\infty}^r \frac{GM}{\tilde{r}^2} d\tilde{r} = \frac{GM}{r},$$

or solving for the velocity

$$v = -\sqrt{\frac{2GM}{r}} = -c\sqrt{\frac{r_S}{r}}. \quad (3.82)$$

The mass m_0 would pass the Schwarzschild's spherical shell for $r = r_S$ and according to (3.82) with the speed of light c , which is not possible according to the special theory of relativity! For $r < r_S$ the velocity would even be greater than light's! The velocity equations have to be modified. What happens then? If one passes through the Schwarzschild's spherical shell in the direction $r \rightarrow 0$, we will never be able to learn from him what he experienced because neither he nor a signal from him to us outside the Schwarzschild's spherical shell will ever reach us. We only can analyse theoretically what happens.

Within the Schwarzschild's Radius From (3.82) for the proper time we have

$$d\tau = -\frac{1}{c} \left(\frac{r}{r_S} \right)^{1/2} dr. \quad (3.83)$$

From this we get the time τ needed when flying through the Schwarzschild's spherical shell to reach the singularity $r = 0$. It is

$$\tau_S \stackrel{\text{def}}{=} -\frac{1}{c} \int_{r_S}^0 \left(\frac{r}{r_S} \right)^{1/2} dr = \frac{2}{3} \frac{r_S}{c}, \quad (3.84)$$

therefore, a finite time.

If a particle is located to the left of the parting line $r = r_S$, then the particle flies left in the direction $r \rightarrow 0$. Photons, starting left of the parting line cannot also exceed this border. Therefore, no light (photon) is coming outward from inside the sphere of radius r_S , a “black hole” is given there. Further, it is interesting that an ensemble of boundary trajectories (each characterized with a minus sign) proceeds in the negative direction of time, thus decreasing time. If one starts, for instance, at the point $(t_0 = 0, r = r_0 < r_S)$ on a ‘minus-trajectory’, then one travels to the past! The result, however, contains no logical contradiction because one cannot act from the inside of the sphere with the radius r_S on the outside: one only passively receives signals, i.e. one can “look into the past”.

3.4.3 Singularities

The solution (3.26) has two singularities: a real singularity at $r = 0$ and an apparent one at $r = r_S$. The apparent singularity is a so-called *coordinate singularity* which could have been avoided by a better coordinate choice. Such a favourable coordinate

system would have been obtained, for example, if we had introduced a different radial coordinate r^* as follows:

$$r = \left(1 + \frac{r_S}{4r^*}\right)^2 r^*.$$

Then

$$\frac{dr}{dr^*} = \left(1 - \frac{r_S}{4r^*}\right) \left(1 + \frac{r_S}{4r^*}\right),$$

so

$$dr = \left(1 - \frac{r_S}{4r^*}\right) \left(1 + \frac{r_S}{4r^*}\right) dr^*.$$

Used in (3.26) this yields

$$ds^2 = \left(\frac{1 - \frac{r_S}{4r^*}}{1 + \frac{r_S}{4r^*}}\right)^2 c^2 dt^2 - \left(1 + \frac{r_S}{4r^*}\right)^4 (dr^{*2} + r^{*2}(d\theta^2 + \sin^2 \theta d\varphi^2)).$$

In these coordinates, in fact, only a singularity at $r^* = 0$ exists!

The authenticity of the singularity at $r = 0$ can be seen by examining the invariants. A function of coordinates is invariant under a transformation if it remains unchanged applying the transformation to the coordinates. This is precisely the mark of a true singularity, which does not depend on the randomly selected coordinate system. Such invariants were already encountered in the investigation of electromagnetic fields and the Lorentz transformation, e.g.

$$-\frac{1}{2} \text{trace}(\mathbf{F}_{B,e}^* \mathbf{F}_{B,e}) = b^2 - e^2 = b'^2 - e'^2$$

and

$$-\frac{1}{4} \text{trace}(\mathbf{F}_{B,e}^* \mathbf{F}_{E,b}) = \mathbf{e}^\top \mathbf{b} = \mathbf{e}'^\top \mathbf{b}'.$$

Such an invariant for our problem now is the so-called Kretschmann's invariant, which is defined with the aid of the modified symmetric Riemannian curvature matrix

$$\mathbf{R}^* \stackrel{\text{def}}{=} (\mathbf{I}_4 \otimes \mathbf{G}^{-1}) \mathbf{R}$$

as follows:

$$I_K \stackrel{\text{def}}{=} \text{trace}(\mathbf{R}^* \mathbf{R}^*). \quad (3.85)$$

Because the Ricci matrix \mathbf{R}_{Ric} equals the zero matrix, it is unsuitable for an invariant. The newly defined matrix \mathbf{R}^* will now be derived step by step for the

Schwarzschild's metric. First, we have

$$\boldsymbol{F} = \begin{pmatrix} 0 & \frac{m}{r^2 h} & 0 & 0 \\ \frac{m}{r^2 h} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{mh}{r^2} & 0 & 0 & 0 \\ 0 & -\frac{m}{r^2 h} & 0 & 0 \\ 0 & 0 & -rh & 0 \\ 0 & 0 & 0 & -rh \sin^2 \theta \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{r} & 0 \\ 0 & \frac{1}{r} & 0 & 0 \\ 0 & 0 & 0 & -\sin \theta \cos \theta \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{r} \\ 0 & 0 & 0 & \cot \theta \\ 0 & \frac{1}{r} & \cot \theta & 0 \end{pmatrix},$$

where

$$m \stackrel{\text{def}}{=} \frac{GM}{c^2} \quad \text{and} \quad h \stackrel{\text{def}}{=} 1 - \frac{2m}{r}.$$

With the help of (2.165) \boldsymbol{R} can be calculated. The result is

$$\boldsymbol{R} = \left(\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{2m}{r^3 h} & 0 & 0 & \frac{2m}{r^3 h} & 0 & 0 & 0 \\ 0 & 0 & \frac{m}{r} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{m \sin^2 \theta}{r} & 0 & 0 & 0 & 0 \\ \hline 0 & -\frac{2mh}{r^3} & 0 & 0 & \frac{2mh}{r^3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{m}{r} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{m \sin^2 \theta}{r} \\ \hline 0 & 0 & \frac{mh}{r^3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{m}{r^3 h} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \frac{mh}{r^3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{m}{r^3 h} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

$$\left(\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{m}{r} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{m \sin^2 \theta}{r} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{m}{r} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{m \sin^2 \theta}{r} & 0 & 0 \\ \hline -\frac{mh}{r^3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{m}{r^3 h} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{2m \sin^2 \theta}{r} & 0 & 0 & \frac{2m \sin^2 \theta}{r} & 0 \\ \hline 0 & 0 & 0 & 0 & -\frac{mh}{r^3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{m}{r^3 h} & 0 & 0 \\ 0 & 0 & 0 & \frac{2m}{r} & 0 & 0 & -\frac{2m}{r} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) .$$

The first, $(4 + 2)$ th, $(8 + 3)$ th and 16th row/column are zero rows/columns, as it must always be in the matrix \mathbf{R} . With \mathbf{R} the following symmetric matrix is obtained

$$(\mathbf{I}_4 \otimes \mathbf{G}^{-1}) \mathbf{R}$$

$$= \left(\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2m}{r^3} & 0 & 0 & -\frac{2m}{r^3} & 0 & 0 & 0 \\ 0 & 0 & -\frac{m}{r^3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{m}{r^3} & 0 & 0 & 0 & 0 \\ \hline 0 & -\frac{2m}{r^3} & 0 & 0 & \frac{2m}{r^3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{m}{r^3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{m}{r^3} \\ \hline 0 & 0 & \frac{m}{r^3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{m}{r^3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \frac{m}{r^3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{m}{r^3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

$$\left(\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{m}{r^3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{m}{r^3} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{m}{r^3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{m}{r^3} & 0 & 0 \\ \hline -\frac{m}{r^3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{m}{r^3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2m}{r^3} & 0 & 0 & -\frac{2m}{r^3} & 0 \\ \hline 0 & 0 & 0 & 0 & -\frac{m}{r^3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{m}{r^3} & 0 & 0 \\ 0 & 0 & 0 & -\frac{2m}{r^3} & 0 & 0 & \frac{2m}{r^3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right),$$

and finally the Kretschmann's invariant is

$$I_K = \text{trace}(\mathbf{R}^* \mathbf{R}^*) = 48 \frac{m^2}{r^6} = 12 \frac{r_S^2}{r^6}.$$

Hence, it is now quite apparent that the only real singularity occurs at $r = 0$!

Event Horizon Detector In [13], a possibility is given, how to determine if one approaches the event horizon (the spherical shell of radius r_S) of a black hole or even exceeds it. To this end, the authors give the invariant

$$I_1 = -\frac{720M^2(2M-r)}{r^9},$$

which for $r = 2M = r_S$ is zero and outside the horizon is positive with a maximum at $r = 9M/4$. An observer falling to the black hole can detect the presence of a horizon by observation of I_1 . If the event horizon is crossed, it is too late for the observer. But he may, if the maximum is exceeded, use it as a warning and quickly initiate the trajectory reversal.

3.4.4 Eddington's Coordinates

The coordinate singularity at $r = r_S$ of the Schwarzschild's metric can be eliminated, for example, by using the following coordinate transformation.

In (3.81), from the Schwarzschild's solution we derived

$$ct = \pm(r + r_S \ln|r - r_S| + c_0).$$

If we define

$$r^* \stackrel{\text{def}}{=} r_S \ln(r - r_S), \quad (3.86)$$

we get from

$$\frac{dr^*}{dr} = \frac{r_S}{r - r_S}$$

the differential

$$dr^* = \frac{r_S}{r} \frac{r}{r - r_S} dr = \frac{r_S}{r} \left(\frac{1}{1 - \frac{r_S}{r}} \right)^{-1} dr. \quad (3.87)$$

It looks like the term

$$-\left(\frac{1}{1 - \frac{r_S}{r}} \right)^{-1} dr^2$$

in the Schwarzschild's metric, which was the reason for the coordinate singularity! If (3.87) were squared, indeed dr^2 would appear, but the parenthesis would have too much negative power. If one installs, however, the term r^* into a new time coordinate t^* , then within the Schwarzschild's metric it would again be multiplied by $\frac{r}{r - r_S}$, i.e. have the “right” power! Now the new time coordinate is set as

$$ct^* \stackrel{\text{def}}{=} ct + r^* = ct + r_S \ln |r - r_S|. \quad (3.88)$$

Differentiation with respect to r gives

$$c \frac{dt^*}{dr} = c \frac{dt}{dr} + \frac{dr^*}{dr}, \quad (3.89)$$

so

$$c dt^* = c dt + dr^* = c dt + \frac{r_S}{r - r_S} dr,$$

i.e.

$$c dt = c dt^* - \frac{r_S}{r - r_S} dr,$$

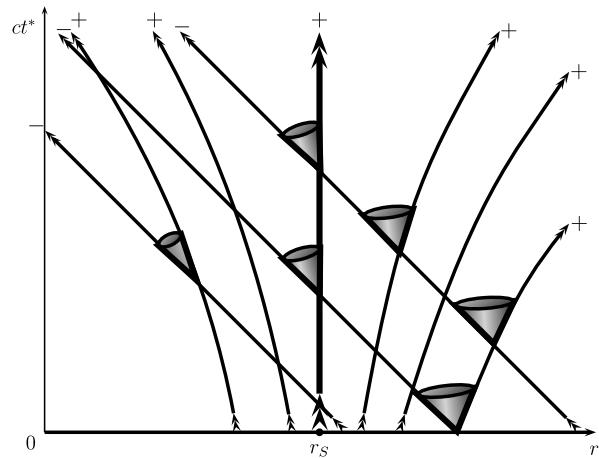
and squared

$$c^2 dt^2 = c^2 dt^{*2} - 2 \frac{r_S}{r - r_S} c dt^* dr + \left(\frac{r_S}{r - r_S} \right)^2 dr^2.$$

Used in the Schwarzschild's metric this yields

$$ds^2 = \left(1 - \frac{r_S}{r} \right) c^2 dt^{*2} - 2 \frac{r_S}{r} c dt^* dr + \frac{r - r_S}{r} \left(\frac{r_S}{r - r_S} \right)^2 dr^2 - \left(\frac{r}{r - r_S} \right) dr^2,$$

Fig. 3.2 Eddington's solution



so finally,

$$\underline{\underline{ds^2 = \frac{r - r_S}{r} c^2 dt^{*2} - 2\frac{r_S}{r} c dt^* dr - \frac{r + r_S}{r} dr^2.}} \quad (3.90)$$

When we remove the restriction $d\theta = d\varphi = 0$, we finally get the full Schwarzschild's metric in Eddington's coordinates

$$\underline{\underline{ds^2 = \frac{r - r_S}{r} c^2 dt^{*2} - 2\frac{r_S}{r} c dt^* dr - \frac{r + r_S}{r} dr^2 - r^2(d\theta^2 + \sin^2\theta d\varphi^2).}} \quad (3.91)$$

This metric has, in fact, a singularity only at $r = 0$!

How will we see movements in the new (ct^*, r) -plane? For photons one has $ds^2 = 0$; so, if we divide (3.90) by $c^2 dt^{*2}$, we receive

$$0 = \frac{r - r_S}{r} - 2\frac{r_S}{r} \frac{dr}{c dt^*} - \frac{r + r_S}{r} \left(\frac{dr}{c dt^*} \right)^2.$$

This quadratic equation has the solutions

$$\frac{dr}{c dt^*} = \frac{-r_S \pm r}{r + r_S} = \frac{r - r_S}{r + r_S} \text{ and } -1, \quad (3.92)$$

or

$$\frac{dr}{dt^*} = \frac{-r_S \pm r}{r + r_S} c = \frac{r - r_S}{r + r_S} c \text{ and } -c. \quad (3.93)$$

The first solution provides for $r = r_S$ the zero slope, i.e. a vertical line; for $r < r_S$ the slope is negative, and it is positive for $r > r_S$. The second solution in the (ct^*, r) -plane is an inclined at 45° straight line. It shows that the velocity of an ingoing photon is constant and equal to c . Overall, one obtains Fig. 3.2.

3.5 Rotating Masses

3.5.1 Ansatz for the Metric Matrix \mathbf{G}

From a non-rotating mass of radius R generated gravitation effect on the outside, i.e. for $r > R$, is described by the Schwarzschild's metric. A non-rigid rotating mass bulges out along its equator, so it cannot be described by a Schwarzschild's metric, which assumes a symmetric spherical mass. A rotating mass has in addition to the mass M also an angular momentum J , which is directly proportional to the angular velocity ω . M and J are the only physical quantities for a rotating black hole that are required for the physical description of the gravitational field and how powerful the rotating mass is (there are possibly millions of solar masses in a black hole in the centre of a galaxy)! The Nobel-prize laureate Chandrasekhar has expressed it as:

Rotating black holes are the most perfect macroscopic objects in the universe. And as the theory of general relativity provides a single, unique, two-parameter solution for their description, they are also the simplest objects.

For the mathematical description, spatial polar coordinates r, θ and φ are selected. The polar axis is the axis of rotation around which the body rotates with constant angular velocity ω . The elements of the metric matrix \mathbf{G} are allowed to depend neither on the angle θ nor on time t . One can also use a priori, for reasons of symmetry, some matrix elements equal to zero. Because if one takes a time reflection $t \rightarrow -t$, this also changes the direction of rotation $\omega = \frac{d\varphi}{dt} \rightarrow -\omega = -\frac{d\varphi}{dt} = \frac{d(-\varphi)}{dt}$. But if we introduce two transformations $t \rightarrow -t$ and $\varphi \rightarrow -\varphi$ simultaneously, then the gravitational field is not allowed to change at all, also the metric matrix \mathbf{G} is not changed. So that being the case, the matrix elements $g_{t\theta}$ and g_{tr} must be zero because in the calculation of the metric here—with the indicated transformations— $dt d\theta \rightarrow -dt d\theta$ and $dt dr \rightarrow -dt dr$. And so one can start from the following the metric matrix which must always be symmetric:

$$\mathbf{G} = \begin{pmatrix} g_{tt} & 0 & 0 & g_{t\varphi} \\ 0 & g_{rr} & g_{r\theta} & 0 \\ 0 & g_{r\theta} & g_{\theta\theta} & 0 \\ g_{t\varphi} & 0 & 0 & g_{\varphi\varphi} \end{pmatrix}.$$

The symmetric 2×2 -matrix in the centre of the matrix \mathbf{G} can be brought to diagonal form with the help of a similarity transformation so that, without loss of generality, one may start with the symmetric matrix:

$$\mathbf{G} = \begin{pmatrix} g_{tt} & 0 & 0 & g_{t\varphi} \\ 0 & g_{rr} & 0 & 0 \\ 0 & 0 & g_{\theta\theta} & 0 \\ g_{t\varphi} & 0 & 0 & g_{\varphi\varphi} \end{pmatrix}.$$

3.5.2 Kerr's Solution in Boyer–Lindquist Coordinates

Without derivation we here directly give as solution the so-called Kerr's metric in Boyer–Lindquist coordinates:

$$\begin{aligned} ds^2 = & \left(1 - \frac{2mr}{\rho^2} \right) c^2 dt^2 + 4ma \frac{r \sin^2 \theta}{\rho^2} c dt d\varphi - \frac{\rho^2}{\Delta} dr^2 \\ & \underline{\underline{- \rho^2 d\theta^2 - \left(r^2 + a^2 + \frac{2mr}{r^2} a^2 \sin^2 \theta \right) \sin^2 \theta d\varphi^2}}, \end{aligned} \quad (3.94)$$

with

$$\Delta \stackrel{\text{def}}{=} r^2 - 2mr + a^2,$$

$$\rho^2 \stackrel{\text{def}}{=} r^2 + a^2 \cos^2 \theta,$$

$$m \stackrel{\text{def}}{=} \frac{MG}{c^2}$$

and the angular momentum (spin) J per unit mass

$$a \stackrel{\text{def}}{=} \frac{J}{m}.$$

As system parameters, in fact, only the two physical parameters occur: the modified mass, m , and the *angular momentum per unit mass*, a ! For $J = 0$, i.e. for a non-rotating mass, one obtains, of course, the Schwarzschild's solution.

A simpler form of the solution is obtained for $M/r \ll 1$ and $a/r \ll 1$, i.e. for weak fields and slow rotation, namely

$$\begin{aligned} ds^2 \cong & \left(1 - \frac{2m}{r} \right) c^2 dt^2 + \frac{4J}{r} \sin^2 \theta c dt d\varphi - \left(1 + \frac{2m}{r} \right) dr^2 \\ & - r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \end{aligned}$$

3.5.3 The Lense–Thirring Effect

Two electric charges q_1 and q_2 of opposite polarity attract according to the Coulomb's law with the force

$$\mathbf{f}_{\text{electr}} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}}.$$

Here the vector $\mathbf{r}_{12} \in \mathbb{R}^3$ shows from q_1 to q_2 . Newton's law of attraction between two masses m_1 and m_2 has almost the same shape, namely

$$\mathbf{f}_{\text{mech}} = G \frac{m_1 m_2}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}}.$$

Going over in both laws from the action at a distance theory to the theory of contiguous action, i.e. the field theory, with the electric field strength vector

$$\mathbf{e}_e \stackrel{\text{def}}{=} \frac{1}{4\pi\epsilon_0} \frac{q_1}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}}$$

one can also write the electric force as

$$\mathbf{f}_{\text{electr}} = \mathbf{e}_e q_2.$$

With the appropriate nomenclature for the mechanical parameters and the field strength

$$\mathbf{e}_m \stackrel{\text{def}}{=} G \frac{m_1}{r_{12}^2} \frac{\mathbf{r}_{12}}{r_{12}},$$

one finally obtains

$$\mathbf{f}_{\text{mech}} = \mathbf{e}_m m_2.$$

If an electric charge is moving, then there is in addition a magnetic field \mathbf{b} , which depends on the speed and the load of the charge. The magnetic field \mathbf{b} acts on the moving charge q having speed \mathbf{v} , generating, together with the electric field \mathbf{e} , the force:

$$\mathbf{f} = q \left(\mathbf{e} + \frac{1}{c} \mathbf{v} \times \mathbf{b} \right).$$

The question now is: Does a moving *mass* have a similar additional effect to another moving mass? This is indeed the case, for example, the precession of gyroscopes in the vicinity of large rotating masses like the Earth. This first was treated by Föppl [11]. The Austrian physicists Lense and Thirring calculated exactly this effect from the gravitational equations of Einstein in 1918.

One speaks in this context of a so-called *gravitomagnetic* field in analogy with classical electromagnetism. To this end, we first define with the help of the angular momentum \mathbf{J} of a spherical rotating mass the field

$$\mathbf{h}(\mathbf{r}) \stackrel{\text{def}}{=} -2 \frac{\mathbf{J} \times \mathbf{r}}{r^3} = -\frac{4GM R^2}{5c^3} \frac{\boldsymbol{\omega} \times \mathbf{r}}{r^3}, \quad (3.95)$$

where $\boldsymbol{\omega}$ is the angular velocity of the rotating mass M with the radius R , and $r > R$ is the distance from the centre of mass. Next we define the *gravitomagnetic field*:

$$\bar{\mathbf{h}} \stackrel{\text{def}}{=} \nabla \times \mathbf{h} = 2 \frac{\mathbf{J} - 3(\mathbf{r}^\top \mathbf{J})\mathbf{r}}{r^3} = \frac{2GM R^2}{5c^2} \frac{3(\boldsymbol{\omega}^\top \mathbf{r})\mathbf{r} - \boldsymbol{\omega} r^2}{r^5}. \quad (3.96)$$

The angular momentum here thus plays the same role as the magnetic dipole moment in electrodynamics, and the vector \mathbf{h} plays the same role as the vector potential. In summary, we obtain an analogue of the Lorentz force, namely

$$\underline{\underline{m \frac{d^2 \mathbf{x}}{dt^2} = m \left(\mathbf{e}_{\text{mech}} + \frac{d\mathbf{x}}{dt} \times \bar{\mathbf{h}} \right)}}. \quad (3.97)$$

In the case of stars orbiting close to a spinning, supermassive black hole, the Lense-Thirring effect should cause the star's orbital plane to precess about the black hole's spin axis. This effect should be detectable within the next few years via astrometric monitoring of stars at the centre of the Milky Way galaxy.

3.6 Summary of Results for the Gravitation of a Spherical Mass

The solution of Einstein's field equation for the *outside* of a spherically symmetric, uniform, time-invariant mass distribution is this Schwarzschild's metric (3.26):

$$\boxed{ds^2 = \left(1 - \frac{2GM}{c^2 r}\right) c^2 dt^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2 \theta d\varphi^2),}$$

and the so-called Schwarzschild's radius is (3.28):

$$\boxed{r_s \stackrel{\text{def}}{=} \frac{2GM}{c^2}.}$$

The Schwarzschild's metric in matrix form is (3.25):

$$\boxed{ds^2 = d\vec{x}^\top \mathbf{G} d\vec{x} = d\vec{x}^\top \begin{pmatrix} 1 - \frac{r_s}{r} & 0 & 0 & 0 \\ 0 & -(1 - \frac{r_s}{r})^{-1} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2 \sin^2 \theta \end{pmatrix} d\vec{x},}$$

with

$$\boxed{d\vec{x} \stackrel{\text{def}}{=} \begin{pmatrix} c dt \\ dr \\ d\theta \\ d\varphi \end{pmatrix}.}$$

This changes the length (3.30) and time (3.31) according to:

$$\boxed{dR = \left(1 - \frac{2GM}{c^2 r}\right)^{-1/2} dr,}$$

and

$$d\tau = \left(1 - \frac{2GM}{c^2 r} \right)^{1/2} dt.$$

For the redshift of spectral lines the frequency ratio is obtained (3.32):

$$\frac{\nu_R}{\nu_T} = \left(\frac{1 - \frac{2GM}{c^2 r_T}}{1 - \frac{2GM}{c^2 r_R}} \right)^{1/2}.$$

The necessary Schwarzschild's density ρ_S for the existence of a *black hole* is

$$\begin{aligned} \rho_S &\stackrel{\text{def}}{=} \frac{M}{\frac{4}{3}\pi r_S^3} = \frac{M}{\frac{4}{3}\pi (\frac{2GM}{c^2})^3} \\ &= \frac{3c^6}{32\pi G^3 M^2} = 2.33 \cdot 10^{71} \cdot M_{[\text{kg}]} \left[\frac{\text{kg}}{\text{cm}^3} \right]. \end{aligned}$$

The Eddington's coordinates for a black hole are given with the new time coordinate (3.88)

$$ct^* \stackrel{\text{def}}{=} ct + r_S \ln |r - r_S|,$$

so

$$c dt^* = c dt + \frac{r_S}{r - r_S} dr.$$

The Schwarzschild's metric in Eddington's coordinates is (3.91):

$$ds^2 = \frac{r - r_S}{r} c^2 dt^{*2} - 2 \frac{r_S}{r} c dt^* dr - \frac{r + r_S}{r} dr^2 - r^2 (d\theta^2 + \sin^2 \theta d\varphi^2).$$

The Lense–Thirring effect for a rotating mass generates a *gravitomagnetic* field in analogy with classical electromagnetism. With the help of the angular momentum \mathbf{J} of a spherical rotating mass, the field is (3.95):

$$\mathbf{h}(\mathbf{r}) \stackrel{\text{def}}{=} -2 \frac{\mathbf{J} \times \mathbf{r}}{r^3} = -\frac{4GMR^2}{5c^3} \frac{\boldsymbol{\omega} \times \mathbf{r}}{r^3},$$

where $\boldsymbol{\omega}$ is the angular velocity of the rotating mass M with the radius R , and $r > R$ is the distance from the centre of mass. Then the *gravitomagnetic field* is defined as (3.96):

$$\bar{\mathbf{h}} \stackrel{\text{def}}{=} \nabla \times \mathbf{h} = 2 \frac{\mathbf{J} - 3(\mathbf{r}^\top \mathbf{J})\mathbf{r}}{r^3} = \frac{2GMR^2}{5c^2} \frac{3(\boldsymbol{\omega}^\top \mathbf{r})\mathbf{r} - \boldsymbol{\omega}r^2}{r^5}.$$

An analogue of the Lorentz force is obtained (3.97):

$$m \frac{d^2\mathbf{x}}{dt^2} = m \left(\mathbf{e}_{\text{mech}} + \frac{d\mathbf{x}}{dt} \times \bar{\mathbf{h}} \right).$$

3.7 Concluding Remark

In the theory of *Special Relativity*, the effects are most clearly visible when the masses move quickly; however, in the theory of *General Relativity*, the effects are the greatest when the mass densities are very large and thus the spatial curvature is very pronounced.

Appendix A

Vectors and Matrices

A.1 Vectors and Matrices

If the speed of a body is given, then its size and the direction need to be identified. For the description of such a directional quantity, *vectors* are used. These vectors in the three dimensional space require three components which, e.g. in a *column vector*, are summarized as follows:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}. \quad (\text{A.1})$$

A second possibility is to present a *transposed* column vector

$$\mathbf{v}^T = (v_1 \quad v_2 \quad v_3), \quad (\text{A.2})$$

a *row vector*.

In another way, we get the concept of the vector when the following purely mathematical problem is considered: Find the solutions of the three coupled equations with four unknowns x_1 , x_2 , x_3 and x_4 :

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = y_1, \quad (\text{A.3})$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = y_2, \quad (\text{A.4})$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 = y_3. \quad (\text{A.5})$$

The four unknowns are summarized in the vector

$$\mathbf{x} \stackrel{\text{def}}{=} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \quad (\text{A.6})$$

the variables y_1 , y_2 and y_3 form the vector

$$\mathbf{y} \stackrel{\text{def}}{=} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad (\text{A.7})$$

and the coefficients a_{ij} are included into the *matrix*

$$\mathbf{A} \stackrel{\text{def}}{=} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}. \quad (\text{A.8})$$

With the two vectors \mathbf{x} and \mathbf{y} and the matrix \mathbf{A} , the system of equations can compactly be written as

$$\mathbf{Ax} = \mathbf{y}. \quad (\text{A.9})$$

If the two systems of equations

$$a_{11}x_1 + a_{12}x_2 = y_1, \quad (\text{A.10})$$

$$a_{21}x_1 + a_{22}x_2 = y_2 \quad (\text{A.11})$$

and

$$a_{11}z_1 + a_{12}z_2 = v_1, \quad (\text{A.12})$$

$$a_{21}z_1 + a_{22}z_2 = v_2 \quad (\text{A.13})$$

are added, one obtains

$$a_{11}(x_1 + z_1) + a_{12}(x_2 + z_2) = (y_1 + v_1), \quad (\text{A.14})$$

$$a_{21}(x_1 + z_1) + a_{22}(x_2 + z_2) = (y_2 + v_2). \quad (\text{A.15})$$

With the aid of vectors and matrices, the two systems of equations can be written as

$$\mathbf{Ax} = \mathbf{y} \quad \text{and} \quad \mathbf{Az} = \mathbf{v}. \quad (\text{A.16})$$

Adding the two equations in (A.16) is formally accomplished as

$$\mathbf{Ax} + \mathbf{Az} = \mathbf{A}(\mathbf{x} + \mathbf{z}) = \mathbf{y} + \mathbf{v}. \quad (\text{A.17})$$

A comparison of (A.17) with (A.14) and (A.15) suggests the following definition of the *addition of vectors*:

Definition:

$$\mathbf{y} + \mathbf{v} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} y_1 + v_1 \\ y_2 + v_2 \\ \vdots \\ y_n + v_n \end{pmatrix}. \quad (\text{A.18})$$

Accordingly, the product of a vector and a real or complex number c is defined by

Definition:

$$c \cdot \mathbf{x} \stackrel{\text{def}}{=} \begin{pmatrix} c \cdot x_1 \\ \vdots \\ c \cdot x_n \end{pmatrix}. \quad (\text{A.19})$$

A.2 Matrices

A.2.1 Types of Matrices

In the first section, the concept of the matrix has been introduced.

Definition: If a matrix \mathbf{A} has n rows and m columns, it is called an $n \times m$ matrix and denoted $\mathbf{A} \in \mathbb{R}^{n \times m}$.

Definition: If \mathbf{A} (with the elements a_{ij}) is an $n \times m$ matrix, then the *transpose* of \mathbf{A} , denoted by \mathbf{A}^T , is the $m \times n$ matrix with the elements $a_{ij}^T = a_{ji}$.

So the matrix (A.8) has the matrix transpose

$$\mathbf{A}^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \\ a_{14} & a_{24} & a_{34} \end{pmatrix}. \quad (\text{A.20})$$

In a *square matrix*, one has $n = m$; and in an $n \times n$ *diagonal matrix*, all the elements a_{ij} , $i \neq j$, outside the main diagonal are equal to zero. An *identity matrix* \mathbf{I} is a diagonal matrix where all elements on the main diagonal are equal to one. An $r \times r$ identity matrix is denoted by \mathbf{I}_r . If a transposed matrix \mathbf{A}^T is equal to the original matrix \mathbf{A} , such a matrix is called *symmetric*. In this case, $a_{ij} = a_{ji}$.

A.2.2 Matrix Operations

If the two systems of equations

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = y_1,$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = y_2,$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = y_n$$

and

$$b_{11}x_1 + b_{12}x_2 + \cdots + b_{1m}x_m = z_1,$$

$$b_{21}x_1 + b_{22}x_2 + \cdots + b_{2m}x_m = z_2,$$

$$\vdots$$

$$b_{n1}x_1 + b_{n2}x_2 + \cdots + b_{nm}x_m = z_n$$

are added, one obtains

$$(a_{11} + b_{11})x_1 + (a_{12} + b_{12})x_2 + \cdots + (a_{1m} + b_{1m})x_m = (y_1 + z_1),$$

$$(a_{21} + b_{21})x_1 + (a_{22} + b_{22})x_2 + \cdots + (a_{2m} + b_{2m})x_m = (y_2 + z_2),$$

$$\vdots$$

$$(a_{n1} + b_{n1})x_1 + (a_{n2} + b_{n2})x_2 + \cdots + (a_{nm} + b_{nm})x_m = (y_n + z_n),$$

or, in vector-matrix notation, with

$$\mathbf{Ax} = \mathbf{y} \quad \text{and} \quad \mathbf{Bx} = \mathbf{z},$$

the same can also be written symbolically as

$$(\mathbf{A} + \mathbf{B})\mathbf{x} = \mathbf{y} + \mathbf{z}. \quad (\text{A.21})$$

A comparison of the last equations suggests the following definition:

Definition: The sum of two $n \times m$ matrices \mathbf{A} and \mathbf{B} is defined by

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nm} \end{pmatrix} \\ &\stackrel{\text{def}}{=} \begin{pmatrix} (a_{11} + b_{11}) & \cdots & (a_{1m} + b_{1m}) \\ \vdots & & \vdots \\ (a_{n1} + b_{n1}) & \cdots & (a_{nm} + b_{nm}) \end{pmatrix}. \end{aligned} \quad (\text{A.22})$$

The sum of two matrices can only be formed when both matrices have the same number of rows and the same number of columns.

If the relations

$$\mathbf{y} = \mathbf{Ax} \quad \text{and} \quad \mathbf{x} = \mathbf{Bz} \quad (\text{A.23})$$

are given, what is the connection between the two vectors \mathbf{y} and \mathbf{z} ? One may write

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = y_1,$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = y_2,$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = y_n$$

and

$$b_{11}z_1 + b_{12}z_2 + \cdots + b_{1\ell}z_\ell = x_1,$$

$$b_{21}z_1 + b_{22}z_2 + \cdots + b_{2\ell}z_\ell = x_2,$$

$$\vdots$$

$$b_{m1}z_1 + b_{m2}z_2 + \cdots + b_{m\ell}z_\ell = x_m,$$

then one obtains, by inserting the x_i 's from the latter system of equations into the former system,

$$a_{11}(b_{11}z_1 + \cdots + b_{1\ell}z_\ell) + \cdots + a_{1m}(b_{m1}z_1 + \cdots + b_{m\ell}z_\ell) = y_1,$$

$$a_{21}(b_{11}z_1 + \cdots + b_{1\ell}z_\ell) + \cdots + a_{2m}(b_{m1}z_1 + \cdots + b_{m\ell}z_\ell) = y_2,$$

$$\vdots$$

$$a_{n1}(b_{11}z_1 + \cdots + b_{1\ell}z_\ell) + \cdots + a_{nm}(b_{m1}z_1 + \cdots + b_{m\ell}z_\ell) = y_n.$$

Combining the terms with z_i , we obtain

$$(a_{11}b_{11} + \cdots + a_{1m}b_{m1})z_1 + \cdots + (a_{11}b_{1\ell} + \cdots + a_{1m}b_{m\ell})z_\ell = y_1,$$

$$(a_{21}b_{11} + \cdots + a_{2m}b_{m1})z_1 + \cdots + (a_{21}b_{1\ell} + \cdots + a_{2m}b_{m\ell})z_\ell = y_2,$$

$$\vdots$$

$$(a_{n1}b_{11} + \cdots + a_{nm}b_{m1})z_1 + \cdots + (a_{n1}b_{1\ell} + \cdots + a_{nm}b_{m\ell})z_\ell = y_n.$$

If, on the other hand, we formally insert the right-hand side of (A.23) into the left equation, we obtain

$$\mathbf{y} = \mathbf{A}\mathbf{Bz} \stackrel{\text{def}}{=} \mathbf{Cz}. \quad (\text{A.24})$$

Definition: The *product* of an $n \times m$ matrix \mathbf{A} with an $m \times \ell$ matrix \mathbf{B} is the $n \times \ell$ matrix \mathbf{C} with the matrix elements

$$c_{ij} = \sum_{k=1}^m a_{ik} b_{kj}, \quad (\text{A.25})$$

for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, \ell$.

The element c_{ij} of the product matrix \mathbf{C} is obtained by multiplying the elements of the i th row of the first matrix \mathbf{A} with the elements of the j th column of the second matrix \mathbf{B} and adding these products. It follows that the number of columns of the first matrix must be equal to the number of rows of the second matrix, so that the matrix multiplication can be executed at all. The product matrix has as many rows as the first matrix and as many columns as the second matrix. It follows that, in general, $\mathbf{AB} \neq \mathbf{BA}$.

We get another matrix operation through the following problem. In

$$\mathbf{Ax} = \mathbf{b}, \quad (\text{A.26})$$

the 3×3 matrix \mathbf{A} and the 3×1 vector \mathbf{b} shall be given. Wanted is the 3×1 vector \mathbf{x} that satisfies the system of equations (A.26). Written out this is the linear system of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3. \end{aligned}$$

Denoting the determinant of the square matrix \mathbf{A} by $\det(\mathbf{A})$, the solutions are obtained by using Cramer's rule

$$x_1 = \frac{1}{\det(\mathbf{A})} \det \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix}, \quad (\text{A.27})$$

$$x_2 = \frac{1}{\det(\mathbf{A})} \det \begin{pmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{pmatrix}, \quad (\text{A.28})$$

$$x_3 = \frac{1}{\det(\mathbf{A})} \det \begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{pmatrix}. \quad (\text{A.29})$$

If we develop in (A.27) the determinant in the numerator with respect to the first column, we obtain

$$\begin{aligned} x_1 &= \frac{1}{\det(\mathbf{A})} \left(b_1 \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - b_2 \det \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} + b_3 \det \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix} \right) \\ &= \frac{1}{\det(\mathbf{A})} (b_1 A_{11} + b_2 A_{21} + b_3 A_{31}) \\ &= \frac{1}{\det(\mathbf{A})} (A_{11} \quad A_{21} \quad A_{31}) \mathbf{b}. \end{aligned} \quad (\text{A.30})$$

Accordingly, we obtain from (A.28) and (A.29)

$$x_2 = \frac{1}{\det(\mathbf{A})} (A_{12} \quad A_{22} \quad A_{32}) \mathbf{b} \quad (\text{A.31})$$

and

$$x_3 = \frac{1}{\det(\mathbf{A})} (A_{13} \quad A_{23} \quad A_{33}) \mathbf{b}. \quad (\text{A.32})$$

Here the *adjuncts* A_{ij} are the determinants which are obtained when the i th row and the j th column of the matrix \mathbf{A} are removed, and from the remaining matrix the determinant is computed and this is multiplied by the factor $(-1)^{i+j}$.

Definition: The adjuncts are summarized in the *adjoint matrix*

$$\text{adj}(\mathbf{A}) \stackrel{\text{def}}{=} \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix}. \quad (\text{A.33})$$

With this matrix, the three equations (A.30) to (A.32) can be written as one equation

$$\mathbf{x} = \frac{\text{adj}(\mathbf{A})}{\det(\mathbf{A})} \mathbf{b}. \quad (\text{A.34})$$

Definition: The $n \times n$ matrix (whenever $\det(\mathbf{A}) \neq 0$)

$$\mathbf{A}^{-1} \stackrel{\text{def}}{=} \frac{\text{adj}(\mathbf{A})}{\det(\mathbf{A})} \quad (\text{A.35})$$

is called the *inverse matrix* of the square $n \times n$ matrix \mathbf{A} .

For a matrix product, the inverse matrix is obtained as

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1} \quad (\text{A.36})$$

because

$$(\mathbf{AB})(\mathbf{B}^{-1} \mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}.$$

A.2.3 Block Matrices

Often large matrices have a certain structure, e.g. when one or more sub-arrays are zero matrices. On the other hand, one can make a *block matrix* from each matrix by drawing vertical and horizontal lines. For a system of equations, one then, for example, obtains

$$\left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1n} \\ \hline A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hline A_{m1} & A_{m2} & \cdots & A_{mn} \end{array} \right) \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_m \end{pmatrix}. \quad (\text{A.37})$$

The A_{ij} 's are called *sub-matrices* and the vectors \mathbf{x}_i and \mathbf{y}_i *sub-vectors*. For two appropriately partitioned block matrices, the product may be obtained by simply carrying out the multiplication as if the sub-matrices were themselves elements:

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right) \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array} \right) = \left(\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hline A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array} \right).$$

In particular, the subdivision into blocks of matrices is of benefit for the calculation of the inverse matrix. Looking at the system of equations

$$\mathbf{Ax}_1 + \mathbf{Bx}_2 = \mathbf{y}_1, \quad (\text{A.38})$$

$$\mathbf{Cx}_1 + \mathbf{Dx}_2 = \mathbf{y}_2, \quad (\text{A.39})$$

or combined into the form

$$\left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right) \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad (\text{A.40})$$

the inverse of the matrix

$$\mathbf{M} = \left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right) \quad (\text{A.41})$$

can be expressed by more easily calculated inverse sub-matrices. When the matrix \mathbf{A} is invertible, one gets from (A.38)

$$\mathbf{x}_1 = \mathbf{A}^{-1} \mathbf{y}_1 - \mathbf{A}^{-1} \mathbf{Bx}_2. \quad (\text{A.42})$$

This, used in (A.39), yields

$$\mathbf{y}_2 = \mathbf{CA}^{-1} \mathbf{y}_1 - (\mathbf{CA}^{-1} \mathbf{B} - \mathbf{D}) \mathbf{x}_2 \quad (\text{A.43})$$

and, solving for \mathbf{x}_2 ,

$$\mathbf{x}_2 = (\mathbf{CA}^{-1} \mathbf{B} - \mathbf{D})^{-1} (\mathbf{CA}^{-1} \mathbf{y}_1 - \mathbf{y}_2). \quad (\text{A.44})$$

Equation (A.44) used in (A.42) yields

$$\mathbf{x}_1 = [A^{-1} - A^{-1}\mathbf{B}(\mathbf{C}A^{-1}\mathbf{B} - \mathbf{D})^{-1}\mathbf{C}A^{-1}]y_1 + A^{-1}\mathbf{B}(\mathbf{C}A^{-1}\mathbf{B} - \mathbf{D})^{-1}y_2. \quad (\text{A.45})$$

Thus, the solution of the system of equations (A.40) is obtained, namely

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \quad (\text{A.46})$$

with

$$\mathbf{M}^{-1} = \left(\begin{array}{c|c} A^{-1} - A^{-1}\mathbf{B}(\mathbf{C}A^{-1}\mathbf{B} - \mathbf{D})^{-1}\mathbf{C}A^{-1} & A^{-1}\mathbf{B}(\mathbf{C}A^{-1}\mathbf{B} - \mathbf{D})^{-1} \\ \hline (\mathbf{C}A^{-1}\mathbf{B} - \mathbf{D})^{-1}\mathbf{C}A^{-1} & -(\mathbf{C}A^{-1}\mathbf{B} - \mathbf{D})^{-1} \end{array} \right), \quad (\text{A.47})$$

and the inverse matrix of \mathbf{M} can be calculated by using the inverse matrices of the smaller sub-matrices \mathbf{A} and $(\mathbf{C}A^{-1}\mathbf{B} - \mathbf{D})$. If the sub-matrix \mathbf{D} is invertible, (A.39) can be solved for \mathbf{x}_2 , and then, in a similar way, also the inverse matrix of \mathbf{M} can be calculated. One then obtains a different form of the inverted matrix, namely

$$\mathbf{M}^{-1} = \left(\begin{array}{c|c} -(\mathbf{B}\mathbf{D}^{-1}\mathbf{C} - \mathbf{A})^{-1} & (\mathbf{B}\mathbf{D}^{-1}\mathbf{C} - \mathbf{A})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \hline \mathbf{D}^{-1}\mathbf{C}(\mathbf{B}\mathbf{D}^{-1}\mathbf{C} - \mathbf{A})^{-1} & \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{B}\mathbf{D}^{-1}\mathbf{C} - \mathbf{A})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{array} \right). \quad (\text{A.48})$$

So there are two different results available for the same matrix. It follows that the corresponding sub-matrices must be equal. From the comparison of the *northwestern* blocks, if \mathbf{D} is replaced by $-\mathbf{D}$, the known matrix-inversion lemma follows:

$$(\mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}\mathbf{A}^{-1}\mathbf{B} + \mathbf{D})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (\text{A.49})$$

A special case occurs when we have a block-triangular matrix, for example, if $\mathbf{C} = \mathbf{O}$. Then we obtain

$$\left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{O} & \mathbf{D} \end{array} \right)^{-1} = \left(\begin{array}{c|c} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \hline \mathbf{O} & \mathbf{D}^{-1} \end{array} \right). \quad (\text{A.50})$$

A.3 The Kronecker-Product

A.3.1 Definitions

Definition: The *Kronecker-product* of two matrices $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{B} \in \mathbb{R}^{p \times q}$ is a matrix $\mathbf{C} \in \mathbb{R}^{np \times mq}$, denoted as

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{C}.$$

Here, the sub-matrix $\mathbf{C}_{ij} \in \mathbb{R}^{p \times q}$, for $i = 1$ to n and $j = 1$ to m , is defined by

$$\mathbf{C}_{ij} \stackrel{\text{def}}{=} a_{ij} \mathbf{B},$$

so that, overall, the matrix \mathbf{C} has the form

$$\mathbf{C} = \begin{pmatrix} a_{11} \mathbf{B} & a_{12} \mathbf{B} & \dots & a_{1m} \mathbf{B} \\ a_{21} \mathbf{B} & a_{22} \mathbf{B} & \dots & a_{2m} \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} \mathbf{B} & a_{n2} \mathbf{B} & \dots & a_{nm} \mathbf{B} \end{pmatrix}.$$

The matrix elements of the product matrix \mathbf{C} can be directly calculated using the following formula

$$c_{i,j} = a_{\lfloor \frac{i-1}{p} \rfloor + 1, \lfloor \frac{j-1}{q} \rfloor + 1} \cdot b_{i - \lfloor \frac{i-1}{p} \rfloor p, j - \lfloor \frac{j-1}{q} \rfloor q},$$

where $\lfloor x \rfloor$ is the integer part of x .

Definition: If a matrix \mathbf{A} is composed of the m columns $\mathbf{a}_i \in \mathbb{C}^n$,

$$\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_m),$$

the *vec*-operator is defined as follows:

$$\text{vec}(\mathbf{A}) \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix} \in \mathbb{C}^{nm}.$$

A.3.2 Some Theorems

The following is very interesting:

Lemma:	$\text{vec}(\mathbf{A} \mathbf{X} \mathbf{B}) = (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{X}).$
---------------	---

(A.51)

Proof Let $\mathbf{B} \in \mathbb{C}^{n \times m}$, then one has

$$\begin{aligned}\mathbf{AXB} &= (\mathbf{Ax}_1 \quad \mathbf{Ax}_2 \quad \dots \quad \mathbf{Ax}_n) \mathbf{B} \\ &= (\mathbf{Ax}_1 \quad \mathbf{Ax}_2 \quad \dots \quad \mathbf{Ax}_n) (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_m) \\ &= \begin{pmatrix} (b_{11}\mathbf{Ax}_1 + b_{21}\mathbf{Ax}_2 + \dots + b_{n1}\mathbf{Ax}_n) & \dots \\ (b_{1m}\mathbf{Ax}_1 + b_{2m}\mathbf{Ax}_2 + \dots + b_{nm}\mathbf{Ax}_n) & \dots \end{pmatrix}.\end{aligned}$$

Applying the vec -operator to the last equation, we obtain

$$\begin{aligned}\text{vec}(\mathbf{AXB}) &= \begin{pmatrix} (b_{11}\mathbf{Ax}_1 + b_{21}\mathbf{Ax}_2 + \dots + b_{n1}\mathbf{Ax}_n) \\ \vdots \\ (b_{1m}\mathbf{Ax}_1 + b_{2m}\mathbf{Ax}_2 + \dots + b_{nm}\mathbf{Ax}_n) \end{pmatrix} \\ &= \begin{pmatrix} b_{11}\mathbf{A} & \dots & b_{n1}\mathbf{A} \\ \vdots & \ddots & \vdots \\ b_{1m}\mathbf{A} & \dots & b_{nm}\mathbf{A} \end{pmatrix} \text{vec}(\mathbf{X}) \\ &= (\mathbf{B}^\top \otimes \mathbf{A}) \text{vec}(\mathbf{X}).\end{aligned}\quad \square$$

As corollaries of this lemma, we get the following results:

$$\boxed{\text{vec}(\mathbf{AX}) = (\mathbf{I} \otimes \mathbf{A}) \text{vec}(\mathbf{X})}. \quad (\text{A.52})$$

Proof Set $\mathbf{B} = \mathbf{I}$ in the lemma. \square

$$\boxed{\text{vec}(\mathbf{XB}) = (\mathbf{B}^\top \otimes \mathbf{I}) \text{vec}(\mathbf{X})}. \quad (\text{A.53})$$

Proof Set $\mathbf{A} = \mathbf{I}$ in the lemma. \square

$$\boxed{\text{vec}(\mathbf{ba}^\top) = (\mathbf{a} \otimes \mathbf{b})}. \quad (\text{A.54})$$

Proof Simply write $\text{vec}(\mathbf{ba}^\top) = \text{vec}(\mathbf{b1a}^\top) = (\mathbf{a} \otimes \mathbf{b}) \text{vec}(\mathbf{1}) = \mathbf{a} \otimes \mathbf{b}$. \square

A.3.3 The Permutation Matrix $\mathbf{U}_{p \times q}$

Definition: The *permutation matrix*

$$\mathbf{U}_{p \times q} \stackrel{\text{def}}{=} \sum_i^p \sum_k^q \mathbf{E}_{ik}^{p \times q} \otimes \mathbf{E}_{ki}^{q \times p} \in \mathbb{R}^{pq \times qp} \quad (\text{A.55})$$

has just one 1 in each column and each row. In the formation of matrix

$$\mathbf{E}_{ik}^{p \times q} \stackrel{\text{def}}{=} \mathbf{e}_i \mathbf{e}_k^T, \quad (\text{A.56})$$

\mathbf{e}_i is the i th column of \mathbf{I}_p and \mathbf{e}_k is the k th column of \mathbf{I}_q . However, only the matrix element $E_{ik} = 1$; the other matrix elements are zeros.

For example, the permutation matrix $\mathbf{U}_{4 \times 4} \in \mathbb{R}^{16 \times 16}$ —often used in this book—has the form

$$\mathbf{U}_{4 \times 4} = \left(\begin{array}{cccc|cccc|cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right). \quad (\text{A.57})$$

Permutation matrices have the following characteristics [4]:

$$\mathbf{U}_{p \times q}^T = \mathbf{U}_{q \times p}, \quad (\text{A.58})$$

$$\mathbf{U}_{p \times q}^{-1} = \mathbf{U}_{q \times p}, \quad (\text{A.59})$$

$$\mathbf{U}_{p \times 1} = \mathbf{U}_{1 \times p} = \mathbf{I}_p, \quad (\text{A.60})$$

$$\mathbf{U}_{n \times n} = \mathbf{U}_{n \times n}^T = \mathbf{U}_{n \times n}^{-1}. \quad (\text{A.61})$$

Permutation matrices are mainly used to change the order of the factors in a Kronecker-product because

$$U_{s \times p}(\mathbf{B} \otimes \mathbf{A})U_{q \times t} = \mathbf{A} \otimes \mathbf{B} \quad \text{if } \mathbf{A} \in \mathbb{R}^{p \times q} \text{ and } \mathbf{B} \in \mathbb{R}^{s \times t}. \quad (\text{A.62})$$

A.3.4 More Properties of the Kronecker-Product

The following important properties are listed also without proof (see [4]):

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}), \quad (\text{A.63})$$

$$(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top, \quad (\text{A.64})$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}. \quad (\text{A.65})$$

A.4 Derivatives of Vectors/Matrices with Respect to Vectors/Matrices

A.4.1 Definitions

Definition: The derivative of a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ with respect to a matrix $\mathbf{M} \in \mathbb{R}^{r \times s}$ is defined as follows:

$$\frac{\partial \mathbf{A}}{\partial \mathbf{M}} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial \mathbf{A}}{\partial M_{11}} & \frac{\partial \mathbf{A}}{\partial M_{12}} & \cdots & \frac{\partial \mathbf{A}}{\partial M_{1s}} \\ \frac{\partial \mathbf{A}}{\partial M_{21}} & \frac{\partial \mathbf{A}}{\partial M_{22}} & \cdots & \frac{\partial \mathbf{A}}{\partial M_{2s}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{A}}{\partial M_{r1}} & \frac{\partial \mathbf{A}}{\partial M_{r2}} & \cdots & \frac{\partial \mathbf{A}}{\partial M_{rs}} \end{pmatrix} \in \mathbb{R}^{nr \times ms}. \quad (\text{A.66})$$

With the $r \times s$ -operator

$$\frac{\partial}{\partial \mathbf{M}} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial}{\partial M_{11}} & \frac{\partial}{\partial M_{12}} & \cdots & \frac{\partial}{\partial M_{1s}} \\ \frac{\partial}{\partial M_{21}} & \frac{\partial}{\partial M_{22}} & \cdots & \frac{\partial}{\partial M_{2s}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial M_{r1}} & \frac{\partial}{\partial M_{r2}} & \cdots & \frac{\partial}{\partial M_{rs}} \end{pmatrix} \quad (\text{A.67})$$

the definition of the derivative (A.66) is also written as

$$\frac{\partial \mathbf{A}}{\partial \mathbf{M}} \stackrel{\text{def}}{=} \frac{\partial}{\partial \mathbf{M}} \otimes \mathbf{A}. \quad (\text{A.68})$$

Thus one can show that

$$\left(\frac{\partial \mathbf{A}}{\partial \mathbf{M}} \right)^\top = \left(\frac{\partial}{\partial \mathbf{M}} \otimes \mathbf{A} \right)^\top = \left(\left(\frac{\partial}{\partial \mathbf{M}} \right)^\top \otimes \mathbf{A}^\top \right) = \frac{\partial \mathbf{A}^\top}{\partial \mathbf{M}^\top}. \quad (\text{A.69})$$

For the derivatives of vectors with respect to vectors one has:

$$\frac{\partial \mathbf{f}^\top}{\partial \mathbf{p}} \stackrel{\text{def}}{=} \frac{\partial}{\partial \mathbf{p}} \otimes \mathbf{f}^\top = \begin{pmatrix} \frac{\partial f_1^\top}{\partial p_1} & \frac{\partial f_2^\top}{\partial p_1} & \cdots & \frac{\partial f_n^\top}{\partial p_1} \\ \frac{\partial f_1^\top}{\partial p_2} & \frac{\partial f_2^\top}{\partial p_2} & \cdots & \frac{\partial f_n^\top}{\partial p_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1^\top}{\partial p_r} & \frac{\partial f_2^\top}{\partial p_r} & \cdots & \frac{\partial f_n^\top}{\partial p_r} \end{pmatrix} \in \mathbb{R}^{r \times n} \quad (\text{A.70})$$

and

$$\frac{\partial \mathbf{f}}{\partial \mathbf{p}^\top} \stackrel{\text{def}}{=} \frac{\partial}{\partial \mathbf{p}^\top} \otimes \mathbf{f} = \left(\frac{\partial}{\partial \mathbf{p}} \otimes \mathbf{f}^\top \right)^\top = \left(\frac{\partial \mathbf{f}^\top}{\partial \mathbf{p}} \right)^\top \in \mathbb{R}^{n \times r}. \quad (\text{A.71})$$

A.4.2 Product Rule

Let $\mathbf{A} = A(\alpha)$ and $\mathbf{B} = B(\alpha)$. Then we obviously have

$$\frac{\partial(\mathbf{AB})}{\partial \alpha} = \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{B} + \mathbf{A} \frac{\partial \mathbf{B}}{\partial \alpha}. \quad (\text{A.72})$$

In addition, (A.66) can be rewritten as:

$$\frac{\partial \mathbf{A}}{\partial \mathbf{M}} = \sum_{i,k} \mathbf{E}_{ik}^{s \times t} \otimes \frac{\partial \mathbf{A}}{\partial m_{ik}}, \quad \mathbf{M} \in \mathbb{R}^{s \times t}. \quad (\text{A.73})$$

Using (A.72) and (A.73), this product rule can be derived:

$$\begin{aligned} \underline{\underline{\frac{\partial(\mathbf{AB})}{\partial \mathbf{M}}}} &= \sum_{i,k} \mathbf{E}_{ik}^{s \times t} \otimes \frac{\partial(\mathbf{AB})}{\partial m_{ik}} = \sum_{i,k} \mathbf{E}_{ik}^{s \times t} \otimes \left(\frac{\partial \mathbf{A}}{\partial m_{ik}} \mathbf{B} + \mathbf{A} \frac{\partial \mathbf{B}}{\partial m_{ik}} \right) \\ &= \left(\frac{\partial}{\partial \mathbf{M}} \otimes \mathbf{A} \right) (\mathbf{I}_t \otimes \mathbf{B}) + (\mathbf{I}_s \otimes \mathbf{A}) \left(\frac{\partial}{\partial \mathbf{M}} \otimes \mathbf{B} \right) \\ &= \underline{\underline{\frac{\partial \mathbf{A}}{\partial \mathbf{M}}}} (\mathbf{I}_t \otimes \mathbf{B}) + (\mathbf{I}_s \otimes \mathbf{A}) \frac{\partial \mathbf{B}}{\partial \mathbf{M}}. \end{aligned} \quad (\text{A.74})$$

A.4.3 Chain Rule

When a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ is a function of a matrix $\mathbf{B} \in \mathbb{R}^{k \times \ell}$ which is again a function of a matrix $\mathbf{M} \in \mathbb{R}^{r \times s}$, then the chain rule is valid [4]:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{M}} A(\mathbf{B}(\mathbf{M})) &= \left(\mathbf{I}_r \otimes \frac{\partial \mathbf{A}}{\partial (\text{vec}(\mathbf{B}^\top))^\top} \right) \left(\frac{\partial \text{vec}(\mathbf{B}^\top)}{\partial \mathbf{M}} \otimes \mathbf{I}_m \right) \\ &= \left(\frac{\partial (\text{vec}(\mathbf{B}))^\top}{\partial \mathbf{M}} \otimes \mathbf{I}_n \right) \left(\mathbf{I}_s \otimes \frac{\partial \mathbf{A}}{\partial \text{vec}(\mathbf{B})} \right).\end{aligned}\quad (\text{A.75})$$

A special case of this is

$$\frac{d\mathbf{A}(\mathbf{x}(t))}{dt} = \frac{\partial \mathbf{A}}{\partial \mathbf{x}^\top} \left(\frac{d\mathbf{x}}{dt} \otimes \mathbf{I}_m \right) = \left(\frac{d\mathbf{x}^\top}{dt} \otimes \mathbf{I}_n \right) \frac{\partial \mathbf{A}}{\partial \mathbf{x}} \in \mathbb{R}^{n \times m}. \quad (\text{A.76})$$

A.5 Differentiation with Respect to Time

A.5.1 Differentiation of a Function with Respect to Time

Suppose a function a that depends on the three space variables x_1, x_2 and x_3 is given. The local space variables themselves are, in turn, dependent on the time parameter t . So it is

$$a = a(\mathbf{x}(t)), \quad (\text{A.77})$$

if the three space variables are summarized in the vector \mathbf{x} .

We want to find the velocity

$$\dot{a} = \frac{da}{dt}. \quad (\text{A.78})$$

To determine this velocity, we first define the total difference

$$\Delta a \stackrel{\text{def}}{=} \frac{\partial a}{\partial x_1} \Delta x_1 + \frac{\partial a}{\partial x_2} \Delta x_2 + \frac{\partial a}{\partial x_3} \Delta x_3. \quad (\text{A.79})$$

After division by Δt , in the limit $\Delta t \rightarrow 0$, one has

$$\dot{a} = \frac{da}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta a}{\Delta t} = \frac{\partial a}{\partial x_1} \dot{x}_1 + \frac{\partial a}{\partial x_2} \dot{x}_2 + \frac{\partial a}{\partial x_3} \dot{x}_3. \quad (\text{A.80})$$

The right-hand side of this equation can be presented by the scalar product of the column vectors $\dot{\mathbf{x}}$ and $\frac{\partial a}{\partial \mathbf{x}}$ in two ways, namely

$\dot{a} = \dot{\mathbf{x}}^\top \frac{\partial a}{\partial \mathbf{x}} = \frac{\partial a}{\partial \mathbf{x}^\top} \dot{\mathbf{x}}.$

(A.81)

A.5.2 Differentiation of a Vector with Respect to Time

If two functions a_1 and a_2 are given, they have the same dependence on time t as $a(t)$ in (A.77), and are summarized in the column vector

$$\mathbf{a} \stackrel{\text{def}}{=} \begin{pmatrix} a_1(\mathbf{x}(t)) \\ a_2(\mathbf{x}(t)) \end{pmatrix}, \quad (\text{A.82})$$

we obtain initially for the derivative with respect to time using (A.81)

$$\dot{\mathbf{a}} = \begin{pmatrix} \dot{a}_1 \\ \dot{a}_2 \end{pmatrix} = \begin{pmatrix} \dot{\mathbf{x}}^\top \frac{\partial a_1}{\partial \mathbf{x}} \\ \dot{\mathbf{x}}^\top \frac{\partial a_2}{\partial \mathbf{x}} \end{pmatrix} = \begin{pmatrix} \frac{\partial a_1}{\partial \mathbf{x}^\top} \dot{\mathbf{x}} \\ \frac{\partial a_2}{\partial \mathbf{x}^\top} \dot{\mathbf{x}} \end{pmatrix}. \quad (\text{A.83})$$

The next-to-last vector in (A.83) can be decomposed as follows:

$$\dot{\mathbf{a}} = \begin{pmatrix} \dot{\mathbf{x}}^\top \frac{\partial a_1}{\partial \mathbf{x}} \\ \dot{\mathbf{x}}^\top \frac{\partial a_2}{\partial \mathbf{x}} \end{pmatrix} = \begin{pmatrix} \dot{\mathbf{x}}^\top & \mathbf{o}_3^\top \\ \mathbf{o}_3^\top & \dot{\mathbf{x}}^\top \end{pmatrix} \begin{pmatrix} \frac{\partial a_1}{\partial \mathbf{x}} \\ \frac{\partial a_2}{\partial \mathbf{x}} \end{pmatrix} = (\mathbf{I}_2 \otimes \dot{\mathbf{x}}^\top) \left(\mathbf{a} \otimes \frac{\partial}{\partial \mathbf{x}} \right). \quad (\text{A.84})$$

Computing the last Kronecker-product, we would formally get $a_i \frac{\partial}{\partial \mathbf{x}}$ which, of course, should be understood as $\frac{\partial a_i}{\partial \mathbf{x}}$. With the help of the permutation matrix $\mathbf{U}_{\alpha \times \beta}$ and the exchange rule (A.62)

$$(\mathbf{A} \otimes \mathbf{B}) = \mathbf{U}_{s \times p} (\mathbf{B} \otimes \mathbf{A}) \mathbf{U}_{q \times t} \quad \text{if } \mathbf{A} \in \mathbb{R}^{p \times q} \text{ and } \mathbf{B} \in \mathbb{R}^{s \times t},$$

in the appendix, the last product in (A.84) can be written

$$\underline{\underline{\dot{\mathbf{a}}}} = \underbrace{[\mathbf{U}_{1 \times 2} (\dot{\mathbf{x}}^\top \otimes \mathbf{I}_2) \mathbf{U}_{2 \times r}]}_{\mathbf{I}_{2r}} \underbrace{[\mathbf{U}_{r \times 2} \left(\frac{\partial}{\partial \mathbf{x}} \otimes \mathbf{a} \right) \mathbf{U}_{1 \times 1}]}_{\frac{\partial \mathbf{a}}{\partial \mathbf{x}}} = \underline{\underline{(\dot{\mathbf{x}}^\top \otimes \mathbf{I}_2) \frac{\partial \mathbf{a}}{\partial \mathbf{x}}}}. \quad (\text{A.85})$$

For the second form in (A.83), one obtains

$$\underline{\underline{\dot{\mathbf{a}}}} = \begin{pmatrix} \frac{\partial a_1}{\partial \mathbf{x}^\top} \dot{\mathbf{x}} \\ \frac{\partial a_2}{\partial \mathbf{x}^\top} \dot{\mathbf{x}} \end{pmatrix} = \left(\mathbf{a} \otimes \frac{\partial}{\partial \mathbf{x}} \right) \dot{\mathbf{x}} = \left[\underbrace{\mathbf{U}_{1 \times 2} \left(\frac{\partial}{\partial \mathbf{x}^\top} \otimes \mathbf{a} \right)}_{\mathbf{I}_2} \underbrace{\mathbf{U}_{1 \times r}}_{\mathbf{I}_r} \right] \dot{\mathbf{x}} = \underline{\underline{\frac{\partial \mathbf{a}}{\partial \mathbf{x}^\top} \dot{\mathbf{x}}}}, \quad (\text{A.86})$$

so that combined, these two possible representations are written as

$$\dot{\mathbf{a}} = (\dot{\mathbf{x}}^\top \otimes \mathbf{I}_2) \frac{\partial \mathbf{a}}{\partial \mathbf{x}} = \underline{\underline{\frac{\partial \mathbf{a}}{\partial \mathbf{x}^\top} \dot{\mathbf{x}}}}.$$

(A.87)

A.5.3 Differentiation of a 2×3 -Matrix with Respect to Time

For the derivative of a 2×3 -matrix with respect to time, with the above results, one obtains

$$\begin{aligned}\dot{\underline{A}} &= \begin{pmatrix} \dot{a}_{11} & \dot{a}_{12} & \dot{a}_{13} \\ \dot{a}_{21} & \dot{a}_{22} & \dot{a}_{23} \end{pmatrix} = \begin{pmatrix} \dot{x}^\top \frac{\partial a_{11}}{\partial x} & \dot{x}^\top \frac{\partial a_{12}}{\partial x} & \dot{x}^\top \frac{\partial a_{13}}{\partial x} \\ \dot{x}^\top \frac{\partial a_{21}}{\partial x} & \dot{x}^\top \frac{\partial a_{22}}{\partial x} & \dot{x}^\top \frac{\partial a_{23}}{\partial x} \end{pmatrix} \\ &= \begin{pmatrix} \dot{x}^\top & \mathbf{o}_3^\top \\ \mathbf{o}_3^\top & \dot{x}^\top \end{pmatrix} \left(A \otimes \frac{\partial}{\partial x} \right) = (\mathbf{I}_2 \otimes \dot{x}^\top) \left(A \otimes \frac{\partial}{\partial x} \right) \\ &= \underbrace{[\mathbf{U}_{1 \times 2}(\dot{x}^\top \otimes \mathbf{I}_2) \mathbf{U}_{2 \times r}]}_{\mathbf{I}_{2r}} \underbrace{[\mathbf{U}_{r \times 2} \left(\frac{\partial}{\partial x} \otimes A \right) \mathbf{U}_{3 \times 1}]}_{\frac{\partial A}{\partial x}} = \underbrace{(\dot{x}^\top \otimes \mathbf{I}_2) \frac{\partial A}{\partial x}}, \quad (\text{A.88})\end{aligned}$$

or, with the second representation in (A.81) for the \dot{a}_{ij} ,

$$\begin{aligned}\dot{\underline{A}} &= \begin{pmatrix} \frac{\partial a_{11}}{\partial x^\top} \dot{x} & \frac{\partial a_{12}}{\partial x^\top} \dot{x} & \frac{\partial a_{13}}{\partial x^\top} \dot{x} \\ \frac{\partial a_{21}}{\partial x^\top} \dot{x} & \frac{\partial a_{22}}{\partial x^\top} \dot{x} & \frac{\partial a_{23}}{\partial x^\top} \dot{x} \end{pmatrix} = \left(A \otimes \frac{\partial}{\partial x^\top} \right) \begin{pmatrix} \dot{x} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \dot{x} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \dot{x} \end{pmatrix} \\ &= \underbrace{[\mathbf{U}_{1 \times 2} \left(\frac{\partial}{\partial x^\top} \otimes A \right) \mathbf{U}_{3 \times r}]}_{\frac{\partial A}{\partial x^\top}} \underbrace{[\mathbf{U}_{r \times 3}(\dot{x} \otimes \mathbf{I}_3) \mathbf{U}_{3 \times 1}]}_{\mathbf{I}_{3r}} = \underbrace{\frac{\partial A}{\partial x^\top} (\dot{x} \otimes \mathbf{I}_3)}. \quad (\text{A.89})\end{aligned}$$

Here the Kronecker-product is also present.

A.5.4 Differentiation of an $n \times m$ -Matrix with Respect to Time

In general, one gets for a matrix $A \in \mathbb{R}^{n \times m}$ and a vector $x \in \mathbb{R}^r$

$$\dot{A} = (\dot{x}^\top \otimes \mathbf{I}_n) \frac{\partial A}{\partial x} \stackrel{\text{and}}{=} \frac{\partial A}{\partial x^\top} (\dot{x} \otimes \mathbf{I}_m) \in \mathbb{R}^{n \times m}. \quad (\text{A.90})$$

The derivation is given below without any comment.

$$\begin{aligned}\dot{A} &= \begin{pmatrix} \dot{x}^\top & \dots & \mathbf{O} \\ \vdots & \ddots & \vdots \\ \mathbf{O} & \dots & \dot{x}^\top \end{pmatrix} \left(A \otimes \frac{\partial}{\partial x} \right) = (\mathbf{I}_n \otimes \dot{x}^\top) \left(A \otimes \frac{\partial}{\partial x} \right) \\ &= \underbrace{[\mathbf{U}_{1 \times n}(\dot{x}^\top \otimes \mathbf{I}_n) \mathbf{U}_{n \times r}]}_{\mathbf{I}_{nr}} \underbrace{[\mathbf{U}_{r \times n} \left(\frac{\partial}{\partial x} \otimes A \right) \mathbf{U}_{m \times 1}]}_{\frac{\partial A}{\partial x}} = \underbrace{(\dot{x}^\top \otimes \mathbf{I}_n) \frac{\partial A}{\partial x}}_{\frac{\partial A}{\partial x}}.\end{aligned}$$

$$\begin{aligned}
\dot{\mathbf{A}} &= \left(\mathbf{A} \otimes \frac{\partial}{\partial \mathbf{x}^\top} \right) \begin{pmatrix} \dot{\mathbf{x}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \dot{\mathbf{x}} \end{pmatrix} = \left(\mathbf{A} \otimes \frac{\partial}{\partial \mathbf{x}^\top} \right) (\mathbf{I}_n \otimes \dot{\mathbf{x}}) \\
&= \underbrace{\left[\underbrace{\mathbf{U}_{1 \times n} \left(\frac{\partial}{\partial \mathbf{x}^\top} \otimes \mathbf{A} \right) \mathbf{U}_{m \times r}}_{\frac{\partial \mathbf{A}}{\partial \mathbf{x}^\top}} \right]}_{\mathbf{I}_{mr}} \underbrace{[\mathbf{U}_{r \times m} (\dot{\mathbf{x}} \otimes \mathbf{I}_m) \mathbf{U}_{m \times 1}]}_{\mathbf{I}_{mr}} = \underbrace{\frac{\partial \mathbf{A}}{\partial \mathbf{x}^\top} (\dot{\mathbf{x}} \otimes \mathbf{I}_n)}_{\mathbf{I}_{mr}}
\end{aligned}$$

A.6 Supplements to Differentiation with Respect to a Matrix

For the derivative of a 4×4 matrix with respect to itself, one has

$$\frac{\partial \mathbf{M}}{\partial \mathbf{M}} = \bar{\mathbf{U}}_{4 \times 4}, \quad (\text{A.91})$$

where $\bar{\mathbf{U}}_{4 \times 4}$ is defined by

$$\begin{aligned}
\bar{\mathbf{U}}_{4 \times 4} &\stackrel{\text{def}}{=} \sum_i^4 \sum_k^4 \mathbf{E}_{ik} \otimes \mathbf{E}_{ik} \\
&= \left(\begin{array}{cccc|cccc|cccc|cccc}
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array} \right). \quad (\text{A.92})
\end{aligned}$$

This fact can be easily made clear just by the definition of the differentiation of a matrix with respect to a matrix. The result is more complex when the matrix $\mathbf{M} = \mathbf{M}^\top$

is symmetric because then

$$\begin{aligned} \frac{\partial \mathbf{M}}{\partial \mathbf{M}} &= \bar{\mathbf{U}}_{4 \times 4} + \mathbf{U}_{4 \times 4} - \sum_i^4 \mathbf{E}_{ii} \otimes \mathbf{E}_{ii} \\ &= \left(\begin{array}{cccc|cccc|cccc|cccc} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{array} \right) . \quad (\text{A.93}) \end{aligned}$$

Appendix B

Some Differential Geometry

From a sheet of letter paper, one can form a cylinder or a cone, but it is impossible to obtain a surface element of a sphere without folding, stretching or cutting. The reason lies in the geometry of the spherical surface: No part of such a surface can be isometrically mapped onto the plane.

B.1 Curvature of a Curved Line in Three Dimensions

In a plane, the tangent vector remains constant when one moves on it: the plane has no curvature. The same is true for a straight line when the tangent vector coincides with the line. If a line in a neighbourhood of one of its points is not a straight line, it is called a *curved* line. The same is valid for a curved surface. We consider curves in the three-dimensional space with the position vector $\mathbf{x}(q)$ of points parametrized by the length q . The direction of a curve \mathcal{C} at the point $\mathbf{x}(q)$ is given by the normalized tangent vector

$$\mathbf{t}(q) \stackrel{\text{def}}{=} \frac{\mathbf{x}'(q)}{\|\mathbf{x}'(q)\|},$$

with $\mathbf{x}'(q) \stackrel{\text{def}}{=} \frac{\partial \mathbf{x}(q)}{\partial q}$. Passing through a curve from a starting point $\mathbf{x}(q_0)$ to an endpoint $\mathbf{x}(q)$ does not change the tangent vector \mathbf{t} in a straight line, the tip of the tangent vector does not move, thus describing a curve of length 0. If the curve is curved, then the tip of the tangent vector describes an arc of length not equal to zero. As *arc length* we call the integral over a curved arc with endpoints q_0 and q ($q > q_0$):

$$\int_{q_0}^q \sqrt{x_1'^2 + x_2'^2 + x_3'^2} dq = \int_{q_0}^q \sqrt{\mathbf{x}'^\top \mathbf{x}'} dq = \int_{q_0}^q \|\mathbf{x}'\| dq. \quad (\text{B.1})$$

This formula arises as follows: Suppose the interval $[q_0, q]$ is divided by the points $q_0 < q_1 < \dots < q_n = q$, then the length σ_n of the inscribed polygon in the curve

arch is

$$\sigma_n = \sum_{k=0}^n \sqrt{[x_1(q_k) - x_1(q_{k+1})]^2 + [x_2(q_k) - x_2(q_{k+1})]^2 + [x_3(q_k) - x_3(q_{k+1})]^2}. \quad (\text{B.2})$$

By the mean value theorem of differential calculus, for every smooth curve between q_k and q_{k+1} there exists a point $q_k^{(i)}$ such that

$$x_i(q_k) - x_i(q_{k+1}) = x'_i(q_k^{(i)})(q_{k+1} - q_k) \quad (\text{B.3})$$

for $i = 1, 2$ and 3 . Inserting (B.3) into (B.2) yields

$$\sigma_n = \sum_{k=0}^n [q_{k+1} - q_k] \sqrt{[x'_1(q_k^{(1)})]^2 + [x'_2(q_k^{(2)})]^2 + [x'_3(q_k^{(3)})]^2}. \quad (\text{B.4})$$

From (B.4) one gets, as $q_{k+1} - q_k \rightarrow 0$, (B.1).

A measure of the curvature of a curve is the rate of change of the direction. The curvature is larger when the change of direction of the tangent vector \mathbf{t} is greater. Generally, we therefore define as a *curvature of a curve* \mathcal{C} at a point $\mathbf{x}(q_0)$

$$\kappa(q_0) \stackrel{\text{def}}{=} \lim_{q \rightarrow q_0} \frac{\text{length of } \mathbf{t}}{\text{length of } \mathbf{x}} = \frac{\|\mathbf{t}'(q_0)\|}{\|\mathbf{x}'(q_0)\|},$$

(B.5)

where

$$\text{length of } \mathbf{t} \stackrel{\text{def}}{=} \int_{q_0}^q \|\mathbf{t}'\| dq$$

and

$$\text{length of } \mathbf{x} \stackrel{\text{def}}{=} \int_{q_0}^q \|\mathbf{x}'\| dq.$$

A straight line has zero curvature. In the case of a circle, the curvature is constant; the curvature is greater, the smaller the radius. The reciprocal value $1/\kappa$ of the curvature is called the *curvature radius*.

B.2 Curvature of a Surface in Three Dimensions

B.2.1 Vectors in the Tangent Plane

Already in the nineteenth century, Gauss investigated how from the measurements on a surface one can make conclusions about its spatial form. He then came to his main result, the *Theorema Egregium*, which states that the Gaussian curvature of a

surface depends only on the internal variables g_{ij} and their derivatives. This result is deduced in the following.

Suppose an area is defined as a function $\mathbf{x}(q_1, q_2) \in \mathbb{R}^3$ of the two coordinates q_1 and q_2 . At a point P of the surface, the tangent plane is, for example, spanned by the two tangent vectors $\mathbf{x}_1 \stackrel{\text{def}}{=} \frac{\partial \mathbf{x}}{\partial q_1}$ and $\mathbf{x}_2 \stackrel{\text{def}}{=} \frac{\partial \mathbf{x}}{\partial q_2}$. If the two tangent vectors \mathbf{x}_1 and \mathbf{x}_2 are linearly independent, any vector in the tangent plane can be decomposed in a linear combination of the two vectors, e.g. as

$$v^1 \mathbf{x}_1 + v^2 \mathbf{x}_2.$$

The scalar product of two vectors from the tangent plane is then defined as

$$\begin{aligned} \underline{\underline{(\mathbf{v} \cdot \mathbf{w})}} &\stackrel{\text{def}}{=} (v^1 \mathbf{x}_1^\top + v^2 \mathbf{x}_2^\top)(w^1 \mathbf{x}_1 + w^2 \mathbf{x}_2) \\ &= \mathbf{v}^\top \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} [\mathbf{x}_1, \mathbf{x}_2] \mathbf{w} = \mathbf{v}^\top \begin{pmatrix} \mathbf{x}_1^\top \mathbf{x}_1 & \mathbf{x}_1^\top \mathbf{x}_2 \\ \mathbf{x}_2^\top \mathbf{x}_1 & \mathbf{x}_2^\top \mathbf{x}_2 \end{pmatrix} \mathbf{w} = \underline{\underline{\mathbf{v}^\top \mathbf{G} \mathbf{w}}}. \end{aligned}$$

As $\mathbf{x}_1^\top \mathbf{x}_2 = \mathbf{x}_2^\top \mathbf{x}_1$, the matrix $\mathbf{G}^\top = \mathbf{G}$ is symmetric. In addition, one has

$$\|\mathbf{v}\| = \sqrt{(\mathbf{v} \cdot \mathbf{v})} = \sqrt{\mathbf{v}^\top \mathbf{G} \mathbf{v}}.$$

Now we want to define the curvature by oriented parallelograms. Let $\mathbf{v} \wedge \mathbf{w}$ be the oriented parallelogram defined by the vectors \mathbf{v} and \mathbf{w} in *that order*. $\mathbf{w} \wedge \mathbf{v} = -\mathbf{v} \wedge \mathbf{w}$ is then the parallelogram with the opposite orientation, $\text{area}(\mathbf{w} \wedge \mathbf{v}) = -\text{area}(\mathbf{v} \wedge \mathbf{w})$. The determinant of the matrix \mathbf{G} is then obtained as

$$\begin{aligned} \underline{\underline{g}} &\stackrel{\text{def}}{=} \det \mathbf{G} = \|\mathbf{x}_1\|^2 \cdot \|\mathbf{x}_2\|^2 - (\mathbf{x}_1 \cdot \mathbf{x}_2)^2 \\ &= \|\mathbf{x}_1\|^2 \cdot \|\mathbf{x}_2\|^2 - \|\mathbf{x}_1\|^2 \cdot \|\mathbf{x}_2\|^2 \cos^2 \Theta \\ &= \|\mathbf{x}_1\|^2 \cdot \|\mathbf{x}_2\|^2 (1 - \cos^2 \Theta) = \underline{\underline{\|\mathbf{x}_1\|^2 \cdot \|\mathbf{x}_2\|^2 \sin^2 \Theta}}. \end{aligned}$$

On the other hand, one has $\|\mathbf{x}_1 \times \mathbf{x}_2\| = \|\mathbf{x}_1\| \cdot \|\mathbf{x}_2\| \cdot \sin \Theta$, and so

$$\underline{\underline{\|\mathbf{x}_1 \times \mathbf{x}_2\|}} = \sqrt{g} = \text{area}(\mathbf{x}_1 \wedge \mathbf{x}_2). \quad (\text{B.6})$$

For the vector product of two vectors \mathbf{v} and \mathbf{w} from the tangent plane, one gets, on the other hand,

$$\begin{aligned} &(v^1 \mathbf{x}_1 + v^2 \mathbf{x}_2) \times (w^1 \mathbf{x}_1 + w^2 \mathbf{x}_2) \\ &= v^1 w^1 \underbrace{(\mathbf{x}_1 \times \mathbf{x}_1)}_0 + v^1 w^2 (\mathbf{x}_1 \times \mathbf{x}_2) + v^2 w^1 \underbrace{(\mathbf{x}_2 \times \mathbf{x}_1)}_{-(\mathbf{x}_1 \times \mathbf{x}_2)} + v^2 w^2 \underbrace{(\mathbf{x}_2 \times \mathbf{x}_2)}_0 \\ &= \underbrace{(v^1 w^2 - v^2 w^1)}_{\stackrel{\text{def}}{=} \det \mathbf{R}} (\mathbf{x}_1 \times \mathbf{x}_2), \end{aligned}$$

or

$$\text{area}(\mathbf{v} \wedge \mathbf{w}) = \det \mathbf{R} \cdot \text{area}(\mathbf{x}_1 \wedge \mathbf{x}_2),$$

so

$$\underline{\underline{\text{area}(\mathbf{v} \wedge \mathbf{w})}} = \det \mathbf{R} \sqrt{g}. \quad (\text{B.7})$$

The area of the parallelogram spanned by two vectors in the tangential surface is thus determined by the vector components and the surface defining the matrix \mathbf{G} .

B.2.2 Curvature and Normal Vectors

For two-dimensional surfaces, one should define the curvature at a point without using the tangent vectors directly because there are infinitely many of them in the tangent plane. However, any smooth surface in \mathbb{R}^3 has at each point a unique normal direction, which is one-dimensional, so it can be described by the unit normal vector. A *normal vector* at $\mathbf{x}(q)$ is defined as the normalized vector perpendicular to the tangent plane:

$$\mathbf{n}(q) \stackrel{\text{def}}{=} \frac{\mathbf{x}_1 \times \mathbf{x}_2}{\|\mathbf{x}_1 \times \mathbf{x}_2\|}.$$

If the surface is curved, the normal vector changes with displacement according to

$$\mathbf{n}_i(q) \stackrel{\text{def}}{=} \frac{\partial \mathbf{n}(q)}{\partial q^i}.$$

These change vectors \mathbf{n}_i lie in the tangent plane because

$$\frac{\partial(\mathbf{n} \cdot \mathbf{n})}{\partial q^i} = 0 = \left(\frac{\partial \mathbf{n}}{\partial q^i} \cdot \mathbf{n} \right) + \left(\mathbf{n} \cdot \frac{\partial \mathbf{n}}{\partial q^i} \right) = 2(\mathbf{n}_i \cdot \mathbf{n}).$$

The bigger the area spanned by the two change vectors \mathbf{n}_1 and \mathbf{n}_2 lying in the tangent plane, the bigger the curvature at the considered point. If Ω is an area of the tangent surface which contains the point under consideration, then the following curvature definition is obvious:

$$\begin{aligned} \kappa(q) &\stackrel{\text{def}}{=} \lim_{\Omega \rightarrow q} \frac{\text{area of } \mathbf{n}(\Omega)}{\text{area of } \Omega} = \lim_{\Omega \rightarrow q} \frac{\iint_{\Omega} \|\mathbf{n}_1(\tilde{q}) \times \mathbf{n}_2(\tilde{q})\| d\tilde{q}^1 d\tilde{q}^2}{\iint_{\Omega} \|\mathbf{x}_1(\tilde{q}) \times \mathbf{x}_2(\tilde{q})\| d\tilde{q}^1 d\tilde{q}^2} \\ &= \frac{\|\mathbf{n}_1(q) \times \mathbf{n}_2(q)\|}{\|\mathbf{x}_1(q) \times \mathbf{x}_2(q)\|} = \frac{\text{area of } \mathbf{n}_1(q) \wedge \mathbf{n}_2(q)}{\text{area of } \mathbf{x}_1(q) \wedge \mathbf{x}_2(q)}. \end{aligned} \quad (\text{B.8})$$

Since both \mathbf{n}_1 and \mathbf{n}_2 lie in the tangent plane, they can be displayed as linear combinations of the vectors \mathbf{x}_1 and \mathbf{x}_2 :

$$\mathbf{n}_1 = -b_1^1 \mathbf{x}_1 - b_1^2 \mathbf{x}_2 \quad \text{and} \quad \mathbf{n}_2 = -b_2^1 \mathbf{x}_1 - b_2^2 \mathbf{x}_2. \quad (\text{B.9})$$

Combining together the coefficients $-b_i^j$ in the matrix $\overline{\mathbf{B}}$, this corresponds to the matrix \mathbf{R} in (B.7), and (B.6) then yields the *Gauss-curvature*

$$\underline{\underline{\kappa(\mathbf{q}) = \det \overline{\mathbf{B}}}}. \quad (\text{B.10})$$

To confirm the *Theorema Egregium* of Gauss, one must now show that $\overline{\mathbf{B}}$ depends only on the inner values g_{ij} and their derivatives!

B.2.3 Theorema Egregium and the Inner Values g_{ij}

First, we examine the changes of the tangent vectors \mathbf{x}_k by looking at their derivatives

$$\mathbf{x}_{jk} \stackrel{\text{def}}{=} \frac{\partial \mathbf{x}_k}{\partial q^j} = \frac{\partial^2 \mathbf{x}}{\partial q^j \partial q^k}, \quad (\text{B.11})$$

which implies that

$$\mathbf{x}_{jk} = \mathbf{x}_{kj}. \quad (\text{B.12})$$

Since the two vectors \mathbf{x}_1 and \mathbf{x}_2 are a basis for the tangent plane and the vector \mathbf{n} is orthonormal to this plane, any vector in \mathbb{R}^3 , including the vector \mathbf{x}_{jk} , can be assembled as a linear combination of these three vectors:

$$\underline{\underline{\mathbf{x}_{jk} = \Gamma_{jk}^1 \mathbf{x}_1 + \Gamma_{jk}^2 \mathbf{x}_2 + b_{jk} \mathbf{n}}}. \quad (\text{B.13})$$

The vectors \mathbf{x}_{jk} can be summarized in a 4×2 -matrix as follows:

$$\frac{\partial^2 \mathbf{x}}{\partial \mathbf{q} \partial \mathbf{q}^\top} = \begin{pmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} \\ \mathbf{x}_{21} & \mathbf{x}_{22} \end{pmatrix} = \boldsymbol{\Gamma}_1 \otimes \mathbf{x}_1 + \boldsymbol{\Gamma}_2 \otimes \mathbf{x}_2 + \mathbf{B} \otimes \mathbf{n}, \quad (\text{B.14})$$

where

$$\boldsymbol{\Gamma}_i \stackrel{\text{def}}{=} \begin{pmatrix} \Gamma_{11}^i & \Gamma_{12}^i \\ \Gamma_{21}^i & \Gamma_{22}^i \end{pmatrix}$$

and

$$\mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

Multiplying (B.14) from the left by the matrix $(\mathbf{I}_2 \otimes \mathbf{n}^\top)$, due to $\mathbf{n}^\top \mathbf{x}_i = 0$ and $(\mathbf{I}_2 \otimes \mathbf{n}^\top)(\mathbf{B} \otimes \mathbf{n}) = \mathbf{B} \otimes (\mathbf{n}^\top \mathbf{n}) = \mathbf{B} \otimes 1 = \mathbf{B}$, one finally obtains

$$\underline{\underline{\mathbf{B} = (\mathbf{I}_2 \otimes \mathbf{n}^\top) \frac{\partial^2 \mathbf{x}}{\partial \mathbf{q} \partial \mathbf{q}^\top} = \begin{pmatrix} \mathbf{n}^\top \mathbf{x}_{11} & \mathbf{n}^\top \mathbf{x}_{12} \\ \mathbf{n}^\top \mathbf{x}_{21} & \mathbf{n}^\top \mathbf{x}_{22} \end{pmatrix}}}. \quad (\text{B.15})$$

i.e. the 2×2 -matrix \mathbf{B} is symmetric. This matrix is called the *Second Fundamental Form* of the surface. The *First Quadratic Fundamental Form* is given by the connection

$$\underline{\underline{\mathbf{G}}} = \frac{\partial \mathbf{x}^T}{\partial \mathbf{q}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T}, \quad (\text{B.16})$$

or more precisely, by the right-hand side of the equation for the squared line element ds of the surface:

$$ds^2 = d\mathbf{q}^T \underline{\underline{\mathbf{G}}} d\mathbf{q}.$$

Gauss designates, as is still in elementary geometry today, the elements of the matrix \mathbf{G} in this way:

$$\mathbf{G} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}.$$

While \mathbf{G} therefore plays a critical role in determining the length of a curve in an area, \mathbf{B} is, as we will see later, decisively involved in the determination of the curvature of a surface. A further representation of the matrix \mathbf{B} is obtained from the derivative of the scalar product of the mutually orthogonal vectors \mathbf{n} and \mathbf{x}_i :

$$\mathbf{n}^T \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} = \mathbf{0}^T;$$

because this derivative is, according to (A.73),

$$\frac{\partial}{\partial \mathbf{q}} \left(\mathbf{n}^T \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} \right) = \frac{\partial \mathbf{n}^T}{\partial \mathbf{q}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} + (\mathbf{I}_2 \otimes \mathbf{n}^T) \frac{\partial^2 \mathbf{x}}{\partial \mathbf{q} \partial \mathbf{q}^T} = \mathbf{0}_{2 \times 2}.$$

With (B.15) we obtain

$$\underline{\underline{\mathbf{B}}} = - \frac{\partial \mathbf{n}^T}{\partial \mathbf{q}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T}, \quad (\text{B.17})$$

i.e. a further interesting form for the matrix \mathbf{B}

$$\underline{\underline{\mathbf{B}}} = - \begin{pmatrix} \mathbf{n}_1^T \mathbf{x}_1 & \mathbf{n}_1^T \mathbf{x}_2 \\ \mathbf{n}_2^T \mathbf{x}_1 & \mathbf{n}_2^T \mathbf{x}_2 \end{pmatrix}. \quad (\text{B.18})$$

The two equations (B.9) can be summarized to a matrix equation as follows:

$$[\mathbf{n}_1 | \mathbf{n}_2] = \frac{\partial \mathbf{n}}{\partial \mathbf{q}^T} = \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} \begin{pmatrix} -b_1^1 & -b_2^1 \\ -b_1^2 & -b_2^2 \end{pmatrix} = \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} \cdot \underline{\underline{\mathbf{B}}}. \quad (\text{B.19})$$

This used in (B.18) together with (B.16) yields

$$\mathbf{B} = -\underline{\underline{\mathbf{B}}}^T \cdot \frac{\partial \mathbf{x}^T}{\partial \mathbf{q}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} = -\underline{\underline{\mathbf{B}}}^T \underline{\underline{\mathbf{G}}}.$$

or transposed

$$\underline{\underline{\underline{B}}} = -\underline{\underline{\underline{G}}}\overline{\underline{\underline{B}}}, \quad (\text{B.20})$$

since \mathbf{B} and \mathbf{G} are both symmetric matrices. Our goal remains to show that the Gaussian curvature depends only on the g_{ij} 's and their derivatives with respect to q_k , i.e., according to (B.10), one must show that for the matrix $\overline{\mathbf{B}}$

$$\kappa(\mathbf{q}) = \det \overline{\mathbf{B}}$$

is valid. We first examine the matrix \mathbf{G} . For this purpose, its elements are differentiated:

$$\frac{\partial g_{ij}}{\partial q_k} = \mathbf{x}_i^\top \mathbf{x}_j + \mathbf{x}_{jk}^\top \mathbf{x}_i.$$

With (B.13) we obtain

$$\begin{aligned} \mathbf{x}_j^\top \mathbf{x}_{ik} &= \Gamma_{ik}^1 \mathbf{x}_j^\top \mathbf{x}_1 + \Gamma_{ik}^2 \mathbf{x}_j^\top \mathbf{x}_2 + b_{ik} \mathbf{x}_j^\top \mathbf{n} \\ &= \Gamma_{ik}^1 g_{j1} + \Gamma_{ik}^2 g_{j2}. \end{aligned}$$

If we define

$$\check{\Gamma}_{ik}^j \stackrel{\text{def}}{=} \Gamma_{ik}^1 g_{j1} + \Gamma_{ik}^2 g_{j2}$$

and assemble all four components into a matrix $\check{\Gamma}_j$, we obtain

$$\begin{pmatrix} \check{\Gamma}_{j1}^1 & \check{\Gamma}_{j1}^2 \\ \check{\Gamma}_{j2}^1 & \check{\Gamma}_{j2}^2 \end{pmatrix} = \begin{pmatrix} \Gamma_{j1}^1 & \Gamma_{j1}^2 \\ \Gamma_{j2}^1 & \Gamma_{j2}^2 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \boldsymbol{\Gamma}_j \mathbf{G}^\top \quad (\text{B.21})$$

or, because of $\mathbf{G}^\top = \mathbf{G}$,

$$\check{\Gamma}_j = \boldsymbol{\Gamma}_j \mathbf{G}. \quad (\text{B.22})$$

It is therefore true that

$$\frac{\partial g_{ij}}{\partial q_k} = \check{\Gamma}_{ik}^j + \check{\Gamma}_{jk}^i. \quad (\text{B.23})$$

With the following expression of three different derivatives, one obtains

$$\frac{\partial g_{ij}}{\partial q_k} + \frac{\partial g_{ik}}{\partial q_j} - \frac{\partial g_{jk}}{\partial q_i} = \check{\Gamma}_{ik}^j + \check{\Gamma}_{jk}^i + \check{\Gamma}_{ij}^k + \check{\Gamma}_{jk}^i - \check{\Gamma}_{ik}^j - \check{\Gamma}_{ji}^k, \quad (\text{B.24})$$

so

$$\check{\Gamma}_{jk}^i = \frac{1}{2} \left(\frac{\partial g_{ij}}{\partial q_k} + \frac{\partial g_{ik}}{\partial q_j} - \frac{\partial g_{jk}}{\partial q_i} \right). \quad (\text{B.25})$$

Multiplying (B.22) from the right with

$$\mathbf{G}^{-1} \stackrel{\text{def}}{=} \begin{pmatrix} g_{11}^{[-1]} & g_{12}^{[-1]} \\ g_{21}^{[-1]} & g_{22}^{[-1]} \end{pmatrix},$$

one gets the relation

$$\Gamma_j = \check{\Gamma}_j \mathbf{G}^{-1}, \quad (\text{B.26})$$

i.e. element by element

$$\Gamma_{jk}^\ell = \frac{1}{2} \sum_i g_{i\ell}^{[-1]} \check{\Gamma}_{jk}^i, \quad (\text{B.27})$$

so with (B.25)

$$\boxed{\Gamma_{jk}^\ell = \frac{1}{2} \sum_i g_{i\ell}^{[-1]} \left(\frac{\partial g_{ij}}{\partial q_k} + \frac{\partial g_{ik}}{\partial q_j} - \frac{\partial g_{jk}}{\partial q_i} \right)}. \quad (\text{B.28})$$

This clarifies the relationship of the Christoffel-symbols Γ_{jk}^ℓ with the g_{ij} and their derivatives. Now the direct relationship of these variables with the Gaussian curvature κ has to be made. One gets this finally by repeated differentiation of \mathbf{x}_{jk} with respect to q^ℓ :

$$\begin{aligned} \mathbf{x}_{jkl} &\stackrel{\text{def}}{=} \frac{\partial \mathbf{x}_{jk}}{\partial q^\ell} = \sum_i \frac{\partial \Gamma_{jk}^i}{\partial q^\ell} \mathbf{x}_i + \sum_i \Gamma_{j\ell}^i \mathbf{x}_{i\ell} + \frac{\partial b_{jk}}{\partial q^\ell} \mathbf{n} + b_{jk} \mathbf{n}_\ell \\ &= \sum_i \left(\frac{\partial \Gamma_{jk}^i}{\partial q^\ell} + \Gamma_{j\ell}^i \Gamma_{p\ell}^i - b_{jk} b_\ell^i \right) \mathbf{x}_i + \left(\frac{\partial b_{jk}}{\partial q^\ell} + \sum_p \Gamma_{jk}^p b_{p\ell} \right) \mathbf{n}. \end{aligned} \quad (\text{B.29})$$

Interchanging in (B.29) k and ℓ , we obtain

$$\mathbf{x}_{j\ell k} = \sum_i \left(\frac{\partial \Gamma_{j\ell}^i}{\partial q^k} + \Gamma_{jk}^i \Gamma_{p\ell}^i - b_{j\ell} b_k^i \right) \mathbf{x}_i + \left(\frac{\partial b_{j\ell}}{\partial q^k} + \sum_p \Gamma_{j\ell}^p b_{pk} \right) \mathbf{n}. \quad (\text{B.30})$$

Subtracting the two third-order derivatives, we obtain

$$\mathbf{0} = \mathbf{x}_{j\ell k} - \mathbf{x}_{jkl} = \sum_i [R_{jk}^{i\ell} - (b_{j\ell} b_k^i - b_{jk} b_\ell^i)] \mathbf{x}_i + (\dots) \mathbf{n}, \quad (\text{B.31})$$

with

$$R_{jk}^{i\ell} \stackrel{\text{def}}{=} \frac{\partial \Gamma_{j\ell}^i}{\partial q^k} - \frac{\partial \Gamma_{jk}^i}{\partial q^\ell} + \sum_p \Gamma_{j\ell}^p \Gamma_{p\ell}^i - \sum_p \Gamma_{jk}^p \Gamma_{pk}^i. \quad (\text{B.32})$$

Since the vectors $\mathbf{x}_1, \mathbf{x}_2$ and \mathbf{n} are linearly independent, the square bracket in (B.31) must be zero, which implies

$$R_{jk}^{i\ell} = b_{j\ell}b_k^i - b_{jk}b_\ell^i. \quad (\text{B.33})$$

Defining

$$\check{R}_{jk}^{i\ell} = \sum_i g_{ih} R_{jk}^{i\ell}, \quad (\text{B.34})$$

one gets

$$\check{R}_{jk}^{i\ell} = g_{1h}b_{j\ell}b_k^1 - g_{1h}b_{jk}b_\ell^1 + g_{2h}b_{j\ell}b_k^2 - g_{2h}b_{jk}b_\ell^2 = b_{j\ell}b_{kh} - b_{jk}b_{\ell h}. \quad (\text{B.35})$$

In particular,

$$\boxed{\check{R}_{12}^{12} = b_{22}b_{11} - b_{21}b_{21} = \det \mathbf{B}.} \quad (\text{B.36})$$

It is therefore true that

$$\kappa(\mathbf{q}) = \det \bar{\mathbf{B}} = \frac{\det \mathbf{B}}{\det \mathbf{G}} = \frac{\check{R}_{12}^{12}}{g},$$

$$\boxed{\kappa(\mathbf{q}) = \frac{\check{R}_{12}^{12}}{g}}, \quad (\text{B.37})$$

which has finally proved the Theorema Egregium because, due to (B.34), $\check{R}_{jk}^{i\ell}$ depends on $R_{jk}^{i\ell}$; according to (B.32), $R_{jk}^{i\ell}$ only depends on Γ_{ij}^k and their derivatives and, in accordance with (B.25), the Γ_{ij}^k 's depend only on the g_{ik} 's and their derivatives. In the form

$$\boxed{\kappa(\mathbf{q}) = \frac{\det \mathbf{B}}{\det \mathbf{G}}}, \quad (\text{B.38})$$

the paramount importance of the two fundamental forms is expressed.

Remarks

1. Euclidean geometry of is based on a number of Axioms that require no proofs. One is the parallel postulate stating that to every line one can draw through a point not belonging to it one and only one other line which lies in the same plane and does not intersect the former line. This axiom is replaced in the *hyperbolic* geometry in that it admits infinitely many parallels. An example is the surface of a hyperboloid. In the *elliptical* geometry, for example, on the surface of an ellipsoid and, as a special case, on a spherical surface, there are absolutely no parallels because all great circles, which are here the “straight lines”, meet in

two points. In Euclidean geometry, the distance between two points with the Cartesian coordinates x_1, x_2, x_3 and $x_1 + dx_1, x_2 + dx_2, x_3 + dx_3$ is simply

$$ds = \sqrt{dx_1^2 + dx_2^2 + dx_3^2},$$

and in the other two geometries this formula is replaced by

$$ds^2 = a_1 dx_1^2 + a_2 dx_2^2 + a_3 dx_3^2,$$

where the coefficients a_i are certain simple functions of x_i , in the hyperbolic case, of course, different than in the elliptic case. A convenient analytical representation of curved surfaces is the above used Gaussian parameter representation $\mathbf{x} = \mathbf{x}(q_1, q_2)$, where Gauss attaches as curved element:

$$ds^2 = E dq_1^2 + 2F dq_1 dq_2 + G dq_2^2.$$

As an example, we introduce the Gauss-specific parameter representation of the unit sphere, with $\theta = q_1$ and $\varphi = q_2$:

$$x_1 = \sin \theta \cos \varphi, \quad x_2 = \sin \theta \sin \varphi, \quad x_3 = \cos \theta.$$

For the arc element of the unit sphere, we obtain

$$ds^2 = d\theta^2 + \sin^2 \theta (d\varphi)^2.$$

Riemann generalized the Gaussian theory of surfaces, which is valid for two-dimensional surfaces in three-dimensional spaces, to p -dimensional hypersurfaces in n -dimensional spaces, i.e. where

$$\mathbf{x} = \mathbf{x}(q_1, \dots, q_p) \in \mathbb{R}^n$$

is a point on the hypersurface. He made in addition the fundamentally important step, to set up a homogeneous quadratic function of dq_i with arbitrary functions of the q_i as coefficients, as the square of the line elements (quadratic form)

$$ds^2 = \sum_{ik} g_{ik} dq_i dq_k = \mathbf{dq}^\top \mathbf{G} \mathbf{dq}.$$

2. The above-occurring $R_{jk}^{i\ell}$ can be used as matrix elements of the 4×4 -matrix \mathbf{R} , the Riemannian *Curvature Matrix*, to be constructed as a block matrix as follows:

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}^{11} & \mathbf{R}^{12} \\ \mathbf{R}^{21} & \mathbf{R}^{22} \end{pmatrix},$$

where the 2×2 sub-matrices have the form:

$$\mathbf{R}^{i\ell} = \begin{pmatrix} R_{11}^{i\ell} & R_{12}^{i\ell} \\ R_{21}^{i\ell} & R_{22}^{i\ell} \end{pmatrix}.$$

In particular, \check{R}_{12}^{12} is the element in the top right corner of the matrix $\check{\mathbf{R}} = \mathbf{G}\mathbf{R}$.

3. Expanding the representation of $\mathbf{x}(\mathbf{q} + \Delta\mathbf{q})$ in a Taylor series, one obtains

$$\mathbf{x}(\mathbf{q} + \Delta\mathbf{q}) = \mathbf{x}(\mathbf{q}) + \sum_i \mathbf{x}_i \Delta q_i + \frac{1}{2} \sum_{i,k} \mathbf{x}_{ik} \Delta q_i \Delta q_k + \sigma(3).$$

Subtracting $\mathbf{x}(\mathbf{q})$ on both sides of this equation and multiplying the result from the left with the transposed normal vector \mathbf{n}^\top , we obtain

$$\begin{aligned} \mathbf{n}^\top [\mathbf{x}(\mathbf{q} + \Delta\mathbf{q}) - \mathbf{x}(\mathbf{q})] &= \sum_i \underbrace{\mathbf{n}^\top \mathbf{x}_i}_{0} \Delta q_i + \frac{1}{2} \sum_{i,k} \underbrace{\mathbf{n}^\top \mathbf{x}_{ik}}_{b_{ik}} \Delta q_i \Delta q_k + \sigma(3) \\ &= \mathbf{n}^\top \Delta \mathbf{x}(\mathbf{q}) \stackrel{\text{def}}{=} \Delta \ell. \end{aligned}$$

Thus

$$\mathrm{d}\ell \approx \frac{1}{2} \sum_{i,k} b_{ik} \mathrm{d}q_i \mathrm{d}q_k.$$

The coefficients of the second fundamental form, i.e. the elements of the matrix \mathbf{B} , Gauss denotes by L, M and N . Then the distance $\mathrm{d}\ell$ of the point $\mathbf{x}(q_1 + \mathrm{d}q_1, q_2 + \mathrm{d}q_2)$ to the tangent surface at the point $\mathbf{x}(q_1, q_2)$ is

$$\mathrm{d}\ell \approx \frac{1}{2} (L \mathrm{d}q_1^2 + 2M \mathrm{d}q_1 \mathrm{d}q_2 + N \mathrm{d}q_2^2).$$

The *normal curvature* κ of a surface at a given point P and in a given direction \mathbf{q} is defined as

$$\kappa \stackrel{\text{def}}{=} \frac{L \mathrm{d}q_1^2 + 2M \mathrm{d}q_1 \mathrm{d}q_2 + N \mathrm{d}q_2^2}{E \mathrm{d}q_1^2 + 2F \mathrm{d}q_1 \mathrm{d}q_2 + G \mathrm{d}q_2^2}. \quad (\text{B.39})$$

The so-defined normal curvature depends, in general, on the chosen direction $\mathrm{d}\mathbf{q}$. Those directions, in which the normal curvatures at a given point assume an extreme value, are named the *main* directions of the surface at this point. As long as we examine real surfaces, the quadratic differential form $E \mathrm{d}q_1^2 + 2F \mathrm{d}q_1 \mathrm{d}q_2 + G \mathrm{d}q_2^2$ is positive definite, i.e. it is always positive for $\mathrm{d}\mathbf{q} \neq \mathbf{0}$. Thus the sign of the curvature depends only on the quadratic differential form $L \mathrm{d}q_1^2 + 2M \mathrm{d}q_1 \mathrm{d}q_2 + N \mathrm{d}q_2^2$ in the numerator of (B.39). There are three cases:

- (a) $LN - M^2 > 0$, i.e. \mathbf{B} is positive definite, and the numerator retains the same sign, in each direction one is looking. Such a point is called an *elliptical* point. An example is any point on an ellipsoid, in particular, of course, on a sphere.
- (b) $LN - M^2 = 0$, i.e. \mathbf{B} is semi-definite. The surface behaves at this point as at an elliptical point except in one direction where $\kappa = 0$. This point is called *parabolic*. An example is any point on a cylinder.

- (c) $LN - M^2 < 0$, i.e. \mathbf{B} is indefinite. The numerator does not keep the same sign for all directions. Such a point is called *hyperbolic*, or a *saddle point*. An example is a point on a hyperbolic paraboloid.

Dividing the numerator and the denominator in (B.39) by dq_2 and introducing $dq_1/dq_2 \stackrel{\text{def}}{=} \lambda$, we obtain

$$\kappa(\lambda) = \frac{L + 2M\lambda + N\lambda^2}{E + 2F\lambda + G\lambda^2} \quad (\text{B.40})$$

and from this the extreme values from

$$\frac{d\kappa}{d\lambda} = 0$$

as those satisfying

$$(E + 2F\lambda + G\lambda^2)(M + N\lambda) - (L + 2M\lambda + N\lambda^2)(F + G\lambda) = 0. \quad (\text{B.41})$$

In this case, the resulting expression for κ is

$$\kappa = \frac{L + 2M\lambda + N\lambda^2}{E + 2F\lambda + G\lambda^2} = \frac{M + N\lambda}{F + G\lambda}. \quad (\text{B.42})$$

Since furthermore

$$E + 2F\lambda + G\lambda^2 = (E + F\lambda) + \lambda(F + G\lambda)$$

and

$$L + 2M\lambda + N\lambda^2 = (L + M\lambda) + \lambda(M + N\lambda),$$

(B.40) can be transformed into the simpler form

$$\kappa = \frac{L + M\lambda}{E + F\lambda}. \quad (\text{B.43})$$

From this the two equations for κ follow:

$$(\kappa E - L) + (\kappa F - M)\lambda = 0,$$

$$(\kappa F - M) + (\kappa G - N)\lambda = 0.$$

These equations are simultaneously satisfied if and only if

$$\det \begin{pmatrix} \kappa E - L & \kappa F - M \\ \kappa F - M & \kappa G - N \end{pmatrix} = 0. \quad (\text{B.44})$$

This can also be written as

$$\det(\kappa \mathbf{G} - \mathbf{B}) = 0. \quad (\text{B.45})$$

This is the solvability condition for the eigenvalue equation

$$\kappa \mathbf{G} - \mathbf{B} = \mathbf{0},$$

which can be transformed into

$$\kappa \mathbf{I} - \mathbf{G}^{-1} \mathbf{B} = \mathbf{0}. \quad (\text{B.46})$$

This results in a quadratic equation for κ . The two solutions are called the *principal curvatures* and are denoted as κ_1 and κ_2 . The Gaussian curvature κ of a surface at a given point is the product of the principal curvatures κ_1 and κ_2 of the surface in this point. According to Vieta's root theorem, the product of the solutions is equal to the determinant of the matrix $\mathbf{G}^{-1} \mathbf{B}$, so finally,

$$\kappa = \kappa_1 \kappa_2 = \det(\mathbf{G}^{-1} \mathbf{B}) = \frac{\det \mathbf{B}}{\det \mathbf{G}} = \frac{LN - M^2}{EG - F^2}.$$

Appendix C

Geodesic Deviation

Geodesics are the lines of general manifolds along which, for example, free particles move. In a flat space the relative velocity of each pair of particles is constant, so that their relative acceleration is always equal to zero. Generally, due to the curvature of space, the relative acceleration is not equal to zero.

The curvature of a surface can be illustrated as follows [21]. Suppose there are two ants on an apple which leave a starting line at the same time and follow with the same speed geodesics which are initially perpendicular to the start line. Initially, their paths are parallel, but, due to the curvature of the apple, they are approaching each other from the beginning. Their distance ξ from one another is not constant, i.e., in general, the relative acceleration of the ants moving on geodesics with constant velocity is not equal zero if the area over which they move is curved. So the curvature can be indirectly perceived through the so-called *geodesic deviation* ξ .

The two neighboring geodesics $x(u)$ and $\check{x}(u)$ have the distance

$$\xi(u) \stackrel{\text{def}}{=} \check{x}(u) - x(u), \quad (\text{C.1})$$

where u is the proper time or distance.

The mathematical descriptions of these geodesics are

$$\ddot{\check{x}} + (\mathbf{I}_4 \otimes \dot{\check{x}}^\top) \check{\Gamma} \dot{\check{x}} = \mathbf{0}, \quad (\text{C.2})$$

$$\ddot{x} + (\mathbf{I}_4 \otimes \dot{x}^\top) \Gamma \dot{x} = \mathbf{0}. \quad (\text{C.3})$$

The Christoffel-matrix $\check{\Gamma}$ is approximated by

$$\check{\Gamma} \approx \Gamma + \frac{\partial \Gamma}{\partial x^\top} (\xi \otimes \mathbf{I}_4). \quad (\text{C.4})$$

Subtracting (C.3) from (C.2) and considering (C.1) and (C.4), one obtains

$$\ddot{\xi} + (\mathbf{I}_4 \otimes \dot{\check{x}}^\top) \check{\Gamma} \dot{\check{x}} - (\mathbf{I}_4 \otimes \dot{x}^\top) \Gamma \dot{x} + (\mathbf{I}_4 \otimes \dot{\check{x}}^\top) \left(\frac{\partial \Gamma}{\partial x^\top} (\xi \otimes \mathbf{I}_4) \right) \dot{x} = \mathbf{0}. \quad (\text{C.5})$$

With $\dot{\tilde{x}} = \dot{\xi} + \dot{x}$ and neglecting quadratic and higher powers of ξ and $\dot{\xi}$, one obtains from (C.5)

$$\ddot{\xi} + (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{x} + (\mathbf{I}_4 \otimes \dot{x}^\top) \boldsymbol{\Gamma} \dot{\xi} + (\mathbf{I}_4 \otimes \dot{x}^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\xi \otimes \mathbf{I}_4) \dot{x} = \mathbf{0}. \quad (\text{C.6})$$

Hence

$$\frac{D\xi}{du} = \dot{\xi} + (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x} \quad (\text{C.7})$$

and

$$\begin{aligned} \frac{D^2\xi}{du^2} &= \frac{D}{du} (\dot{\xi} + (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x}) \\ &= \ddot{\xi} + \frac{d}{du} \{(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x}\} + (\mathbf{I}_4 \otimes [\dot{\xi} + (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x}]^\top) \boldsymbol{\Gamma} \dot{x} \\ &= \ddot{\xi} + \frac{d}{du} \{(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x}\} + (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{x} + (\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x}]^\top) \boldsymbol{\Gamma} \dot{x}. \end{aligned} \quad (\text{C.8})$$

For the second term, by (C.3), one gets

$$\begin{aligned} \frac{d}{du} \{(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x}\} &= (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{x} + (\mathbf{I}_4 \otimes \xi^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\dot{x} \otimes \mathbf{I}_4) \dot{x} + (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \ddot{x} \\ &= (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{x} + (\mathbf{I}_4 \otimes \xi^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\dot{x} \otimes \mathbf{I}_4) \dot{x} \\ &\quad - (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} (\mathbf{I}_4 \otimes \dot{x}^\top) \boldsymbol{\Gamma} \dot{x}. \end{aligned} \quad (\text{C.9})$$

Equation (C.9) used in (C.8) yields

$$\begin{aligned} \frac{D^2\xi}{du^2} &= \ddot{\xi} + (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{x} + (\mathbf{I}_4 \otimes \xi^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\dot{x} \otimes \mathbf{I}_4) \dot{x} - (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} (\mathbf{I}_4 \otimes \dot{x}^\top) \boldsymbol{\Gamma} \dot{x} \\ &\quad + (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{x} + (\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{x}]^\top) \boldsymbol{\Gamma} \dot{x}. \end{aligned} \quad (\text{C.10})$$

Remark Since the sub-matrices $\boldsymbol{\Gamma}_i$ are symmetric, it is generally true that

$$(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \mathbf{b} = (\mathbf{I}_4 \otimes \mathbf{b}^\top) \boldsymbol{\Gamma} \mathbf{a}. \quad (\text{C.11})$$

In addition, one has $(\mathbf{I}_4 \otimes \mathbf{a}^\top) \boldsymbol{\Gamma} \mathbf{b} = \overline{\boldsymbol{\Gamma}}(\mathbf{I}_4 \otimes \mathbf{a}) \mathbf{b} = \overline{\boldsymbol{\Gamma}}(\mathbf{b} \otimes \mathbf{a})$ and $(\mathbf{I}_4 \otimes \mathbf{b}^\top) \boldsymbol{\Gamma} \mathbf{a} = \overline{\boldsymbol{\Gamma}}(\mathbf{I}_4 \otimes \mathbf{b}) \mathbf{a} = \overline{\boldsymbol{\Gamma}}(\mathbf{a} \otimes \mathbf{b})$, thus, due to (C.11),

$$\overline{\boldsymbol{\Gamma}}(\mathbf{b} \otimes \mathbf{a}) = \overline{\boldsymbol{\Gamma}}(\mathbf{a} \otimes \mathbf{b}). \quad (\text{C.12})$$

With (C.11), one has from (C.10)

$$\ddot{\xi} + (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{x} + (\mathbf{I}_4 \otimes \dot{x}^\top) \boldsymbol{\Gamma} \dot{\xi}$$

$$\begin{aligned}
&= \frac{D^2\xi}{du^2} - (\mathbf{I}_4 \otimes \xi^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\dot{\mathbf{x}} \otimes \mathbf{I}_4) \dot{\mathbf{x}} \\
&\quad + (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} (\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}} - (\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}]^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}. \tag{C.13}
\end{aligned}$$

For $(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} (\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}$ one can write

$$(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} (\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}} = \underline{\underline{(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \bar{\boldsymbol{\Gamma}} (\mathbf{I}_4 \otimes \dot{\mathbf{x}}) \dot{\mathbf{x}}}}, \tag{C.14}$$

and the expression $(\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}]^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}$ can be rewritten as

$$\begin{aligned}
&(\mathbf{I}_4 \otimes [(\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}]^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}} \\
&= \bar{\boldsymbol{\Gamma}} (\mathbf{I}_4 \otimes (\mathbf{I}_4 \otimes \xi^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}}) \dot{\mathbf{x}} \\
&= \bar{\boldsymbol{\Gamma}} (\mathbf{I}_{16} \otimes \xi^\top) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma} \dot{\mathbf{x}}) \dot{\mathbf{x}} = \underline{\underline{(\mathbf{I}_4 \otimes \xi^\top) (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma}) (\mathbf{I}_4 \otimes \dot{\mathbf{x}}) \dot{\mathbf{x}}}}. \tag{C.15}
\end{aligned}$$

With (C.14) (in somewhat modified form) and (C.15) one obtains for (C.13)

$$\begin{aligned}
&\ddot{\xi} + (\mathbf{I}_4 \otimes \dot{\xi}^\top) \boldsymbol{\Gamma} \dot{\mathbf{x}} + (\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \boldsymbol{\Gamma} \dot{\xi} \\
&= \frac{D^2\xi}{du^2} - (\mathbf{I}_4 \otimes \xi^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\dot{\mathbf{x}} \otimes \mathbf{I}_4) \dot{\mathbf{x}} \\
&\quad + (\mathbf{I}_4 \otimes \xi^\top) [\boldsymbol{\Gamma} \bar{\boldsymbol{\Gamma}} - (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma})] (\dot{\mathbf{x}} \otimes \dot{\mathbf{x}}). \tag{C.16}
\end{aligned}$$

Equation (C.16) used in (C.6) provides

$$\begin{aligned}
\frac{D^2\xi}{du^2} &= -(\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\xi \otimes \mathbf{I}_4) \dot{\mathbf{x}} + (\mathbf{I}_4 \otimes \xi^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\dot{\mathbf{x}} \otimes \mathbf{I}_4) \dot{\mathbf{x}} \\
&\quad + (\mathbf{I}_4 \otimes \xi^\top) [\boldsymbol{\Gamma} \bar{\boldsymbol{\Gamma}} - (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma})] (\dot{\mathbf{x}} \otimes \dot{\mathbf{x}}). \tag{C.17}
\end{aligned}$$

As the 16×16 -matrix $\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top}$ is symmetric, the first term of the right-hand side can be transformed as follows:

$$\begin{aligned}
(\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\xi \otimes \mathbf{I}_4) \dot{\mathbf{x}} &= (\mathbf{I}_4 \otimes \dot{\mathbf{x}}^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} U_{4 \times 4} (\mathbf{I}_4 \otimes \xi) \dot{\mathbf{x}} \\
&= (\mathbf{I}_4 \otimes \xi^\top) \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} U_{4 \times 4} (\mathbf{I}_4 \otimes \dot{\mathbf{x}}) \dot{\mathbf{x}}.
\end{aligned}$$

This in (C.17) provides

$$\begin{aligned}
\frac{D^2\xi}{du^2} &= (\mathbf{I}_4 \otimes \xi^\top) \left[\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} - \frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} U_{4 \times 4} \right] (\dot{\mathbf{x}} \otimes \mathbf{I}_4) \dot{\mathbf{x}} \\
&\quad + (\mathbf{I}_4 \otimes \xi^\top) [\boldsymbol{\Gamma} \bar{\boldsymbol{\Gamma}} - (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4) (\mathbf{I}_4 \otimes \boldsymbol{\Gamma})] (\dot{\mathbf{x}} \otimes \dot{\mathbf{x}}), \tag{C.18}
\end{aligned}$$

and finally,

$$\frac{D^2 \xi}{du^2} = (\mathbf{I}_4 \otimes \xi^\top) \underbrace{\left[\frac{\partial \boldsymbol{\Gamma}}{\partial \mathbf{x}^\top} (\mathbf{I}_{16} - \mathbf{U}_{4 \times 4}) + (\boldsymbol{\Gamma} \bar{\boldsymbol{\Gamma}} - (\bar{\boldsymbol{\Gamma}} \otimes \mathbf{I}_4)(\mathbf{I}_4 \otimes \boldsymbol{\Gamma})) \right]}_{-\mathbf{R}} (\dot{\mathbf{x}} \otimes \dot{\mathbf{x}}). \quad (\text{C.19})$$

Using a slightly modified Riemannian curvature matrix \mathbf{R} , we finally obtain for the dynamic behavior of the geodesic deviation

$$\frac{D^2 \xi}{du^2} + (\mathbf{I}_4 \otimes \xi^\top) \mathbf{R} (\dot{\mathbf{x}} \otimes \dot{\mathbf{x}}) = \mathbf{0}. \quad (\text{C.20})$$

In a flat manifold, i.e. in a gravity-free space one has $\mathbf{R} \equiv \mathbf{0}$ and in Cartesian coordinates $D/du = d/du$ so that (C.20) reduces to the equation $d^2 \xi / du^2 = \mathbf{0}$ whose solution is the linear relationship $\xi(u) = \dot{\xi}_0 \cdot u + \xi_0$. If $\mathbf{R} \neq \mathbf{0}$, gravity exists and the solution of (C.20) is nonlinear, curved.

Appendix D

Another Ricci-Matrix

The Ricci-matrix \mathbf{R}_{Ric} is now defined as the sum of the sub-matrices on the main diagonal of \mathbf{R}

$$\mathbf{R}_{\text{Ric}} \stackrel{\text{def}}{=} \sum_{v=0}^3 \mathbf{R}^{vv}. \quad (\text{D.1})$$

Analogously, we define

$$\check{\mathbf{R}}_{\text{Ric}} \stackrel{\text{def}}{=} \sum_{v=0}^3 \check{\mathbf{R}}^{vv}. \quad (\text{D.2})$$

From (2.194) it can immediately be read that the Ricci-matrix $\check{\mathbf{R}}_{\text{Ric}}$ is *symmetric* because $\check{R}_{\alpha\beta}^{\gamma\gamma} = \check{R}_{\beta\alpha}^{\gamma\gamma}$. It is also true that

$$\mathbf{R} = (\mathbf{G}^{-1} \otimes \mathbf{I}_4) \check{\mathbf{R}},$$

so

$$\mathbf{R}^{\gamma\delta} = (\mathbf{g}_\gamma^{-T} \otimes \mathbf{I}_4) \check{\mathbf{R}}^\delta = \sum_{v=0}^3 g_{\gamma v}^{[-1]} \check{\mathbf{R}}^{v\delta}, \quad (\text{D.3})$$

where \mathbf{g}_γ^{-T} is the γ th row of \mathbf{G}^{-1} and $\check{\mathbf{R}}^\delta$ is the matrix consisting of the sub-matrices in the δ th block column of $\check{\mathbf{R}}$, i.e. the matrix elements are

$$R_{\alpha\beta}^{\gamma\delta} = \sum_{v=0}^3 g_{\gamma v}^{[-1]} \check{R}_{\alpha\beta}^{v\delta}. \quad (\text{D.4})$$

With the help of (D.3), the Ricci-matrix is obtained as

$$\mathbf{R}_{\text{Ric}} = \sum_\gamma \mathbf{R}^{\gamma\gamma} = \sum_\gamma \sum_v g_{\gamma v}^{[-1]} \check{\mathbf{R}}^{vv}, \quad (\text{D.5})$$

i.e. for the components one has

$$R_{\text{Ric},\alpha\beta} = \sum_{\gamma} \sum_{\nu} g_{\gamma\nu}^{[-1]} \check{R}_{\alpha\beta}^{\nu\gamma}, \quad (\text{D.6})$$

or with (2.173)

$$R_{\text{Ric},\alpha\beta} = \sum_{\gamma} \sum_{\nu} g_{\gamma\nu}^{[-1]} \check{R}_{\nu\gamma}^{\alpha\beta}. \quad (\text{D.7})$$

The *curvature scalar* R is obtained from the Ricci-matrix by taking the trace

$$R \stackrel{\text{def}}{=} \sum_{\alpha} R_{\text{Ric},\alpha\alpha} = \sum_{\alpha} \sum_{\gamma} \sum_{\nu} g_{\gamma\nu}^{[-1]} \check{R}_{\nu\gamma}^{\alpha\alpha} = \sum_{\gamma} \sum_{\nu} g_{\gamma\nu}^{[-1]} \check{R}_{\text{Ric},\nu\gamma}. \quad (\text{D.8})$$

Conversely, we obtain a corresponding relationship

$$\check{R}_{\alpha\beta}^{\gamma\delta} = \sum_{\nu=0}^3 g_{\gamma\nu} R_{\alpha\beta}^{\nu\delta}. \quad (\text{D.9})$$

From (2.165) it directly follows that

$$R_{\text{Ric},\alpha\beta} = \sum_{\gamma=0}^3 \left(\frac{\partial}{\partial x_{\beta}} \Gamma_{\alpha\gamma}^{\gamma} - \frac{\partial}{\partial x_{\gamma}} \Gamma_{\alpha\beta}^{\gamma} + \sum_{\nu=0}^3 \Gamma_{\beta\nu}^{\gamma} \Gamma_{\gamma\alpha}^{\nu} - \sum_{\nu=0}^3 \Gamma_{\gamma\nu}^{\gamma} \Gamma_{\alpha\beta}^{\nu} \right) \quad (\text{D.10})$$

and from (2.168)

$$\check{R}_{\text{Ric},\alpha\beta} = \sum_{\gamma=0}^3 \left(\frac{\partial}{\partial x_{\beta}} \check{\Gamma}_{\alpha\gamma}^{\gamma} - \frac{\partial}{\partial x_{\gamma}} \check{\Gamma}_{\alpha\beta}^{\gamma} + \sum_{\nu=0}^3 \Gamma_{\alpha\beta}^{\nu} \check{\Gamma}_{\nu\gamma}^{\gamma} - \sum_{\nu=0}^3 \Gamma_{\alpha\gamma}^{\nu} \check{\Gamma}_{\gamma\beta}^{\beta} \right). \quad (\text{D.11})$$

Symmetry of the Ricci-Matrix R_{Ric} Even if \mathbf{R} itself is not symmetric, the from \mathbf{R} derived Ricci-matrix R_{Ric} is symmetric; this will be shown in the following. The symmetry will follow from the components equation (D.10) of the Ricci-matrix. One sees immediately that the second and fourth summands are symmetric in α and β .

The symmetry of the term $\sum_{\gamma=0}^3 \frac{\partial}{\partial x_{\beta}} \Gamma_{\alpha\gamma}^{\gamma}$ in α and β is not seen directly. This can be checked using the Laplace-expansion theorem for determinants.¹ Developing the determinant of \mathbf{G} along the γ th row yields

$$g \stackrel{\text{def}}{=} \det(\mathbf{G}) = g_{\gamma 1} A_{\gamma 1} + \cdots + g_{\gamma \beta} A_{\gamma \beta} + \cdots + g_{\gamma n} A_{\gamma n},$$

¹The sum of the products of all elements of a row (or column) with their adjuncts is equal to the determinant's value.

where $A_{\gamma\beta}$ is the element in the γ th row and β th column of the adjoint of \mathbf{G} . If $g_{\beta\gamma}^{[-1]}$ is the $(\beta\gamma)$ th element of the inverse of \mathbf{G} , then $g_{\beta\gamma}^{[-1]} = \frac{1}{g} A_{\gamma\beta}$, so $A_{\gamma\beta} = g g_{\beta\gamma}^{[-1]}$. Thus we obtain

$$\frac{\partial g}{\partial g_{\gamma\beta}} = A_{\gamma\beta} = g g_{\beta\gamma}^{[-1]},$$

or

$$\delta g = g g_{\beta\gamma}^{[-1]} \delta g_{\gamma\beta},$$

or

$$\frac{\partial g}{\partial x_\alpha} = g g_{\beta\gamma}^{[-1]} \frac{\partial g_{\gamma\beta}}{\partial x_\alpha},$$

i.e.

$$\frac{1}{g} \frac{\partial g}{\partial x_\alpha} = g_{\beta\gamma}^{[-1]} \frac{\partial g_{\gamma\beta}}{\partial x_\alpha}. \quad (\text{D.12})$$

Using (2.62), on the other hand, one has

$$\sum_{\gamma=0}^3 \Gamma_{\alpha\gamma}^\gamma = \sum_{\gamma=0}^3 \sum_{\beta=0}^3 \frac{g_{\beta\gamma}^{[-1]}}{2} \left(\frac{\partial g_{\gamma\beta}}{\partial x_\alpha} + \frac{\partial g_{\alpha\beta}}{\partial x_\gamma} - \frac{\partial g_{\alpha\gamma}}{\partial x_\beta} \right),$$

i.e. the last two summands cancel out and it remains to deal with

$$\sum_{\gamma=0}^3 \Gamma_{\alpha\gamma}^\gamma = \sum_{\gamma=0}^3 \sum_{\beta=0}^3 \frac{1}{2} g_{\beta\gamma}^{[-1]} \frac{\partial g_{\gamma\beta}}{\partial x_\alpha}.$$

It follows from (D.12) that

$$\underbrace{\sum_{\gamma=0}^3 \frac{\partial}{\partial x_\beta} \Gamma_{\alpha\gamma}^\gamma}_{\sum_{\gamma=0}^3 \sum_{\beta=0}^3 \frac{1}{\sqrt{|g|}} \frac{\partial^2 \sqrt{|g|}}{\partial x_\alpha \partial x_\beta}} = \sum_{\gamma=0}^3 \sum_{\beta=0}^3 \frac{1}{\sqrt{|g|}} \frac{\partial^2 \sqrt{|g|}}{\partial x_\alpha \partial x_\beta}. \quad (\text{D.13})$$

But this form is immediately seen symmetric in α and β .

Now it remains to show that the third term in (D.10) is symmetric. Its expression is

$$\sum_{\gamma=0}^3 \sum_{v=0}^3 \Gamma_{\beta v}^\gamma \Gamma_{\gamma\alpha}^v.$$

Now one can see that this term is symmetric because

$$\sum_{\gamma, v=0}^3 \Gamma_{\beta v}^\gamma \Gamma_{\gamma\alpha}^v = \sum_{\gamma, v=0}^3 \Gamma_{v\beta}^\gamma \Gamma_{\alpha\gamma}^v = \sum_{v, \gamma=0}^3 \Gamma_{\gamma\beta}^v \Gamma_{\alpha\gamma}^v.$$

All this shows that the Ricci-matrix \mathbf{R}_{Ric} is symmetric.

Divergence of the Ricci-Matrix R_{Ric} Multiplying the Bianchi-identities (2.199) in the form of

$$\frac{\partial}{\partial x_\kappa} R_{\alpha\beta}^{\nu\delta} + \frac{\partial}{\partial x_\beta} R_{\alpha\delta}^{\nu\kappa} + \frac{\partial}{\partial x_\delta} R_{\alpha\kappa}^{\nu\beta} = 0$$

with $g_{\gamma\nu}$ and summing over ν , we obtain in \mathcal{P} , since there $\frac{\partial G}{\partial x} = \mathbf{0}$,

$$\frac{\partial}{\partial x_\kappa} \sum_{\nu=0}^3 g_{\gamma\nu} R_{\alpha\beta}^{\nu\delta} + \frac{\partial}{\partial x_\beta} \sum_{\nu=0}^3 g_{\gamma\nu} R_{\alpha\delta}^{\nu\kappa} + \frac{\partial}{\partial x_\delta} \sum_{\nu=0}^3 g_{\gamma\nu} R_{\alpha\kappa}^{\nu\beta} = 0.$$

Combined with (D.9) this becomes

$$\underline{\underline{\frac{\partial}{\partial x_\kappa} \check{R}_{\alpha\beta}^{\gamma\delta} + \frac{\partial}{\partial x_\beta} \check{R}_{\alpha\delta}^{\gamma\kappa} + \frac{\partial}{\partial x_\delta} \check{R}_{\alpha\kappa}^{\gamma\beta} = 0.}}$$
(D.14)

The second term, according to (2.172), may be written as

$$-\frac{\partial}{\partial x_\beta} \check{R}_{\alpha\kappa}^{\gamma\delta}.$$

Substituting now $\gamma = \delta$ and summing over γ , one obtains

$$\frac{\partial}{\partial x_\kappa} \check{R}_{\text{Ric},\alpha\beta} - \frac{\partial}{\partial x_\beta} \check{R}_{\text{Ric},\alpha\kappa} + \sum_{\gamma=0}^3 \frac{\partial}{\partial x_\gamma} \check{R}_{\alpha\kappa}^{\gamma\beta} = 0. \quad (\text{D.15})$$

In the third summand, one can, according to (2.171), replace $\check{R}_{\alpha\kappa}^{\gamma\beta}$ by $-\check{R}_{\gamma\kappa}^{\alpha\beta}$. If we set $\alpha = \beta$ and sum over α , we obtain for (D.15) with the trace $\check{R} \stackrel{\text{def}}{=} \sum_{\alpha=0}^3 \check{R}_{\text{Ric},\alpha\alpha}$ of the Riccati-matrix \check{R}_{Ric}

$$\frac{\partial}{\partial x_\kappa} \check{R} - \sum_{\alpha=0}^3 \frac{\partial}{\partial x_\alpha} \check{R}_{\text{Ric},\alpha\kappa} - \sum_{\gamma=0}^3 \frac{\partial}{\partial x_\gamma} \check{R}_{\text{Ric},\gamma\kappa} = 0. \quad (\text{D.16})$$

If in the last sum the summation index γ is replaced by α , we can finally summarize:

$$\underline{\underline{\frac{\partial}{\partial x_\kappa} \check{R} - 2 \sum_{\alpha=0}^3 \frac{\partial}{\partial x_\alpha} \check{R}_{\text{Ric},\alpha\kappa} = 0.}}$$
(D.17)

One would get the same result, if one were to start with the equation:

$$\frac{\partial}{\partial x_\kappa} \check{R}_{\alpha\beta}^{\gamma\delta} - 2 \frac{\partial}{\partial x_\beta} \check{R}_{\alpha\kappa}^{\gamma\delta} = 0. \quad (\text{D.18})$$

Indeed, if we set $\delta = \gamma$ and sum over γ , we get first

$$\frac{\partial}{\partial x_\kappa} \check{R}_{\text{Ric},\alpha\beta} - 2 \frac{\partial}{\partial x_\beta} \check{R}_{\text{Ric},\alpha\kappa} = 0.$$

If we now set $\alpha = \beta$ and sum over α , we will again arrive at (D.17).

A different result is obtained when starting from (D.18) (with ν instead of γ) first, multiplying this equation by $g_{\gamma\nu}^{[-1]}$ to get

$$\frac{\partial}{\partial x_\kappa} g_{\gamma\nu}^{[-1]} \check{R}_{\alpha\beta}^{\nu\delta} - 2 \frac{\partial}{\partial x_\beta} g_{\gamma\nu}^{[-1]} \check{R}_{\alpha\kappa}^{\nu\delta} = 0,$$

then again setting $\gamma = \delta$ and summing over γ and ν and noting (D.6):

$$\begin{aligned} & \sum_\gamma \sum_\nu \frac{\partial}{\partial x_\kappa} g_{\gamma\nu}^{[-1]} \check{R}_{\alpha\beta}^{\nu\gamma} - 2 \sum_\gamma \sum_\nu \frac{\partial}{\partial x_\beta} g_{\gamma\nu}^{[-1]} \check{R}_{\alpha\kappa}^{\nu\gamma} \\ &= \frac{\partial}{\partial x_\kappa} R_{\text{Ric},\alpha\beta} - 2 \frac{\partial}{\partial x_\beta} R_{\text{Ric},\alpha\kappa} = 0. \end{aligned}$$

If we now take $\alpha = \eta$ and sum over α , we finally obtain the important relationship

$$\underline{\underline{\frac{\partial}{\partial x_\kappa} R - 2 \sum_\alpha \frac{\partial}{\partial x_\alpha} R_{\text{Ric},\alpha\kappa} = 0.}} \quad (\text{D.19})$$

These are the four equations for the four spacetime coordinates x_0, \dots, x_3 .

Finally, this overall result can be represented as

$$\underline{\underline{\vec{\nabla}^\top \left(\mathbf{R}_{\text{Ric}} - \frac{1}{2} R \mathbf{I}_4 \right) = \mathbf{0}^\top.}} \quad (\text{D.20})$$

References

1. V.I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978)
2. G. Barton, *Introduction to the Relativity Principle* (Wiley, New York, 1999)
3. H. Bondi, *Relativity and Common Sense* (Dover Publications, Dover, 1986)
4. J.W. Brewer, Kronecker products and matrix calculus in system theory. IEEE Trans. Circuits Syst. 772–781 (1978)
5. J.J. Callahan, *The Geometry of Spacetime* (Springer, Berlin, 2000)
6. S. Carroll, *Spacetime and Geometry: An Introduction to General Relativity* (Addison Wesley, San Francisco, 2003)
7. S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon Press, Oxford, 1983)
8. I. Ciufolini, J.A. Wheeler, *Gravitation and Inertia* (Princeton University Press, Princeton, 1995)
9. R. D’Inverno, *Introducing Einstein’s Relativity* (Clarendon Press, New York, 1992)
10. A. Einstein, *Relativity, The Special and the General Theory*, 2nd edn. (Edit Benei Noaj, 2007)
11. A. Föppl, Über einen Kreiselversuch zur Messung der Umdrehungsgeschwindigkeit der Erde. Sitzungsber. Bayer. Akad. Wiss. **34**, 5–28 (1904)
12. J. Foster, J.D. Nightingale, *A Short Course in General Relativity* (Springer, Berlin, 1995)
13. R.G. Gass, F.P. Esposito, L.C.R. Wijewardhana, L. Witten, Detecting event horizons and stationary surfaces (1998). [arXiv:gr-qc/9808055v](https://arxiv.org/abs/gr-qc/9808055v)
14. R. Geroch, *Relativity from A to B* (Chicago University Press, Chicago, 1981)
15. O. Gron, S. Hervik, *Einstein’s General Theory of Relativity* (Springer, New York, 2007)
16. O. Gron, A. Naess, *Einstein’s Theory* (Springer, New York, 2011)
17. J.B. Hartle, *Gravity: An Introduction to Einstein’s General Relativity* (Addison Wesley, San Francisco, 2003)
18. S.W. Hawking, G.F.R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973)
19. I.R. Kenyon, *General Relativity* (Oxford University Press, Oxford, 1990)
20. L.D. Landau, E.M. Lifschitz, *Classical Theory of Fields*: 2. Course of Theoretical Physics Series (Butterworth Heinemann, Stoneham, 1987)
21. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973)
22. R.A. Mould, *Basic Relativity* (Springer, Berlin, 2002)
23. J. Natario, *General Relativity Without Calculus* (Springer, Berlin, 2011)
24. W. Rindler, *Essential Relativity* (Springer, Berlin, 1977)
25. L. Sartori, *Understanding Relativity: A Simplified Approach to Einstein’s Theories* (University of California Press, Berkeley, 1996)
26. B.F. Schutz, *A First Course in General Relativity* (Cambridge University Press, Cambridge, 1985)
27. H. Stephani, *General Relativity* (Cambridge University Press, Cambridge, 1977)

28. J. Stewart, *Advanced General Relativity* (Cambridge University Press, Cambridge, 1991)
29. T. Takeuchi, *An Illustrated Guide to Relativity* (Cambridge University Press, Cambridge, 2010)
30. E.F. Taylor, J.A. Wheeler, *Spacetime Physics* (Freeman, New York, 1992)
31. E.F. Taylor, J.A. Wheeler, *Black Holes* (Addison Wesley, Reading, 2000)
32. E.F. Taylor, J.A. Wheeler, *Exploring Black Holes: Introduction to General Relativity* (Addison Wesley, San Francisco, 2000)
33. M. von Laue, *Die Relativitätstheorie. Erster Band: Die spezielle Relativitätstheorie*, 7th edn. (Vieweg, Wiesbaden, 1961)
34. R. Wald, *General Relativity* (Chicago University Press, Chicago, 1984)
35. R. Wald, *Space, Time and Gravity: Theory of the Big Bang and Black Holes* (University Press, Chicago, 1992)
36. S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972)
37. W.J. Wild, A matrix formulation of Einstein's vacuum field equations, [gr-qc/9812095](https://arxiv.org/abs/gr-qc/9812095), 31 Dec. 1998

Index

A

- Acceleration, 3, 26
 - vector
 - fourdimensional, 27
- Action functional
 - Hilbert's, 101
- Action-at-a-distance-equation, 98
- Aether, 7
- Affine mapping, 6

B

- Bianchi, 186
- Black hole, 109, 116, 127, 129
 - rotating, 138
- Block-transposed matrix, 63
- Brown dwarf, 128

C

- Calculus of variation, 57, 58
- Carbon, 128
- Chandrasekhar, 138
- Change
 - length, 117
- Charge, 6
- Circle
 - great, 173
- Conservation of mass, 44
- Constant
 - gravitational, 97
- Contraction, 13
 - time, 118
- Coordinate
 - Boyer–Lindquist, 139
 - Eddington's, 137, 142
 - spherical, 109
- Coordinate system, 51
- Covariant, 99

Current, 32

- Curvature, 78
 - Gaussian, 166, 177
 - definition, 168
 - Gauss, 169
 - normal, 175
 - principal, 177

Curvature matrix

- Riemannian, 174

Curvature scalar, 92

- Curve
 - geodesic, 57, 68
 - smooth, 166

Curved space, 66

Cylinder, 175

D

- Definite
 - positive, 175
- Density
 - energy, 41
 - Lagrange, 102
 - scalar, 102
 - Schwarzschild's, 128, 142
- Derivation, 160
 - covariant, 69
 - partial, 69
- Derivative
 - absolute, 68
 - covariant, 75
- Deviation
 - geodesic, 179
- Direction, 165
 - main, 175
 - normal, 168
- Displacement
 - parallel, 79

- Dynamics
Newton's, 101
- E**
Earth, 128
Eddington, 124
Effect
 Lense-Thirring, 109, 139
Einstein, 7, 93, 101
Einsteins field equation, 98
Electric field, 6, 32
Ellipsoid, 173, 175
Energy, 30
 kinetic, 31
Equation
 Einstein's, 102
 Einstein's field, 98
 Euler/Lagrange, 103
 Euler's, 44, 46
 Maxwell's, 32
 Poisson, 78, 94, 97, 101
Equation of motion
 relativistic, 29
Equivalence principle, 99
Event, 1
 time-like, 64
Experiment
 Michelson-Morley, 7, 15
- F**
Field
 gravitomagnetic, 140
Field equation
 Einstein's, 101
Field strength
 electric, 38
Fluid
 perfect, 44
Force
 Lorentz-, 141
 centrifugal, 66, 115
 Coriolis, 66
 electromagnetic, 7, 39
 gravitational, 44
Form
 quadratic differential, 175
 second fundamental, 175
Four-vector, 4
Fundamental law, 3
Fusion
 nuclear, 127
- G**
Galilei, 3
- Galilei-transformation, 1
Gauss, 78, 166, 174
General relativity, 51
General relativity principle, 7
Geodesic, 78, 179
Geometry
 differential, 78, 165
 elliptical, 173
 Euclidean, 173, 174
 hyperbolic, 173
Gravitational field, 52, 101
Gravitomagnetic field, 140
Gravity, 51
Gtransformation
 coordinate, 51
- H**
Hilbert, 101
Hyperboloid, 173
Hypersurface, 174
- I**
Identity
 Bianchi, 88, 92
Induction, 32
 magnetic, 38
Inertial system, 3, 53
Integral
 action, 103
 variated, 58
Interstellar cloud, 127
Invariance, 51, 67
Invariant, 99
Invariant
 electromagnetic, 38
 Kretschmann's, 132, 135
- K**
Kronecker-product, 154, 160
- L**
Light cone, 17, 18
Light wave propagation, 7
Line
 curved, 165
 geodesic, 68
 straight, 57
Line element, 51
Liquid
 neutron, 129
Lorentz-transformation, 11
- M**
Magnetic field, 6

- M**
- Mass
 - rotating, 138
 - solar, 128
 - spherical, 109
 - Matrix
 - inversion-lemma, 153
 - adjoint, 151
 - block, 152
 - Christoffel, 55, 66, 69, 71, 88, 100, 102
 - curvature, 95
 - diagonal, 147
 - differentiation, 157
 - Einstein's, 93, 98
 - electromagnetic energy-momentum-, 42
 - energy-momentum, 41, 97, 98
 - Faraday's, 33
 - identity, 147
 - inverse, 151
 - Jacobi, 52, 55, 65
 - Maxwell's, 34
 - mechanical energy-momentum-, 43
 - Minkowski, 99
 - permutation, 54, 156, 160
 - product, 150
 - Ricci, 89, 90, 92, 97, 98, 101, 113
 - Ricci-, 184
 - Riemannian curvature, 81, 85, 86, 89, 98, 102, 112
 - rotation, 5, 6
 - symmetric, 147
 - total energy-momentum-, 47
 - transposed, 147
 - Maxwell, 6
 - Mechanic
 - classical, 2
 - Metric
 - Kerr's, 139
 - Schwarzschild's, 116
 - Metric matrix, 52
 - Milky Way, 129
 - Minkowski, 19
 - Mirror, 8
- N**
- Neutron star, 127
 - Newton, 1, 94, 98
 - Nuclear solar fuel, 128
- O**
- Observer, 2, 8
 - Operator
 - Laplace, 78
 - vec-, 63
 - Oxygen, 128
- P**
- Paraboloid
 - hyperbolic, 176
 - Paradox
 - twin, 15
 - Parallelogram, 168
 - Phenomena
 - electromagnetic, 7
 - Photon, 19, 53
 - Plane
 - tangent, 166, 168
 - Point
 - elliptical, 175
 - hyperbolic, 176
 - parabolic, 175
 - saddle, 176
 - Polygon
 - inscribed, 165
 - Position
 - initial, 53
 - Position vector, 1
 - Postulate
 - parallel, 173
 - Potential
 - gravitational, 77, 97, 115
 - Principle
 - covariance, 99
 - relativity, 53
 - Product
 - Kronecker-, 154
 - Proper time, 20
- Q**
- Quadratic form, 57
- R**
- Radiation, 31
 - Radius
 - curvature, 166
 - Schwarzschild's, 116, 128, 141
 - Rate of change, 166
 - Rectilinear, 3
 - Redshift, 120
 - Reference frame, 1
 - Reference system, 2
 - Relativity principle
 - Galilei, 3
 - Rest mass, 30
 - Riemann, 78, 174
 - Rotation, 6
- S**
- Scalar
 - curvature, 90, 101, 107

- Schwarzschild, 109
- Shell
 - Schwarzschild's spherical, 131
- Shift
 - color, 120
- Ship
 - moving, 3
- Simultaneity, 12
- Singularity, 116, 131
 - coordinate, 131, 136
- Solution
 - Schwarzschild's, 116, 128
- Space
 - curved, 66
 - homogeneous, 5
 - isotropic, 5
 - n-dimensional, 78
- Space-time vector, 22
- Spaceship, 8
- Spacetime, 19
- Spacetime-interval, 51
- Sphere, 175
 - unit, 174
- Star, 127
- Supernova, 128
- Surface
 - curved, 165
 - first fundamental form, 170
 - hyper-, 78
 - second fundamental form, 170
 - spherical, 173
- Symbol
 - Christoffel-, 172
- Symmetric, 91
- System
 - inertial, 3
 - rotating, 64, 97
- T**
- Tangent vector, 68
- Tension
 - Maxwell, 7
- Theorem
 - Laplace expansion, 90, 104
 - Pythagorean, 8
- Theorema Egregium, 166, 169
- Theory
 - Lagrange/Hamilton, 103
- Thought experiment, 8
- Time
 - proper, 19, 53, 64
- Time dilation, 15
- Time shifting, 5
- Total difference, 159
- Train, 3
- Transformation
 - coordinate, 72
 - Galilei, 1, 4, 5
 - general Galilei, 5
 - similarity, 75
- Transport
 - parallel, 78
- V**
- Variation
 - calculus of, 57
- Variational principle, 101
- Vec-operator, 64
- Vector
 - Poynting-, 41
 - energy-momentum, 31
 - extended, 6
 - normal, 168
 - normalized tangent, 165
 - spacetime, 53
- Velocity
 - Lorentz-transformation, 22
 - initial, 53
- Velocity addition, 20
- Velocity of light, 52
- Virgo galaxy, 129
- W**
- White dwarf, 127, 128
- World-aether, 7

Günter Ludyk

Quantum Mechanics in Matrix Form



Springer

Quantum Mechanics in Matrix Form

Günter Ludyk

Quantum Mechanics in Matrix Form



Springer

Günter Ludyk
Physics and Electrical Engineering
University of Bremen
Bremen
Germany

ISBN 978-3-319-26364-9 ISBN 978-3-319-26366-3 (eBook)
<https://doi.org/10.1007/978-3-319-26366-3>

Library of Congress Control Number: 2017954872

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In memory of my parents

Martha and Paul Ludyk

Preface and Introduction

Though this be madness, yet there is method in't.

Hamlet, William Shakespeare

Matrix mechanics was introduced in 1925 by the German physicist Werner Heisenberg¹ [13]. However, the American Nobel laureate Steven Weinberg² writes about this publication [26]:

If the reader is mystified at what Heisenberg was doing, he or she is not alone. I have tried several times to read the paper that Heisenberg wrote on returning from Helgoland, and, although I think I understand quantum mechanics, I have never understood Heisenberg's motivations for the mathematical steps in his paper.

For years, Heisenberg and his colleagues had been struggling with a problem that had been raised in 1913 by Niels Bohr³ in his atomic theory: why do electrons in atoms occupy only certain permitted orbits with certain well-defined energies? Heisenberg took a completely new approach to this question. Because the trajectory of an electron in an atom cannot be directly observed, he decided to deal only with measurable quantities (namely the allowed energies for the quantum states of all electrons in the atom, as well as the frequencies with which an atom spontaneously jumps from one of these quantum states to a different state while emitting a light particle, i.e., a photon). Heisenberg introduced a sort of “table” of these frequencies. He performed mathematical operations on it, which led to new tables for the various physical quantities such as position, velocity, or the square of the velocity of an electron.

To be more precise, the table entries were the so-called transition amplitudes, that is, quantities whose squares specify a transition probability. When returning

¹Werner Heisenberg, 1901–1976, German physicist, Nobel Prize 1932.

²Steven Weinberg, * 1933, American physicist, Nobel Prize 1979.

³Niels Bohr, 1885–1962, Danish physicist, Nobel Prize 1922.

from Helgoland (where he first had this crucial idea) to Göttingen, Heisenberg found out that the operations he applied to these tables were well known to mathematicians. The tables were called *matrices*, and the operations that he used to get from the table representing the electron velocity to the table representing the square was named *matrix multiplication*. Starting from the known dependence between the energy of a particle and its velocity and position in a simple system, Heisenberg could calculate a table of the system's energies in its different quantum states, similar to Newton's⁴ calculation of the energy of a planet based on its position and velocity.

At the time, Heisenberg was constantly in touch with some influential theoretical physicists, including the German researchers Max Born⁵ and Pascual Jordan⁶ and the English physicist Paul Dirac.⁷ Until the end of the year 1925, they transformed Heisenberg's ideas into a comprehensive and systematic version of quantum mechanics, which today we refer to as *matrix mechanics*. With the help of the new matrix mechanics, the German physicist Wolfgang Pauli⁸ managed in the following January to solve the paradigmatic problem in atomic physics, namely the calculation of the energy levels of the hydrogen atom. His calculations proved the earlier ad hoc results of Bohr.

H.S. Green, an employee of Max Born, writes in [12]:

Most books on quantum theory emphasize the wave mechanical approach (of Schrödinger⁹), probably because it is supposed to be easier to understand for those who already have a solid knowledge on differential equations.

In this book, however, we restrict ourselves to the algebraic method using matrices and only briefly describe Schrödinger's wave mechanics in order to show the equivalence with Heisenberg's matrix mechanics. By implementing numerical algorithms in standard software such as MAPLE or MATHEMATICA, matrices and matrix equations are easy to handle these days [23].

I would like to thank Dr. Claus Ascheron from Springer for his kind assistance during the compilation of this manuscript. Last but not least, I would like to express my gratitude towards my wife Renate, without whom this book would have never been published.

Bremen, Germany

Günter Ludyk

⁴Isaac Newton, 1642–1727.

⁵Max Born, 1882–1970, German physicist, Nobel Prize 1954.

⁶Pascual Jordan, 1902–1980, German physicist.

⁷Paul Dirac, 1902–1984, English physicist, Nobel Prize 1933.

⁸Wolfgang Pauli, 1900–1958, German physicist, Nobel Prize 1945.

⁹Erwin Schrödinger, 1887–1961, Austrian physicist, Nobel Prize 1933.

Notations

Important definitions, facts, and theorems are shown in boxes. Important **intermediate results** are double underlined.

Scalars are written in normal font:

$$a, b, c, \alpha, \beta, \gamma, \dots$$

Vectors are written as lowercase letters in bold font:

$$\mathbf{x}, \mathbf{p}, \mathbf{v}, \dots$$

Vectors in four-dimensional spacetime (\mathbb{R}^4) are written as bold lowercase letters with an arrow:

$$\vec{x}, \vec{v}, \vec{u}, \dots$$

Matrices are written as uppercase letters in bold font:

$$\mathbf{X}, \mathbf{P}, \mathbf{R}, \mathbf{I}, \dots$$

Matrix vectors are written as uppercase letters in bold Fraktur font:

$$\mathfrak{X}, \mathfrak{E}, \mathfrak{B}, \dots$$

Matrix vectors are block matrices:

$$\mathfrak{X} \stackrel{\text{def}}{=} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}.$$

The identity matrix I_n of size n is an n -by- n matrix in which all the elements on the main diagonal are equal to 1 and all other elements are equal to 0:

$$\mathbf{I}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

If the components are real numbers (\mathbb{R}), the **transpose** of a vector x or a matrix A is written as x^T and A^T , respectively. If the components are complex numbers (\mathbb{C}), the transpose is written as x^\dagger and A^\dagger , respectively. In this case, the components of the transpose row vector are the complex conjugate. Note that this is necessary in order for the scalar product $x^\dagger x$ to be a real number. The same reasoning applies to the transpose matrix elements.

Contents

1	Quantum Theory Prior to 1925	1
1.1	Bohr–Sommerfeld Quantization Rule	1
1.2	Sommerfeld’s Derivation	2
1.3	Bohr’s Postulates	7
1.4	Atom Sizes	7
2	Heisenberg’s Year 1925	11
2.1	Spectral Lines	11
2.2	Introduction of Matrices	15
2.3	Problems	20
3	Expansion of the Matrix Method	23
3.1	Commutation Relation	23
3.2	Systems with Several Degrees of Freedom	28
3.3	Transformations	29
3.4	Problems	31
4	Observables and Uncertainty Relations	33
4.1	State Vector	33
4.2	The Stern-Gerlach Experiment	34
4.3	States and Postulates	35
4.4	Projection Matrices	36
4.5	Probabilistic Interpretation	38
4.6	Density Matrix	39
4.6.1	Definitions	39
4.6.2	Mixed States	40
4.6.3	Examples	42
4.7	Time Evolution of the Expectation Value	43
4.8	Heisenberg’s Uncertainty Principle	44
4.9	Problems	46

5	The Harmonic Oscillator	47
5.1	Physics of the Harmonic Oscillator	47
5.2	Expectation Values and Variances	56
5.3	Problems	57
6	Angular Momentum	59
6.1	The Matrix Vector of the Angular Momentum	59
6.2	Eigenvalues and Eigenvectors of L^2 and L_3	63
6.2.1	Commutativity of L^2 and L_3	63
6.2.2	Eigenvalues and Eigenvectors	64
6.2.3	Maximum and Minimum Eigenvalues	65
6.2.4	Orientation of the Angular Momentum Vectors	68
6.2.5	The Matrices L^2 and L_3	69
6.2.6	The Matrices L_+, L_-, L_1 and L_2	71
6.3	Problems	75
7	Wolfgang Pauli and the Hydrogen Atom	77
7.1	Basic Matrices and Matrix Vectors	77
7.2	Introduction of the Matrix Vector \mathfrak{A}	79
7.3	The Hydrogen Spectrum	83
7.4	Problems	84
8	Spin	85
8.1	Magnetic Fields and Light	85
8.2	Derivation of the Zeeman Effect (Without Spin)	85
8.3	Symmetry Considerations	87
8.4	Symmetry and the Spin	91
8.5	Spin- $\frac{1}{2}$ Systems and Spinors	93
8.6	Adding Angular Momenta	94
8.6.1	Clebsch–Gordan Coefficients	94
8.6.2	Clebsch–Gordan Coefficients on the Internet	98
8.7	Spin–Orbit Coupling	98
8.8	Problems	104
9	Atoms in Electromagnetic Fields	105
9.1	Normal Zeeman Effect	105
9.2	Anomalous Zeeman Effect	105
9.2.1	Weak Field Limit	105
9.2.2	Strong Magnetic Field	109
9.3	Problem	109
10	Many Particle Systems	111
10.1	Composed Systems	111
10.1.1	Systems with Two Distinguishable Particles	111
10.1.2	Systems with N Distinguishable Subsystems	113

10.1.3 Entangled Systems	114
10.2 Indistinguishable Subsystems	115
10.2.1 Interchanging Two Particles	119
10.2.2 Interchanging Three Identical Particles	124
10.2.3 Interchanging N Identical Particles	125
10.3 Problems	128
11 Equivalence of Matrix and Wave Mechanics	129
11.1 The De Broglie Wavelength	129
11.2 Operators in the Schrödinger Formalism	130
11.3 Schrödinger's Wave Mechanics	131
11.4 Equivalence of Heisenberg and Schrödinger Pictures	134
11.5 Example: The Harmonic Oscillator	139
11.6 Problems	140
12 Relativistic Quantum Mechanics	141
12.1 Special Relativity	141
12.1.1 Four-Dimensional Spacetime	141
12.1.2 Lorentz Transformation	142
12.1.3 Velocity and Its Lorentz Transformation	143
12.1.4 Momentum and Its Lorentz Transformation	144
12.1.5 Equation of Motion and Force	144
12.1.6 Energy and Rest Mass	146
12.2 The Dirac Equation	147
12.2.1 The Wave Equation for a Free Particle	147
12.2.2 Invariant Form of the Dirac Equation	150
12.2.3 Solution of the Dirac Equation	151
12.2.4 Dirac's Interpretation of the Negative Energy	153
12.3 Problems	154
Appendix A: Solutions to Problems	155
Appendix B: The Kronecker Product	183
Appendix C: Fourier Decomposition of Periodic Functions	185
Appendix D: Laplace–Runge–Lenz Vector	187
Appendix E: Permutation	191
Appendix F: Determinants	193
Appendix G: Dirac's Bra-Ket Notation	195
Appendix H: Proofs of Pauli's Formulas	197
Appendix I: Physical Quantities and Units	207
References	209
Index	211

Chapter 1

Quantum Theory Prior to 1925

Abstract A description is given of the “Older Quantum Mechanics” as introduced by Nils Bohr and Arnold Sommerfeld in the years before 1925. Bohr’s postulates are formulated and the atom size is derived from them.

1.1 Bohr–Sommerfeld Quantization Rule

Before the behavior of atoms could be described with the help of quantum mechanics, Bohr and Sommerfeld¹ explained the spectra of simple atoms based on Bohr’s atom model (“Older Quantum Mechanics”). In this model, spectral lines appear as energy differences between two “discrete” electron orbits. The Bohr–Sommerfeld quantization rule requires that not only the equation of motion holds for the orbit of the electron around the nucleus, but that for each additional circulation

$$\int p \, dx = nh \quad (n = 1, 2, \dots) \tag{1.1}$$

applies, where p is the momentum, and the position x runs through one complete circle. As does the angular momentum, this integral has the dimension location times moment and is an action. The constant

$$h = 2\pi \cdot \hbar = 2\pi \cdot 1.054572 \cdot 10^{-34} Js = 6.6260755 \cdot 10^{-34} Js$$

is the *Planck quantum of action* or *Planck’s constant*.² In other words, the action of each stationary electron orbit in the atom is quantized. The action can only occur as an integer multiple of Planck’s constant. A detailed derivation of this principle by Sommerfeld [22] is presented in the following section.

¹Arnold Sommerfeld, 1868–1951, German physicist.

²Max Planck, 1858–1947, German physicist, Nobel Prize 1919.

1.2 Sommerfeld's Derivation

Let us consider an arbitrarily moving point mass with momentum

$$\mathbf{p} = m\mathbf{v}. \quad (1.2)$$

With $\mathbf{v} = \dot{\mathbf{q}}$ we get

$$\mathbf{p} = m\dot{\mathbf{q}}. \quad (1.3)$$

It is important to note that the dynamic triple of the momentum coordinates \mathbf{p} is considered in addition to the geometric triple of the position coordinates \mathbf{q} . Newton supplies

$$\dot{\mathbf{p}} = \mathbf{f} = -\frac{\partial E_{pot}}{\partial \mathbf{q}}, \quad (1.4)$$

where it was assumed that the force \mathbf{f} can be derived from a potential energy E_{pot} (which is a function of the position coordinates \mathbf{q}). Using (1.3), the kinetic energy is obtained as

$$E_{kin} = \frac{m}{2}\dot{\mathbf{q}}^\top \dot{\mathbf{q}} = \frac{1}{2m}\mathbf{p}^\top \mathbf{p}.$$

The total energy as a function of the q_k and p_k is also called the Hamilton function or Hamiltonian H . We hence have

$$H(\mathbf{q}, \mathbf{p}) = E_{kin} + E_{pot}, \quad \frac{\partial H}{\partial \mathbf{q}} = \frac{\partial E_{pot}}{\partial \mathbf{q}}, \quad \frac{\partial H}{\partial \mathbf{p}} = \frac{\partial E_{kin}}{\partial \mathbf{p}} = \frac{\mathbf{p}}{m}.$$

As a result, the basic (1.3) and (1.25) can be written as

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}}.$$

(1.5)

Assuming that the kinetic energy is expressed as a function of \mathbf{q} and $\dot{\mathbf{q}}$, the equation

$$\mathbf{p} = \frac{\partial E_{kin}}{\partial \dot{\mathbf{q}}} \quad (1.6)$$

is identical to (1.3).

The values of the coordinates \mathbf{q} and \mathbf{p} determine the state of the system. To understand how the state of motion of the system depends on the location \mathbf{q} and the speed/momentum \mathbf{p} , let us imagine a single mass point with three degrees of freedom. Its three location coordinates and its three momentum coordinates can be plotted as Cartesian coordinates in a state space of six dimensions, such that each point of this state space represents a potential state of our mass point. Accordingly, for a system with d degrees of freedom one gets a state space with $2d$ dimensions.

In a first step, we can restrict ourselves to systems with only one degree of freedom. In this case, the general state space is a simple (two-dimensional) state plane. In this state plane, we can introduce q and p as orthogonal coordinates. Afterwards, we can construct *state trajectories* in this state plane, that is, the sequence of points corresponding to the successive states of motion of the system. Taking each point as the initial condition, we could draw such state trajectories and cover the entire state plane with them, whereby the trajectories are arbitrarily close to each other. It is characteristic for quantum theory, however, that a discrete subset of state trajectories is singled out from the infinite amount of potential trajectories. For defining this subset, we first consider a surface area of the state plane that is bounded by any two state trajectories, and we associate this area with the corresponding state. Afterwards we construct our bundle such that the area of the state between two neighboring curves is always equal to Planck's constant h . In other words, h has the meaning of an elementary area of the state. Let us from now on consider this meaning the true definition of Planck's quantum h .

We illustrate these abstract ideas with the help of two important special cases, namely the oscillator and the rotator. The linear *oscillator* is a spring attached to a point mass m at rest position, whereby the mass can only move in one direction $x = q$ from its rest position. Due to the spring, the mass experiences a restoring force, but no damping. The oscillator is the simplest model of a vibration center, as it is used in optics to describe a "quasi-elastically bounded electron." The oscillator is specified as a *harmonic oscillator* if we want to emphasize that only a certain eigen-oscillation is allowed. The oscillation frequency of the oscillator (i.e., the number of free vibrations in one time unit) is ν . The oscillation process then looks like

$$x = q = a \sin 2\pi\nu t, \quad (1.7)$$

where a is the amplitude of the oscillation. In this case, the momentum p is simply $m\dot{q}$; that is,

$$p = 2\pi\nu m a \cos 2\pi\nu t. \quad (1.8)$$

Eliminating t from (1.7) and (1.8), we obtain an ellipse as the state trajectory in the p, q -plane:

$$\frac{q^2}{a^2} + \frac{p^2}{b^2} = 1, \quad (1.9)$$

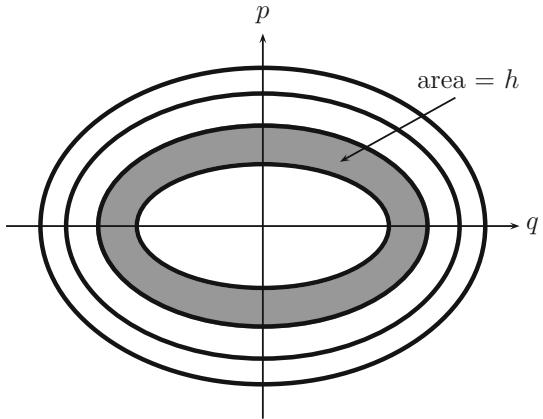
where the semi-minor axis b has the value

$$b = 2\pi\nu m a. \quad (1.10)$$

The area of the ellipse is

$$a b \pi = 2\pi^2 \nu m a^2.$$

Fig. 1.1 State plane of the linear oscillator



Note that this entity equals E/ν , where E is the constant vibrational energy. To see this, let us calculate E at the time $t = 0$, where the potential energy is equal to 0 and the kinetic energy is

$$\frac{m}{2}a^2(2\pi\nu)^2 = E. \quad (1.11)$$

Therefore in fact

$$a b \pi = \frac{E}{\nu}. \quad (1.12)$$

Figure 1.1 shows a family of similar ellipses in the state plane (p, q) that is obtained by changing E accordingly. After all, (1.10) ensures that the ratio b/a has the constant value $2\pi\nu m$. In the image, the ellipses follow one another in such a way that the resulting ellipse rings have the same area h . If we denote the difference of the energy constants for two consecutive ellipses with ΔE , we obtain from (1.12)

$$h = \frac{\Delta E}{\nu}, \quad \Delta E = h \nu. \quad (1.13)$$

By numbering the ellipses as $0, 1, \dots, n$ and calling the corresponding energies E_0, E_1, \dots, E_n , it follows from (1.13)

$$E_n = E_0 + h \nu n. \quad (1.14)$$

In the classical theory, all points of the state plane are equal and represent possible states of the oscillator. In quantum theory, however, those states are distinguished whose points lie on one of those ellipses. Such states are called the stationary states of the oscillator, because they can be continuously passed through without energy loss of the oscillator. In a way, they represent the charged mass point without radiation.

From time to time, however, the oscillator changes its energy. In particular, it emits energy once its state moves onto a smaller ellipse. Similarly, it absorbs energy once the state jumps to a larger ellipse. Note, however, that both emission and absorption occur in multiples of the energy quantum $\hbar\nu$.

Let us generalize this insight to any mechanical system with one degree of freedom. *The image point of the system in the state plane is restricted to certain quantum theoretically special “quantized” state trajectories, each of which includes an elementary region of size h with the following trajectory. The n th trajectory includes (when closed) the area*

$$J = \iint dp dq, \quad (1.15)$$

where the integral covers the inside of the n th trajectory. Carrying out the integration with respect to p yields

$$J = \int p dq, \quad (1.16)$$

where the integral is bound by the n th trajectory itself. We call (1.16) the *state integral* or *action*.

The final formulation of the quantum hypothesis is obtained by requiring that the difference of the state integrals for two consecutive trajectories must be equal to h :

$$\Delta J = h, \quad J = J_0 + n h. \quad (1.17)$$

This requirement selects a discrete (infinite) number of real, quantum theoretically possible trajectories out of the *continuous* manifold of all possible mechanical movements. Unlike this general version of quantum hypothesis, the original energy quantum hypothesis as formulated by Planck with respect to thermal radiation is only valid for the oscillator. The above evaluation of the state integral (1.15) was very easy because we could directly calculate the area of the ellipses by the formula $a b \pi$. In particular, (1.17) yields (assuming $J_0 = 0$) a formulation that is analogous to (1.14):

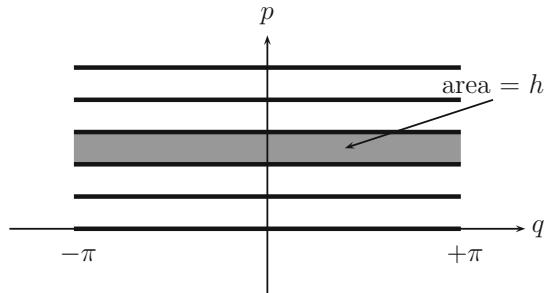
$$J = \int p dq = n h. \quad (1.18)$$

Now let us move on to the *rotator*. A rotator is a point mass m that rotates uniformly on a circle of radius a around a fixed center. We can use the rotation angle $\varphi = q$ as a natural location coordinate. Starting from an arbitrary initial position $\varphi = 0$, the particle position on the circle is labeled by the angle φ . Because the velocity of the point mass m is equal to $a\dot{q}$, the kinetic energy is given by

$$E_{kin} = \frac{m}{2} a^2 \dot{q}^2. \quad (1.19)$$

For a uniform rotation, the potential energy is independent of φ . Also, we can assume that it does not depend on a either, inasmuch as a is constant during the motion. We

Fig. 1.2 State plane of the rotator



can therefore write

$$E_{pot} = \text{const.}$$

According to (1.6) and (1.19), the momentum corresponding to \dot{q} is

$$p = m a^2 \dot{q}. \quad (1.20)$$

Because $\dot{q} = \text{const}$, this momentum p is also constant, in accordance with the equations of motion (1.5). Therefore, the state trajectory of the rotator in the state plane (p, q) is a line that is parallel to the q -axis; see Fig. 1.2.

Note that the state trajectory is not a closed curve in this case. Therefore we must first define what we understand as the area of the state trajectory. We remark that the state of the rotator (i.e., its location on the trajectory as well as the direction of its momentum) is repeated after each loop. The true state trajectory is therefore not an infinite, but a finite line that is repeatedly passed through from left to right. In the q -direction, the state area of the rotator has only the length 2π . Casually speaking, one can cut out the line $q = \pm\pi$ and glue it together to form a cylinder. The area of the cylinder between the n th and the $(n-1)$ th state trajectory is then given (similar to a rectangle with the baseline 2π) by $2\pi(p_n - p_{n-1})$. This area must now be equal to h . For the area between the n th and the zero state trajectory (the latter being represented by the q -axis), we then get

$$2\pi p_n = n h. \quad (1.21)$$

This area plays the same role as the closed curves in the case of the oscillator.

As we can see, the rotator is *not quantized in energy quanta* (as the oscillator), but according to its angular momentum. More precisely, the angular momentum of a rotator is an integral multiple of $h/2\pi$. We can now calculate the (kinetic) energy of the rotator based on (1.19) and (1.20) as

$$E_{kin} = \frac{p \dot{q}}{2}.$$

From (1.21), with $\nu = \dot{q}/2\pi$ we further get:

$$E_{kin} = \frac{n h}{2} \frac{\dot{q}}{2\pi} = \frac{n h \nu}{2}. \quad (1.22)$$

Here, ν is the winding number of the rotator, that is, the number of full rotations in one time unit. This number is the analogue of the oscillation frequency of the oscillator.

1.3 Bohr's Postulates

Niels Bohr derived his model by adding three postulates to Rutherford's model, namely:

1. Electrons move on stable orbits around the nucleus. Contrary to the prediction of the theory of electrodynamics, the electron radiates no energy (in the form of electromagnetic radiation) when moving on the orbits.
2. The radius of the electron orbit does not change continuously, but abruptly. When the electron jumps from one orbit to another, electromagnetic radiation is emitted (or absorbed). The frequency of the radiation is given by Max Planck's relationship between energy and the frequency of light. According to this relationship, if E_n is the energy of the initial state and E_m is the energy of the target state, then a photon is emitted with frequency

$$\nu = \frac{(E_m - E_n)}{h}. \quad (1.23)$$

3. Electron orbits are only stable if the orbital angular momentum L of the electron is an integral multiple of the reduced Planck constant $\hbar = \frac{h}{2\pi}$: $L = n\hbar$. This postulate is often called the *selection condition*.

In addition, the classical equation of motion applies, that is, a centripetal Coulomb force acts.

1.4 Atom Sizes

Bohr's atom model allows a comparison of a series of numerical predictions with experimental results, especially with respect to the position of the lines in the hydrogen spectrum. The model treats the electron as a pointlike particle that is attracted by the opposite electric charge of the nucleus. This force shapes the path of the electron according to the laws of classical mechanics on circular orbits. That is why in Bohr's atomic model the distance of an electron to the nucleus is also referred to as the

classical atomic radius. The angular momentum L of a particle with mass m and velocity v along a circular orbit of radius r is

$$L = mvr.$$

A centripetal force acts on the moving electron according to

$$F_{\text{centr}} = \frac{mv^2}{r}.$$

On the other hand, the electron with the elementary charge e experiences a force in the electric field of the proton according to Coulomb's law:

$$F_{\text{el}} = \frac{e^2}{4\pi\varepsilon_0 r^2}.$$

This Coulomb force keeps the electron on the circular orbit; that is, the two forces must be equal:

$$F_{\text{el}} = F_{\text{centr}} \Leftrightarrow \frac{e^2}{4\pi\varepsilon_0 r^2} = \frac{m_e v^2}{r}. \quad (1.24)$$

Hereby, the angular momentum must satisfy the above-postulated selection condition. A more sustainable formulation of this condition is that the length of the orbit ($2\pi r$) must be an integer multiple of the electron's elementary wavelength (or *de Broglie wavelength*), because otherwise destructive interference would occur. Along a permitted orbit, the electron forms a standing matter wave:

$$\lambda_{\text{dB}} = \frac{h}{p} = \frac{h}{m_e v} \Rightarrow 2\pi r = n \frac{h}{m_e v}.$$

Solving for v yields

$$v = \frac{n\hbar}{m_e r}.$$

This velocity v can be substituted in the above equation for the forces:

$$\begin{aligned} \frac{e^2}{4\pi\varepsilon_0 r^2} &= \frac{m_e \left(\frac{n\hbar}{m_e r} \right)^2}{r} \Leftrightarrow \frac{e^2}{4\pi\varepsilon_0 r^2} = \frac{m_e n^2 \hbar^2}{m_e^2 r^3} \Leftrightarrow \\ r_n &= n^2 \frac{\varepsilon_0 \cdot h^2}{m_e e^2 \pi}. \end{aligned} \quad (1.25)$$

The smallest radius ($n = 1$) is denoted as *Bohr's atomic radius*:

$$r_1 = \frac{\varepsilon_0 h^2}{m_e e^2 \pi} \approx 5.29 \cdot 10^{-11} \text{ m}. \quad (1.26)$$

For each higher electron orbit, we get

$$r_n = n^2 \cdot r_1 \approx n^2 \cdot 5.29 \cdot 10^{-11} \text{ m}. \quad (1.27)$$

The energy E_n of the electron on the n th orbit consists of kinetic as well as potential energy, whereby the potential energy is negative:

$$E_n = E_{kin} + E_{pot} = \frac{m_e \cdot v_n^2}{2} - \frac{1}{4\pi\varepsilon_0} \cdot \frac{e^2}{r_n}. \quad (1.28)$$

From (1.24), it follows that

$$v_n^2 = \frac{e^2}{4\pi\varepsilon_0 \cdot r_n \cdot m_e}.$$

We substitute this term in (1.28), which provides

$$E_n = -\frac{1}{2} \frac{1}{4\pi\varepsilon_0} \frac{e^2}{r_n}.$$

With (1.26) and (1.27), we get

$$E_n = -\frac{1}{2} \frac{1}{4\pi\varepsilon_0} \frac{e^2}{n^2 r_1} = -\frac{m_e e^4}{8\varepsilon_0^2 h^2} \frac{1}{n^2} = -13.6 \text{ eV} \cdot \frac{1}{n^2}. \quad (1.29)$$

In a hydrogen atom, an electron can only assume these discrete energy values, where the integer n denotes its current trajectory.

Chapter 2

Heisenberg's Year 1925

Abstract Starting from the known facts on spectral lines up to 1925, the crucial new ideas of Heisenberg are presented which led him to the introduction of his matrix quantum mechanics.

2.1 Spectral Lines

Nineteenth century research on spectral lines can be considered the starting point for quantum theory. These investigations are based on two measurable variables, namely the frequency and the brightness of the spectral lines. After all, the emitted light consists of very sharp frequencies. In the year 1850, Kirchhoff¹ and Bunsen² had discovered that chemical elements produce such characteristic lines in their spectra.

In the visible region of the *hydrogen spectrum*, four lines can be observed (the wavelengths get closer as they decrease). In 1885, the Swiss mathematician Balmer³ (a teacher at a lyceum) discovered that the wavelength λ of these lines can be calculated with the simple formula

$$\lambda = A \left(\frac{n^2}{n^2 - 4} \right),$$

where $A = 364.56 \times 10^{-9}$ m and $n = 3, 4, 5$, or 6 . This leads to the following wavelengths (in nanometers):

- 656.279 nm (red)
- 486.133 nm (blue-green)
- 434.047 nm (violet)
- 410.174 nm (violet)

¹Gustav Robert Kirchhoff, 1824–1887, German physicist.

²Robert Wilhelm Eberhard Bunsen, 1811–1899, German chemist.

³Johann Jakob Balmer, 1825–1898.

Table 2.1 Measured or calculated hydrogen spectral lines in the Balmer series

Transition $n \rightarrow m = 2$	$3 \rightarrow 2$	$4 \rightarrow 2$	$5 \rightarrow 2$	$6 \rightarrow 2$	$7 \rightarrow 2$	$8 \rightarrow 2$
Name	H_α	H_β	H_γ	H_δ	H_ϵ	H_ζ
Measured (nm)	656.2793	486.1327	434.0466	410.1738	397.0075	388.8052
Calculated	656.278	486.132	434.045	410.1735	397.0074	388.8057
Colour	Red	Blue-Green	Violet	Violet	Violet	Violet

For the wavelengths λ , we obtain the *Balmer formula*

$$\frac{1}{\lambda} = R \left(\frac{1}{4} - \frac{1}{n^2} \right),$$

where $R = 4/A$ is the so-called *Rydberg*⁴ *constant*.⁵ With explicit numbers, it reads

$$\frac{1}{\lambda} = 10973731 \left(\frac{1}{4} - \frac{1}{n^2} \right) [m^{-1}].$$

Table 2.1 is a table of the currently known spectral lines in the Balmer series, measured or calculated with the Balmer formula.

Five years later, in 1890, Rydberg generalized the Balmer equation into the Rydberg formula

$$\frac{1}{\lambda} = R \left(\frac{1}{m^2} - \frac{1}{n^2} \right)$$

with $m = 1, 2, \dots$ and $n = m + 1, m + 2, \dots$. For $m = 1$, one obtains the Lyman series. For $m = 2$, we get again Balmer's series, and $m = 3$ yields the Paschen series. For $m = 4$, the Brackett series is obtained, and $m = 5$ leads to the Pfund series. Each series is named after its discoverer; see Table 2.2 and Fig. 2.1.

For the spectra of other elements, different Rydberg constants are obtained. The Swiss physicist Ritz⁶ discovered in 1908 that one can derive new lines from the known spectral lines of an element without having to modify any constants. Here is why: from the above formulas it becomes obvious that the resulting frequencies ν depend on two integers, namely m and n , in other words,

$$\nu(m, n) = Rc \left(\frac{1}{m^2} - \frac{1}{n^2} \right). \quad (2.1)$$

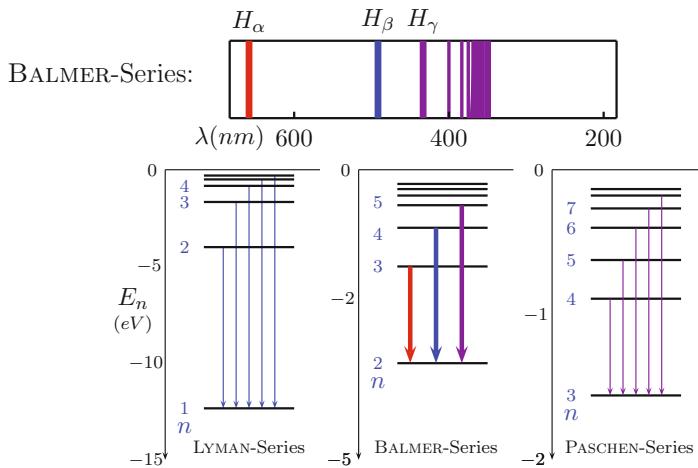
⁴Johannes Robert Rydberg, 1854–1919, Swedish physicist.

⁵Today, R is an accurately known fundamental constant with $R_\infty = 10973731.568539(55) m^{-1}$. The index ∞ indicates that an infinitely large nucleus mass is assumed.

⁶Walter Ritz, 1878–1909, Swiss theoretical physicist.

Table 2.2 Hydrogen spectral lines series

Name	n_1	n_2	Formula	Spectrum range/Colour
LYMAN-Series	1	2, 3, 4, ...	$\tilde{\nu} = R \left(1 - \frac{1}{n_2^2} \right)$	Vacuum-UV (121 nm → 91 nm)
BALMER-Series	2	3, 4, 5, ...	$\tilde{\nu} = R \left(\frac{1}{2^2} - \frac{1}{n_2^2} \right)$	red, blue-green, 4× violet, then transition to near-UV → 365 nm
PASCHEN-Series	3	4, 5, 6, ...	$\tilde{\nu} = R \left(\frac{1}{3^2} - \frac{1}{n_2^2} \right)$	IR-A (1875 nm → 820 nm)
BRACKETT-Series	4	5, 6, 7, ...	$\tilde{\nu} = R \left(\frac{1}{4^2} - \frac{1}{n_2^2} \right)$	IR-B (4050 nm → 1460 nm)
PFUND-Series	5	6, 7, 8, ...	$\tilde{\nu} = R \left(\frac{1}{5^2} - \frac{1}{n_2^2} \right)$	IR-B (7457 nm → 2280 nm)

**Fig. 2.1** The visible part of the hydrogen spectrum of the Balmer series (above). Energy levels and transitions (below)

Adding two different frequencies $\nu(m_1, n_1)$ and $\nu(m_2, n_2)$ of a spectrum, one obtains

$$\nu(m_1, n_1) + \nu(m_2, n_2) = Rc \left(\frac{1}{m_1^2} - \frac{1}{n_1^2} + \frac{1}{m_2^2} - \frac{1}{n_2^2} \right). \quad (2.2)$$

For $n_1 = m_2$ we obtain the new frequency according to the *Ritz combination principle*:

$$\underline{\underline{\nu(m_1, n_2)}} = \nu(m_1, n_1) + \nu(n_1, n_2) = R \cdot c \left(\frac{1}{m_1^2} - \frac{1}{n_2^2} \right). \quad (2.3)$$

And then the time of Niels Bohr came! It can be safely assumed that he was inspired by the Rydberg formula

$$\nu(m, n) = R \cdot c \left(\frac{1}{m^2} - \frac{1}{n^2} \right).$$

We can write this formula without brackets:

$$\nu(m, n) = R \cdot c \frac{1}{m^2} - R \cdot c \frac{1}{n^2}. \quad (2.4)$$

On the other hand, we have

$$E = h \cdot \nu,$$

and therefore

$$\nu = \frac{E}{h}.$$

It seems very plausible to rewrite for (2.4)

$$\nu(m, n) = \frac{E_m}{h} - \frac{E_n}{h}. \quad (2.5)$$

From this follows

$$E_k = \frac{h \cdot R \cdot c}{k^2}.$$

Note that the dimensions in the formula are consistent: the action quantum h has dimension Js , $R \cdot c$ has dimension s^{-1} and k is a dimensionless integer, so E_k has indeed the dimension J .

We need to emphasize once again that in quantum theory the mechanical behavior of an atom is characterized by two basic parameters, namely the energy E_n of the stationary state n and the probability per unit time $A(n, m)$ for the spontaneous transition from state n to the state m . In a spectroscopic study, we measure the radiation as emitted by the atom, that is, the line spectrum. Quantum theory then puts the mechanical properties (namely E_n and $A(n, m)$) into relation to the spectral characteristics (namely the frequency and intensity of the emitted light). Hereby, the transition energy $E_n - E_m$ determines the frequency of the light, and the transition probability $A(n, m)$ determines its intensity.

The emission of radiation by an atom is the result of electrons jumping between two discrete electron orbits. The transition probability determines the occurrence of a quantum jump. The radiation emitted during the transition $n \rightarrow m$ then has the frequency $\nu(n, m)$. Therefore, a mechanical energy loss $E_m - E_n$ occurs during

the transition $n \rightarrow m$. Because⁷ the photon energy is $h\nu(n, m) = \hbar\omega(n, m)$, energy conservation requires $E_m = E_n + \hbar\omega(n, m)$; that is,

$$E_m - E_n = \hbar\omega(n, m). \quad (2.6)$$

Note that a single photon only generates a flash of light on a spectral line. The complete spectral line is produced only if many atoms emit many photons. The number of photons with frequency $\nu(n, m)$ that arrives at a given area of the spectrometer per unit time determines the intensity of the line. In other words, the line intensity is a function of the number of jumping electrons, that is, the transition rate $A(n, m)$.

Let us consider a set of atoms, each in the state n . Then the light intensity $P(n, m)$ of the transition $n \rightarrow m$ defined by the amount of energy emitted per unit time Δt and per atom is

$$P(n, m) \stackrel{\text{def}}{=} \frac{1}{N_n} \frac{\Delta E(n, m)}{\Delta t}, \quad (2.7)$$

where N_n is the number of atoms in the state n , and $\Delta E(n, m)$ is the energy consumption of all atoms accomplishing the transition $n \rightarrow m$ in the time interval Δt . Again, the conservation of energy requires

$$\Delta E(n, m) = \Delta N(n, m) \hbar\omega(n, m),$$

where $\Delta N(n, m)$ is the number of atoms jumping from n to m in the time interval Δt . For large N_n , the portion of jumping atoms equals the probability for an atom to jump, namely

$$\Delta N(n, m)/N_n = A(n, m)\Delta t.$$

$A(n, m)$ is therefore the probability per unit time. From (2.7) we get

$$P(n, m) = \frac{\Delta N(n, m) \hbar\omega(n, m)}{N_n \Delta t} = A(n, m) \hbar\omega(n, m). \quad (2.8)$$

This result implies that intense lines are very probable transitions, and weak lines are improbable transitions.

2.2 Introduction of Matrices

Werner Heisenberg started off from the principle that concepts which do not correspond to *physically observable facts* should not be used in a theoretical description. Heisenberg therefore banished the idea of electron orbits with fixed radii and orbital

⁷ $\hbar \stackrel{\text{def}}{=} \frac{h}{2\pi}$.

periods, because these quantities cannot be observed. Instead, he postulated that the theory was to be constructed with the help of abstract quadratic schemes. Rather than describing the motion by a time-dependent coordinate $x(t)$, he suggested determining a scheme of transition amplitudes x_{mn} . Heisenberg's theory is entirely based on measurable quantities, namely the frequencies and strengths of the spectral lines of atoms.

From Rydberg's formula it follows that each frequency $\nu(n, m)$ of the observed spectrum can be written as the difference of two energy terms E_n and E_m :

$$h\nu(n, m) = E_n - E_m. \quad (2.9)$$

This immediately leads to the Ritz combination principle

$$\nu(n, k) + \nu(k, m) = \frac{1}{h} ((E_n - E_k) + (E_k - E_m)) = \frac{1}{h} (E_n - E_m) = \nu(n, m), \quad (2.10)$$

where the ν s are the observable frequencies of the spectrum. From (2.9), we also find that

$$\nu(n, m) = \frac{1}{h} (E_n - E_m) = -\frac{1}{h} (E_m - E_n) = -\nu(m, n), \quad (2.11)$$

and that

$$\nu(n, n) = \frac{1}{h} (E_n - E_n) = 0. \quad (2.12)$$

For an objective observer, the frequencies and intensities of the spectral lines are the only available data of what is happening inside an atom. Because the occurring frequencies depend on two terms, it makes perfect sense to arrange the frequencies in a table. The first row contains all frequencies that are generated from E_0 , that is,

$$\nu(0, 0), \nu(0, 1), \nu(0, 2), \dots$$

The second row contains frequencies that can be generated starting from E_1 , namely

$$\nu(1, 0), \nu(1, 1), \nu(1, 2), \dots$$

The columns always contain all frequencies (in rising order) that can be generated upon reaching a certain level. Eventually, the table can be written in the form of a matrix Ω as

$$\Omega = 2\pi \begin{pmatrix} \nu(0, 0) & \nu(0, 1) & \nu(0, 2) & \dots \\ \nu(1, 0) & \nu(1, 1) & \nu(1, 2) & \dots \\ \nu(2, 0) & \nu(2, 1) & \nu(2, 2) & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \quad (2.13)$$

For a harmonically bound electron, the position $x(n, t)$, where the stationary state is labeled by n , can be written as a Fourier series

$$x(n, t) = \sum_{\alpha=-\infty}^{\infty} a_{\alpha} e^{i\omega(n)t}. \quad (2.14)$$

Heisenberg observed that the α th component of the classical motion corresponds to the quantum mechanical transition from the state n to the state $n - \alpha$. So Heisenberg replaced the classical component

$$a_{\alpha} e^{i\omega(n)t}$$

by

$$a(n, n - \alpha) e^{i\omega(n,n-\alpha)t}.$$

In order to account for the transition from one stationary state n to another $(n - \alpha)$, he replaced (2.14) by

$$x \rightarrow a(n, n - \alpha) e^{i\omega(n,n-\alpha)t},$$

or

$$x_{nm} = a(n, m) e^{i\omega(n,m)t}, \quad (2.15)$$

and summed over transition components as in (2.14). Heisenberg represented the position by a set of transition components, x_{mn} , and he replaced $x_{\alpha}(n)$ by x_{mn} and $\alpha\omega(n)$ by $\omega(n, m)$.

Additionally Heisenberg modified the old Bohr–Sommerfeld quantization rule (1.18)

$$J = \int p dq = \int m \dot{x}^2 dt = h n,$$

integrated over a full period of the motion. If one expressed this equation in terms of the Fourier series (2.14) for $x(n, t)$ one would obtain

$$h n = 2\pi m \sum_{\alpha=-\infty}^{\infty} |a_{\alpha}(n)|^2 \alpha^2 \omega_n. \quad (2.16)$$

The presence of the integer n in (2.16) seemed for Heisenberg to be an arbitrary condition, and he concluded that this equation must be replaced by a new condition and that the new condition must be about the transition between states. By differentiation (2.16) with respect to n Heisenberg found

$$h = 2\pi m \sum_{\alpha=-\infty}^{\infty} \alpha \frac{d}{dn} (\alpha \omega_n |a_{\alpha}|^2).$$

Heisenberg replaced the derivative by a difference:

$$h = 4\pi m \sum_{\alpha=0}^{\infty} \{|a(n, n+\alpha)|^2 \omega(n, n+\alpha) - |a(n, n-\alpha)|^2 \omega(n, n-\alpha)\}. \quad (2.17)$$

This is Heisenberg's quantum condition; it relates the amplitudes of different lines within an atomic spectrum.

How is the quantity $(x(t))^2$ to be represented in quantum mechanics, which appears, for example, by modeling an anharmonic oscillator? The answer in classical theory is obviously:

$$b_{\beta}(n) e^{i\omega(n)\beta t} = \sum_{\alpha=-\infty}^{\infty} a_{\alpha} a_{\beta-\alpha} e^{i\omega(n)(\alpha+\beta-\alpha)t}, \quad (2.18)$$

so that

$$(x(t))^2 = \sum_{\beta=-\infty}^{\infty} b_{\beta}(n) e^{i\omega(n)\beta t}. \quad (2.19)$$

To Heisenberg it seemed that in quantum mechanics the simplest and most natural assumption would be to replace (2.18) by:

$$b(n, n-\beta) e^{i\omega(n,n-\beta)t} = \sum_{\alpha=-\infty}^{\infty} a(n, n-\alpha) a(n-\alpha, n-\beta) e^{i\omega(n,n-\beta)t}. \quad (2.20)$$

Max Born studied Heisenberg's manuscript and discovered that Heisenberg's symbolic multiplication was nothing but matrix multiplication.

By collecting all possible transitions in a matrix (similar to the frequencies ν in Ω), we obtain a matrix of the form⁸

$$\mathbf{X} = (a(n, m) e^{i2\pi\nu(n,m)t}).$$

Due to $\nu(n, m) = -\nu(m, n)$ and $a(m, n) = a^*(n, m)$, we get $x(n, m) = x^*(m, n)$; that is, the matrix \mathbf{X} is an Hermitian⁹ matrix. When transposing an Hermitian matrix \mathbf{X} , each component becomes its complex conjugate value. If we introduce a matrix multiplication as $\mathbf{X}\mathbf{X}^\dagger$, the elements of the product matrix are

$$a(n, m)a(m, n) = a(n, m)a^*(n, m) = |a(n, m)|^2. \quad (2.21)$$

⁸The matrix components $x(n, m) = a(n, m) e^{i2\pi\nu(n,m)t}$ are not to be confused with the classical coefficients $a_{\alpha} e^{i\alpha\omega t}$ of a Fourier series (see Appendix C), where we sum up from $\alpha = -\infty$ to $\alpha = +\infty$ in order to obtain the periodic function $x(t)$.

⁹Charles Hermite, 1822–1901, French mathematician.

Born and Jordan postulated in [5] that (2.21) is the *likelihood* for the transitions $n \rightleftharpoons m$ from the atomic state n into the state m , and vice versa.

The description is based on the idea of state transitions, therefore the diagonal elements of the matrix X must vanish; that is,

$$\nu(n, n) = 0$$

for all n . After all, no transition takes place from n to n . Altogether we get the form

$$X = \begin{pmatrix} 0 & a(0, 1) e^{2\pi i \nu(0,1)t} & a(0, 2) e^{2\pi i \nu(0,2)t} & \dots \\ a(1, 0) e^{2\pi i \nu(1,0)t} & 0 & a(1, 2) e^{2\pi i \nu(1,2)t} & \dots \\ a(2, 0) e^{2\pi i \nu(2,0)t} & a(2, 1) e^{2\pi i \nu(2,1)t} & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (2.22)$$

Note that it is not summed over any coefficients as in a Fourier series. Rather, all the transition components that are collected in the matrix X reproduce all the possibilities for transitions of the system from the perspective of quantum theory. The aim of quantum theory is to create a mathematical model such that:

1. It allows the calculation of measurable frequency spectra of atoms (namely frequency and intensity of the spectral lines).
2. The classical theory is contained in the limit $\hbar \rightarrow 0$, where \hbar is Planck's quantum of action.

For the time derivatives of the elements $x(n, m)$ of the matrix X one gets

$$\dot{x}(n, m) = 2\pi i \nu(n, m) a(n, m) e^{2\pi i \nu(n, m)t}. \quad (2.23)$$

We introduce the diagonal matrix E with the matrix elements

$$E(n, m) \stackrel{\text{def}}{=} \delta_{n,m} E_n, \quad (2.24)$$

that is,

$$E = \begin{pmatrix} E_0 & 0 & 0 & \dots \\ 0 & E_1 & 0 & \dots \\ 0 & 0 & E_2 & \ddots \\ 0 & 0 & 0 & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

With (2.9), we then get for (2.23)

$$\dot{x}(n, m) = 2\pi i \nu(n, m) x(n, m) = \frac{2\pi i}{\hbar} (E_n - E_m) x(n, m). \quad (2.25)$$

Because E is diagonal, we have for matrix elements

$$E_n x(n, m) = (\mathbf{E} \mathbf{X})(n, m)$$

and

$$E_m x(n, m) = (\mathbf{X} \mathbf{E})(n, m).$$

Therefore we get

$$\dot{x}(n, m) = \frac{2\pi i}{\hbar} ((\mathbf{E} \mathbf{X})(n, m) - (\mathbf{X} \mathbf{E})(n, m)),$$

and because this applies to every element of the matrix

$$\dot{\mathbf{X}} = \frac{2\pi i}{\hbar} (\mathbf{E} \mathbf{X} - \mathbf{X} \mathbf{E}).$$

(2.26)

This is the simplest form of the so-called *quantum mechanical equation of motion*, or *Heisenberg's equation of motion*. Here for the first time a *commutator* enters in the form

$$[\mathbf{E} \mathbf{X}] \stackrel{\text{def}}{=} (\mathbf{E} \mathbf{X} - \mathbf{X} \mathbf{E}).$$

In the theory of quantum mechanics such commutators play an important role, as we show later. For a single matrix element, (2.26) means (2.25). Dividing by $2\pi i x(n, m)$ again yields the frequency condition $\hbar\nu(n, m) = E_n - E_m$.

2.3 Problems

- 2.1 Hermitian Matrices:** Under which conditions is the product of two Hermitian matrices again an Hermitian matrix?
- 2.2 Eigenvectors:** Show that eigenvectors belonging to different eigenvalues are linearly independent.
- 2.3 Eigenvalues of an Hermitian Matrix:** Show that the eigenvalues of an Hermitian matrix are real.
- 2.4 Eigenvalues of a Unitary Matrix:** What general property do the eigenvalues of a unitary matrix have?
- 2.5 Eigenvectors:** In an N -dimensional space, N linear independent vectors \mathbf{a}_j are given. Construct a set of N normalized orthogonal vectors \mathbf{e}_j .

2.6 Normalized Eigenvectors: The two linear independent vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ are given. Which transformation matrix T transforms these two vectors into $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$? Which vectors would be obtained by using the method of Problem 2.5?

Chapter 3

Expansion of the Matrix Method

Abstract We describe the general ideas concerning the matrix method, which was further developed by Born, Heisenberg, and Jordan and extended to systems with several degrees of freedom.

3.1 Commutation Relation

In classical mechanics, the behavior of a system is described by canonical variables $p_1, p_2, \dots, q_1, q_2, \dots$, and its dynamical structure is contained in the Hamilton function $H(p_1, p_2, \dots, q_1, q_2, \dots)$. This scheme turned out to be very suitable for quantum theory as well. However, the canonical variables have to be replaced by more general mathematical objects, namely matrices. Because one wants to end up with the classical Hamilton theory, it makes sense to call the entries of the matrix X position coordinates and the correspondingly structured matrix $P = m\dot{X}$ a momentum matrix, where m is the particle mass.

Max Born and Pascual Jordan, later together with Werner Heisenberg, further expanded the matrix method. For this, we start with the two matrices X and P . The matrices do not need to be commutative; that is, we do not require $XP = PX$. Therefore the difference

$$[P, X] \stackrel{\text{def}}{=} (PX - X\bar{P}), \quad (3.1)$$

called the commutator of the two matrices, is normally not zero. Differentiation of the commutator with respect to the time yields

$$\frac{d}{dt}[P, X] = [\dot{P}, X] + [P, \dot{X}]. \quad (3.2)$$

Newton's law states $m\ddot{X} = \dot{P} = f(X)$. Substitution in (3.2) provides

$$\frac{d}{dt}[P, X] = \underbrace{[f(X), X]}_{f(X)\bar{X} - Xf(X)} + \underbrace{[P, (1/m)\bar{P}]}_{(1/m)\bar{P}^2 - (1/m)\bar{P}^2} = \mathbf{0}. \quad (3.3)$$

Note that $f(\mathbf{X})$ is composed of the matrix \mathbf{X} and powers of the same. In this case, $f(\mathbf{X})$ commutes with \mathbf{X} ; that is,

$$f(\mathbf{X})\mathbf{X} = \mathbf{X}f(\mathbf{X}).$$

The commutator $[\mathbf{P}, \mathbf{X}]$ is thus a constant matrix! What is the form of this constant matrix? Its elements consist of sums of products of the form

$$\dot{x}(\ell, j)x(j, k) = i\omega(\ell, j)x(\ell, j)x(j, k)$$

and

$$x(\ell, j)\dot{x}(j, k) = i\omega(j, k)x(\ell, j)x(j, k).$$

A general matrix element of the commutator of $d\mathbf{X}/dt$ with \mathbf{X} is of the form

$$(\mathbf{X}\dot{\mathbf{X}} - \dot{\mathbf{X}}\mathbf{X})_{\ell k} = i \sum_j (\omega(\ell, j) - \omega(j, k)) x(\ell, j)x(j, k). \quad (3.4)$$

The time derivative of this commutator is

$$\begin{aligned} \frac{d}{dt}(\mathbf{X}\dot{\mathbf{X}} - \dot{\mathbf{X}}\mathbf{X})_{\ell k} &= i^2 \sum_j (\omega(\ell, j) - \omega(j, k)) (\omega(\ell, j) + \omega(j, k)) x(\ell, j)x(j, k) \\ &= i^2 \omega_{\ell k} \sum_j (\omega(\ell, j) - \omega(j, k)) x(\ell, j)x(j, k) = 0, \end{aligned}$$

by the Ritz combination principle. It was already proved that this quantity vanishes. An off-diagonal component of this result is $\omega(\ell, k) \neq 0$; therefore the remaining sum must vanish. But this is just an off-diagonal element of (3.4). Therefore only diagonal elements of (3.4) are nonzero. It follows that $[\mathbf{P}, \mathbf{X}]$ is a *diagonal matrix*.

What are the diagonal elements of the matrix $[\mathbf{P}, \mathbf{X}]$? Let us use the energy matrix

$$\mathbf{H} = \frac{1}{2m} \mathbf{P}^2 + V(\mathbf{q}) \quad (3.5)$$

for the total energy of the system. The total energy is constant, that is, $\dot{\mathbf{H}} = \mathbf{0}$, thus the energy matrix \mathbf{H} must also be constant. In fact, it must be a diagonal matrix like the commutator. Furthermore, the diagonal elements $H(i, i)$ must be related to the i th stationary state. It is therefore plausible to choose $H(i, i)$ to be exactly equal to the constant energy of the system state:

$$H(i, i) = E_i. \quad (3.6)$$

According to (2.26), we have

$$\dot{\mathbf{X}} = \frac{2\pi i}{\hbar} (\mathbf{H}\mathbf{X} - \mathbf{X}\mathbf{H}), \quad (3.7)$$

and with $\mathbf{P} = m\dot{\mathbf{X}}$ we get

$$\mathbf{P} = \frac{2\pi i m}{\hbar} [\mathbf{H}, \mathbf{X}]. \quad (3.8)$$

Because \mathbf{H} represents the total energy, it must be of the form

$$\mathbf{H} = \frac{1}{2m} \mathbf{P}^2 + V(\mathbf{X}), \quad (3.9)$$

with the potential energy $V(\mathbf{X})$. The commutator of the potential energy $V(\mathbf{X})$ with \mathbf{X} is equal to zero and inserting (3.9) into (3.8), one obtains

$$\mathbf{P} = \frac{\pi i}{\hbar} [\mathbf{P}^2, \mathbf{X}]. \quad (3.10)$$

With

$$[\mathbf{P}^2, \mathbf{X}] = \mathbf{P}^2 \mathbf{X} - \mathbf{X} \mathbf{P}^2 = \mathbf{P}^2 \mathbf{X} - \mathbf{P} \mathbf{X} \mathbf{P} + \mathbf{P} \mathbf{X} \mathbf{P} - \mathbf{X} \mathbf{P}^2 = \mathbf{P} [\mathbf{P}, \mathbf{X}] + [\mathbf{P}, \mathbf{X}] \mathbf{P}$$

Equation (3.10) can be simplified to

$$\mathbf{P} = \frac{\pi i}{\hbar} (\mathbf{P} [\mathbf{P}, \mathbf{X}] + [\mathbf{P}, \mathbf{X}] \mathbf{P}). \quad (3.11)$$

If we replace $[\mathbf{P}, \mathbf{X}]$ by the diagonal matrix \mathbf{D} (as discussed above), we get

$$\mathbf{P} = \frac{\pi i}{\hbar} (\mathbf{P} \mathbf{D} + \mathbf{D} \mathbf{P}). \quad (3.12)$$

Componentwise, this reads as

$$p(n, m) d_m + d_n p(n, m) = \frac{\hbar}{\pi i} p(n, m). \quad (3.13)$$

Under the assumption that $p(n, m) \neq 0$ for $n \neq m$, (3.13) supplies

$$d_m + d_n = \hbar / (\pi i).$$

Inasmuch as this condition applies to all m, n , it follows that $d_m = d_n = \hbar / (2\pi i)$, or

$$\mathbf{D} = \hbar / (2\pi i) \mathbf{I}. \quad (3.14)$$

Thus we have the final result for the commutation relation:

$$\boxed{\mathbf{P}X - X\mathbf{P} = \frac{h}{2\pi i}\mathbf{I}}. \quad (3.15)$$

Note that you can just as well ignore this “derivation” and simply *postulate* this commutation relation, as Born and Jordan did later. By the way, Dirac used a similar strategy with his “ q -numbers”.

From (3.15), the following commutation relations for powers of X and \mathbf{P} can be derived:

$$\mathbf{X}^n\mathbf{P} - \mathbf{P}\mathbf{X}^n = n\frac{i}{2\pi}\mathbf{X}^{n-1}, \quad (3.16)$$

and

$$\mathbf{P}^n\mathbf{X} - \mathbf{X}\mathbf{P}^n = -n\frac{i}{2\pi}\mathbf{P}^{n-1}. \quad (3.17)$$

Proof of (3.16) (by induction): The case $n = 1$ corresponds to (3.15). Now suppose that (3.16) is true for n . Multiplication from the left with X yields

$$\mathbf{X}^{n+1}\mathbf{P} - \mathbf{X}\mathbf{P}\mathbf{X}^n = n\frac{i}{2\pi}\mathbf{X}^n.$$

Calculate $\mathbf{X}\mathbf{P}$ from (3.15) and use it in the second term, which yields

$$\mathbf{X}^{n+1}\mathbf{P} - \left(\mathbf{P}\mathbf{X} + \frac{h}{2\pi i}\mathbf{I}\right)\mathbf{X}^n = n\frac{i}{2\pi}\mathbf{X}^n.$$

We can now bring the second term in parentheses to the right side, which yields (3.16) for $n + 1$. **q.e.d.** Note that (3.17) can be proved similarly.

From the commutation relation, we can derive a further useful relation. Let $f(\mathbf{P}, \mathbf{X})$ be any function of \mathbf{P} and \mathbf{X} . Then

$$\boxed{f\mathbf{X} - \mathbf{X}f = \frac{h}{2\pi i}\frac{\partial f}{\partial \mathbf{P}}} \quad (3.18)$$

and

$$\boxed{\mathbf{P}f - f\mathbf{P} = \frac{h}{2\pi i}\frac{\partial f}{\partial \mathbf{X}}}. \quad (3.19)$$

Proof Suppose that (3.18) and (3.19) were correct for any two functions f_1 and f_2 . Then they are also true for $f_1 + f_2$ and $f_1 \cdot f_2$. This fact is trivial for $f_1 + f_2$. For $f_1 \cdot f_2$, a short calculation shows

$$\begin{aligned} f_1 f_2 X - X f_1 f_2 &= f_1 (f_2 X - X f_2) + (f_1 X - X f_1) f_2 = \\ &= \frac{h}{2\pi i} \left(f_1 \frac{\partial f_2}{\partial P} + \frac{\partial f_1}{\partial P} f_2 \right) = \frac{h}{2\pi i} \frac{\partial(f_1 f_2)}{\partial P}, \end{aligned}$$

corresponding to $P f_1 f_2 - f_1 f_2 P$. Now these relations (3.18) and (3.19) obviously apply for P and X . Therefore, they also apply for each function f that can be expressed in powers of P and X . **q.e.d.**

If we solve any H for $P = p(X, H)$, and then choose $P = f$ in (3.18) we find

$$P X - X P = \frac{h}{2\pi i} \frac{\partial P}{\partial P} = \frac{h}{2\pi i} I. \quad (3.20)$$

Thus we have shown that the commutation relation (3.15) is valid in general for any Hamiltonian H .

Differentiation with Respect to a Matrix

When it comes to differentiating a matrix function with respect to a matrix, we use a very basic type of differentiation, namely if $F(X, Y, Z, \dots)$ is a function of the independent matrices X, Y, Z, \dots , then we define the *partial derivative* with respect to the matrix X as

$$\frac{\partial F}{\partial X} \stackrel{\text{def}}{=} \lim_{\epsilon \rightarrow 0} \frac{F(X + \epsilon I, Y, Z, \dots) - F(X, Y, Z, \dots)}{\epsilon}.$$

With this definition, matrix differentiations give similar results to “normal” differentiation; for example,

$$\frac{dX}{dX} = I$$

and

$$\begin{aligned} \frac{dX^2}{dX} &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \sum_k [(X_{nk} + \epsilon \delta_{nk})(X_{km} + \epsilon \delta_{km}) - X_{nk} X_{km}] = \\ &= (2X_{nm}) = 2X. \end{aligned}$$

For the differentiation of a product of two matrix functions, we get

$$\frac{\partial}{\partial X}(FG) = \frac{\partial F}{\partial X}G + F\frac{\partial G}{\partial X}. \quad (3.21)$$

This results from

$$\begin{aligned} \mathbf{F}(X + \epsilon \mathbf{I}, Y, \dots) \mathbf{G}(X + \epsilon \mathbf{I}, Y, \dots) - \mathbf{F}(X, Y, \dots) \mathbf{G}(X, Y, \dots) &= \\ = [\mathbf{F}(X + \epsilon \mathbf{I}, Y, \dots) - \mathbf{F}(X, Y, \dots)] \mathbf{G}(X + \epsilon \mathbf{I}, Y, \dots) & \\ + \mathbf{F}(X, Y, \dots) [\mathbf{G}(X + \epsilon \mathbf{I}, Y, \dots) - \mathbf{G}(X, Y, \dots)]. \end{aligned}$$

From (3.21), it follows directly that

$$\frac{dX^n}{dX} = nX^{n-1}.$$

Therefore one can differentiate polynomials and power series of matrices similar to “normal” differentiation. For the exponential of a matrix

$$e^X \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} \frac{X^i}{i!}$$

one obtains by termwise differentiation

$$\frac{d e^X}{dX} = e^X.$$

3.2 Systems with Several Degrees of Freedom

Thus far, we know that \mathbf{P} equals $m\dot{\mathbf{X}}$. Putting this formula in the commutation relation (3.15) provides a new form of the commutation relation:

$$\boxed{X \dot{X} - \dot{X} X = \frac{h}{2\pi m i} \mathbf{I}.} \quad (3.22)$$

From this form, it becomes obvious that the commutation relation refers to the entity X and its time derivative. However, this is not always the case, as we show later in the case of angular momentum. We can use this knowledge to generalize our findings to systems with several degrees of freedom:

$$X_k \mathbf{P}_k - \mathbf{P}_k X_k = \frac{-h}{2\pi i} \mathbf{I}. \quad (3.23)$$

Because the degree of freedom X_k does not depend on X_i for $i \neq k$, we get for $i \neq k$

$$X_k \mathbf{P}_i - \mathbf{P}_i X_k = \mathbf{0}, \quad (3.24)$$

$$\mathbf{X}_k \mathbf{X}_i - \mathbf{X}_i \mathbf{X}_k = \mathbf{0}, \quad (3.25)$$

and

$$\mathbf{P}_k \mathbf{P}_i - \mathbf{P}_i \mathbf{P}_k = \mathbf{0}. \quad (3.26)$$

In general, if $\mathbf{F}(\mathbf{P}, \mathbf{X})_k$ is a function of \mathbf{P}_k and \mathbf{X}_k , then the following relations hold.

$$[\mathbf{X}_k, \mathbf{F}] = -\frac{\hbar}{2\pi i} \frac{\partial \mathbf{F}}{\partial \mathbf{P}_k}, \quad (3.27)$$

$$[\mathbf{P}_k, \mathbf{F}] = \frac{\hbar}{2\pi i} \frac{\partial \mathbf{F}}{\partial \mathbf{X}_k}. \quad (3.28)$$

In classical physics, a mechanical system is characterized by its energy as a function of the momenta and coordinates. In quantum mechanics, a system is similarly characterized by specifying the *energy function* (energy matrix) $\mathbf{H}(\mathbf{P}, \mathbf{X})$. In close analogy to the classical equations of motion¹

$$\dot{x} = \frac{\partial H}{\partial p} = \{H, x\} \quad \text{and} \quad \dot{p} = -\frac{\partial H}{\partial x} = \{H, p\} \quad (3.29)$$

we use as quantum mechanical equations of motion these matrix relations, following from (3.7), (3.27), and (3.28):

$$\boxed{\dot{\mathbf{X}}_k = \frac{\partial \mathbf{H}}{\partial \mathbf{P}_k} = \frac{2\pi i}{\hbar} [\mathbf{H}, \mathbf{X}_k]}, \quad (3.30)$$

$$\boxed{\dot{\mathbf{P}}_k = -\frac{\partial \mathbf{H}}{\partial \mathbf{X}_k} = \frac{2\pi i}{\hbar} [\mathbf{H}, \mathbf{P}_k]}. \quad (3.31)$$

3.3 Transformations

Matrix differential equations of the form

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X},$$

¹Here, we use the Poisson–Jacobi bracket symbol:

$$\{F, G\} = -\{G, F\} = \sum_k \left(\frac{\partial F}{\partial p_k} \frac{\partial G}{\partial x_k} - \frac{\partial G}{\partial p_k} \frac{\partial F}{\partial x_k} \right).$$

where A is a constant matrix, have the solution

$$X(t) = \Phi(t)X(0),$$

with the transition matrix [15]

$$\Phi(t) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{1}{k!} (At)^k = \exp(At).$$

For the matrix equation

$$\dot{X} = \frac{i}{\hbar} [\mathbf{H}, X] \quad (3.32)$$

we can define a similar solution approach:

$$\boxed{\Phi(t) = \exp\left(\frac{i}{\hbar} \mathbf{H}t\right)}, \quad (3.33)$$

$$\boxed{X(t) = \Phi(t)X(0)\Phi^{-1}(t)}, \quad (3.34)$$

where $X(0)$ is the value of X at time $t = 0$. With $\dot{\Phi}(t) = \frac{i}{\hbar} \mathbf{H}\Phi(t)$, we then get indeed

$$\begin{aligned} \dot{X} &= \dot{\Phi}(t)X(0)\Phi^{-1}(t) + \Phi(t)X(0)\dot{\Phi}^{-1}(t) = \\ &= \frac{i}{\hbar} \mathbf{H}X(t) + \Phi(t)X(0)\left(-\frac{i}{\hbar} \mathbf{H}\Phi^{-1}\right) = \frac{i}{\hbar} [\mathbf{H}X(t) - X(t)\mathbf{H}]. \end{aligned}$$

Note that we assumed the commutativity of \mathbf{H} and $\Phi^{-1}(t) = \Phi(-t)$, which can be readily verified by expanding the series (3.33) of $\Phi(t)$.

The following summary [14] turns out to be very useful. If the matrices \mathbf{P}_k, X_k satisfy the canonical commutation rules (3.24)–(3.26) and also have the property that their Hamilton function $\mathbf{H}(\mathbf{P}, X)$ is a diagonal matrix, the canonical equations of motion (3.30) and (3.31) are satisfied. In other words, if only (3.24)–(3.26) are satisfied and \mathbf{H} is a diagonal matrix, you have already solved a given quantum mechanical problem!

By introducing the transformations

$$X \Rightarrow X' = \mathbf{T}XT^*$$

$$\mathbf{P} \Rightarrow \mathbf{P}' = \mathbf{T}\mathbf{P}\mathbf{T}^*$$

with

$$\mathbf{T} = \exp(-i\mathbf{H}t/\hbar) = \Phi(-t),$$

the coordinates and momenta become independent of time. However, a formerly constant vector \mathbf{v} now becomes a time-dependent vector $\mathbf{T}\mathbf{v} = \exp(-i\mathbf{H}t/\hbar)\mathbf{v}$. But $\exp(-i\mathbf{H}t/\hbar)\mathbf{v}$ is the solution of the differential equation

$$\dot{\mathbf{v}}(t) = -\frac{i}{\hbar}\mathbf{H}\mathbf{v}(t). \quad (3.35)$$

This is the Schrödinger equation for the state vector \mathbf{v} ; see Chap. 11. In quantum mechanics in the Heisenberg picture the state vector \mathbf{v} (see Chap. 4) does not change with time, whereas an observable \mathbf{A} satisfies the Heisenberg equation

$$\frac{d}{dt}\mathbf{A} = \frac{i}{\hbar}[\mathbf{H}, \mathbf{A}].$$

3.4 Problems

3.1 Commutation Relations: What is

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]]?$$

3.2 Powers in Commutation Relations: Show that

$$[X, \mathbf{P}^n] = n i \hbar \mathbf{P}^{n-1}.$$

3.3 Anti-Commutation Relation for Hermitian Matrices: Show that for Hermitian matrices A and B , the sum $AB + BA$ is also an Hermitian matrix.

3.4 Coordinate Transformation: What does X look like when a coordinate transformation is applied according to (3.34) with a diagonal matrix \mathbf{H} ?

3.5 Exponential of an Hermitian Matrix: Show that for an Hermitian matrix \mathbf{H} the matrix $\exp(i\mathbf{H})$ is a unitary matrix.

3.6 Commutation Relation: Show that for Hermitian matrices A and B , the matrix $i\hbar C = [AB - BA]$ is an Hermitian matrix.

3.7 Commutation Relation: Show that

$$a)[A, BC] = B[A, C] + [A, B]C$$

and

$$b)[AB, C] = A[B, C] + [A, C]B.$$

3.8 Exponential of a Nilpotent Matrix: What is the exponential matrix $\exp(tN)$ for the nilpotent matrix

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}?$$

Chapter 4

Observables and Uncertainty Relations

Abstract State vectors and other matrices are introduced. We also define projection matrices for the interpretation of experiments and density matrices for the description of mixed states. Also, Heisenberg's famous uncertainty relation is derived and interpreted.

4.1 State Vector

In classical physics, the dynamic behaviour of a system is completely described through the state vector

$$\xi \stackrel{\text{def}}{=} \begin{pmatrix} x \\ p \end{pmatrix}.$$

The temporal behaviour of the system is determined by a system of first order differential equations, the Hamilton equations

$$\dot{\xi} = \begin{pmatrix} \dot{x} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial x} \end{pmatrix}.$$

The state variables (e.g., in astronomy, the space coordinates x_i and the momentum components p_i of a celestial body) can be measured simultaneously at each given time. These physical entities are real numbers that can be measured with a finite precision, e.g. Δx_i . More specifically, the measured value is a multiple of this finite precision, i.e. $x_i = n_i \cdot \Delta x_i$, where n_i is an integer. Once the product $\Delta x_i \Delta p_i$ of the accuracies Δx_i and Δp_i reach the magnitude of the Planck constant h (which is obviously the case in quantum mechanics), x_i and p_i can no longer be measured simultaneously with arbitrary precision.

4.2 The Stern-Gerlach Experiment

In their famous Stern-Gerlach experiment in 1922, the physicists Otto Stern¹ and Walther Gerlach² were the first to observe the directional quantization of angular momenta of atoms. The experiment is one of the most fundamental experiments in physics, and it is repeatedly referred to for explaining quantum mechanical effects that cannot be understood in the framework of classical physics.

In this experiment, a beam of (electrically neutral) silver atoms is sent through a vacuum between the poles of a magnet. One of the poles has the form of a cutting edge (parallel to the beam), the other pole looks like a groove in a flat plane. This setup implies that the magnetic field is strongly inhomogeneous in the direction transverse to the beam. Eventually, the silver atoms are displayed on a screen. It turns out that the silver atoms can be found in two separate patches. In other words, the magnetic field splits the beam into two separate sub-beams.

Here is a quantum mechanical explanation of the experiment. The silver atom has a magnetic moment $\vec{\mu}$, which points into the same direction as its angular momentum \vec{S} . The magnetic field can be represented as

$$\vec{B}_{\text{total}} = \vec{B}_{\text{homogeneous}} + \vec{B}_{\text{inhomogeneous}} = (B_{\text{homogeneous}} + B_{\text{inhomogeneous}}) \cdot \vec{e}_z.$$

The angular momentum has a quantum number $\frac{1}{2}$. Therefore, only the settings $\frac{-\hbar}{2}$ or $\frac{+\hbar}{2}$ are allowed in the z-direction (also called “spin down” and “spin up”). In contrast, a classical angular momentum could point in any direction with respect to this axis. In the inhomogeneous field, the force

$$\vec{F} = -\nabla(-\vec{\mu} \cdot \vec{B}_{\text{total}}) = (\vec{\mu} \cdot \nabla)\vec{B}_{\text{total}} = \begin{pmatrix} 0 \\ 0 \\ \mu_z \cdot \frac{\partial B}{\partial z} \end{pmatrix}$$

acts on the magnetic moment of the atom. Since $\vec{\mu}$ is proportional to \vec{S} , the z-component of $\vec{\mu}$ can only have a positive value or a negative value of equal size. Depending on the orientation of the angular momentum, a force that is of the same value but opposite in direction therefore acts perpendicular to the flight direction of the atom. As a result, the beam splits into two sub-beams. From a classical point of view, the magnetic moment $\vec{\mu}$ can occupy any angle with respect to the z-axis. The deflection force would therefore also have a continuously distributed value, and the silver atoms would be visible in a continuous strip.

¹Otto Stern, 1888–1969, German physicist, Nobel Prize 1943.

²Walther Gerlach, 1889–1979, German physicist.

4.3 States and Postulates

In the Stern-Gerlach experiment, we can assign two states to the atoms:

$$\text{atoms with } \mu_z > 0 : \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$\text{atoms with } \mu_z < 0 : \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The total state is thus characterized by a vector of the form

$$\xi = \frac{1}{\sqrt{\alpha^2 + \beta^2}} \left(\alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \frac{1}{\sqrt{\alpha^2 + \beta^2}} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Quantum Mechanics is an axiomatic theory because it is well-grounded on few postulates. A postulate is a statement, also known as an axiom, which is taken to be true without proof. Postulates are the basic structure from which lemmas and theorems are derived.

Postulate 1 *The physical state of a quantum system at a time t_0 is completely described through a state vector ξ .*

In quantum mechanics, all physically measurable properties are described by matrices, for example X_i and P_i .

Postulate 2 *Each physically measurable property A is described by a Hermitian matrix A in the state space. A is called the observable.*

Since Hermitian matrices have only real eigenvalues, measurement results are always real, as expected.

Postulate 3 *The measurement of the physical property A always returns an eigenvalue of the matrix A .*

A quantum state \mathbf{a} is called an eigenstate of the matrix \mathbf{A} if the action of the matrix on the state returns the same state multiplied by some eigenvalue λ , that is, $\mathbf{A}\mathbf{a} = \lambda\mathbf{a}$. If the quantum system is in the state \mathbf{a} , then a measurement of the observable \mathbf{A} will give the result λ . Note that λ must be a real number (since anything that is physically measurable is a real number).

As an example, quantized results are observed in the Stern-Gerlach experiment. As we just saw, the distribution of possible magnetic moments μ_z is not continuous (as predicted by the classical theory), but limited to two values. These are exactly the eigenvalues ± 1 of the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It follows that a state variable can be fully described by a linear combination of eigenvectors. In order for the eigenvalues (which are now the measurements results) to be always real, the associated matrices must always be Hermitian. For example, the condition $X_i = \bar{X}_i^\top = X_i^\dagger$ must hold (i.e., the transposed matrix with complex conjugate elements is equal to the original matrix).

It can be shown that Hermitian matrices (that correspond to physical entities) have a complete set³ of eigenvectors. Such matrices are also called *observables*, and any possible measurement result of a physical entity corresponding to the observable \mathbf{A} can only be an eigenvalue λ_i of \mathbf{A} . If the state vector ξ can be composed of eigenvectors \mathbf{e}_i of \mathbf{A} , i.e.

$$\xi = \sum_i c_i \mathbf{e}_i, \quad (4.1)$$

then the probability for the measurement of \mathbf{A} to give the result λ_i is equal to $|c_i|^2$ (under the condition that the eigenvectors are orthonormal). For the case that the state vector is equal to an eigenvector \mathbf{e}_i , you will surely find the measurement result λ_i .

4.4 Projection Matrices

If the eigenvectors \mathbf{e}_i are normalized (i.e., $\mathbf{e}_i^\dagger \mathbf{e}_j = 1$ for $i = j$ and $\mathbf{e}_i^\dagger \mathbf{e}_j = 0$ for $i \neq j$), one obtains the coefficients c_j by multiplying (4.1) from the left with the row vector \mathbf{e}_j^\dagger :

$$\mathbf{e}_j^\dagger \xi = \sum_i c_i \mathbf{e}_j^\dagger \mathbf{e}_i = c_j. \quad (4.2)$$

³A complete set of eigenvectors is a set of eigenvectors so that every vector is a linear combination of the eigenvectors.

If we multiply this equation from the left with the column vector \mathbf{e}_j , we obtain

$$\mathbf{e}_j \mathbf{e}_j^\dagger \xi = \mathbf{P}_j \xi = \sum_i c_i \mathbf{e}_j \mathbf{e}_j^\dagger \mathbf{e}_i = \mathbf{e}_j \mathbf{e}_j^\dagger \sum_i c_i \mathbf{e}_i = \mathbf{P}_j \sum_i c_i \mathbf{e}_i = c_j \mathbf{e}_j. \quad (4.3)$$

The matrix

$$\mathbf{P}_j \stackrel{\text{def}}{=} \mathbf{e}_j \mathbf{e}_j^\dagger \quad (4.4)$$

is called *projection matrix*, since it projects the state vector ξ on the eigenvector \mathbf{e}_j . Projection matrices, also called *projection operators*, have the following properties:

- $\mathbf{P}_j^2 = \mathbf{P}_j$,
since $\mathbf{P}_j^2 = (\mathbf{e}_j \mathbf{e}_j^\dagger)(\mathbf{e}_j \mathbf{e}_j^\dagger) = \mathbf{e}_j (\mathbf{e}_j^\dagger \mathbf{e}_j) \mathbf{e}_j^\dagger = \mathbf{e}_j \mathbf{e}_j^\dagger = \mathbf{P}_j$.
- $\sum_i \mathbf{P}_i = \mathbf{I}$.

Therefore the choice

$$\mathbf{P}_2 = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and so on define valid projection matrices.

To some extent, the projection matrix describes the preparation of an experiment, since they are *mathematical models* for the experiment!

If we add a shield behind the magnetic field of the Stern-Gerlach apparatus such that the lower radiation component is blocked, then only the upper portion can reach the detector, i.e. only “eigenvectors”

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The projection matrix is thus composed as

$$\mathbf{P}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and we have

$$\mathbf{P}_1 \xi = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \left(\alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

4.5 Probabilistic Interpretation

We can specify the probability p_i that the result of a measurement is a_i . In order to determine this probability, a large number of measurements must be carried out under steady state conditions. The result of such measurement series is the *expectation value* $\langle A \rangle$ of the observable A . e.g., the spectral line wavelengths of an atom are the mean values based on many transitions from one atom state to another. In the state representation

$$\xi = \sum_i c_i \mathbf{e}_i,$$

the element $c_i \mathbf{e}_i$ describes the *possibility* that the measurement of the appropriate entity related to A gives the eigenvalue a_i . If ξ and \mathbf{e}_i are normalized, then according to Max Born the number $c_i^* c_i = |c_i|^2$ is the probability that the measurement gives the eigenvalue a_i . Of course, the probabilities are positive and add up to 1:

$$\sum_j c_j^* c_j = 1.$$

The measured mean value of A is then given by

$$\sum_j c_j^* c_j a_j = \xi^\dagger A \xi,$$

where it was agreed that if the coefficient c_i is complex, one uses the transposed vector ξ^\dagger as

$$\xi^\dagger = \sum_i c_i^* \mathbf{e}_i^\top,$$

The above formula can be easily verified, since the eigenvalue equation $A\mathbf{e}_i = a_i \mathbf{e}_i$ yields

$$A\xi = A \sum_i c_i \mathbf{e}_i = \sum_i c_i A \mathbf{e}_i = \sum_i c_i a_i \mathbf{e}_i.$$

Multiplying with ξ^\dagger from the left, we finally get (with $\mathbf{e}_i^\dagger \mathbf{e}_j = \delta_{ij}$)

$$\xi^\dagger A \xi = \left(\sum_j c_j^* \mathbf{e}_j^\dagger \right) \left(\sum_i c_i a_i \mathbf{e}_i \right) = \sum_j |c_j|^2 a_j = \langle A \rangle. \quad (4.5)$$

Let us sum up our findings:

The probability to measure the eigenvalue a_j of A in the state ξ is given by $|c_j^2|$. $c_j e_j$ is the projection of the normalized vector ξ on the normalized eigenvector e_j . The average value of the observable A for a system being in the state ξ is $\langle A \rangle = \xi^\dagger A \xi$. The expectation values are extremely important in quantum mechanics. If the measurement of the observable A is repeated a large number of times, the average of all the results will approach $\langle A \rangle$. The same is true if we measure the observable A in many independent yet identical systems simultaneously.

4.6 Density Matrix

4.6.1 Definitions

Until now, we considered systems that consist of different particles in specific states. Let now an *ensemble* be a collection of many identical particles. They may be in different states. However, assume the statistical distribution of these states is known. Albert Einstein was an early advocate of the ensemble interpretation [28]:

“The attempt to conceive the quantum-theoretical description as the complete description of the individual systems leads to unnatural theoretical interpretations, which become immediately unnecessary if one accepts the interpretation that the description refers to ensembles of systems and not to individual systems.”

While the state of a single particle is described by a vector, the state of an ensemble of particles is best described by a *state matrix*, also called *density matrix*. The density matrix D is obtained as follows. In the definition (4.5) of the expectation value $\langle A \rangle$, we can insert the identity matrix in the special form

$$I = \sum_i I_i = \sum_i e_i e_i^\top,$$

where I_i is a diagonal matrix that is 1 only in the i -th diagonal element and zero otherwise:

$$\begin{aligned} \langle A \rangle &= \xi^\dagger A \xi = \xi^\dagger A I \xi = \xi^\dagger A \left(\sum_i I_i \right) \xi = \\ &= \xi^\dagger A \left(\sum_i e_i e_i^\top \right) \xi = \sum_i \underbrace{\xi^\dagger A e_i}_{\xi^\dagger A e_i} \underbrace{e_i^\top \xi}_{e_i^\top \xi} = \sum_i \underbrace{e_i^\top \xi}_{e_i^\top \xi} \underbrace{\xi^\dagger A e_i}_{\xi^\dagger A e_i}. \end{aligned} \quad (4.6)$$

Now we can introduce the density matrix \mathbf{D} as

$$\mathbf{D} \stackrel{\text{def}}{=} \xi \xi^\dagger = \sum_i c_i \mathbf{e}_i \cdot \sum_j c_j^* \mathbf{e}_j^\top = \sum_{i,j} c_i \mathbf{e}_i \cdot c_j^* \mathbf{e}_j^\top. \quad (4.7)$$

The matrix elements of \mathbf{D} in the \mathbf{e} -basis are

$$d(k, \ell) = \mathbf{e}_k^\top \mathbf{D} \mathbf{e}_\ell = \mathbf{e}_k^\top \sum_{i,j} c_i \mathbf{e}_i \cdot c_j^* \mathbf{e}_j^\top = c_k c_\ell^*.$$

Generally the density matrix is not diagonal. The *trace of a matrix* is defined as the sum of the elements on the main diagonal. It can be written by means of the normalized eigenvectors \mathbf{e}_ν as

$$\text{trace}(X) \stackrel{\text{def}}{=} \sum_\nu \mathbf{e}_\nu^\top X \mathbf{e}_\nu. \quad (4.8)$$

With this definition, we finally get from (4.6) and (4.7)

$$\boxed{\langle A \rangle = \text{trace}(\mathbf{D}A).} \quad (4.9)$$

The expectation value of a dynamical variable A , represented by the matrix A , in the state \mathbf{D} , is (4.9).

If ξ is a normalized state vector (i.e. length 1), it follows from (4.7)

$$\text{trace}(\mathbf{D}) = \sum_i |c_i|^2 = 1. \quad (4.10)$$

Also, we get

$$\mathbf{D}^2 = \mathbf{D}, \quad (4.11)$$

since

$$\mathbf{D}^2 = \xi \underbrace{\xi^\dagger \xi}_{1} \xi^\dagger = \xi \xi^\dagger = \mathbf{D}.$$

4.6.2 Mixed States

If all particles of the considered ensemble are in a definite quantum state, it is called a *pure state*. In this case, the measurement result according to (4.9) is obviously the expectation value. If, however, the particles are in different quantum states, we call it

a *mixed state*. Now, if N_ν out of N particles are in the state ξ_ν , then the probability⁴ to pick such a particle from the entire ensemble equals

$$p_\nu = \frac{N_\nu}{N},$$

which certainly assumes

$$\sum_\nu p_\nu = 1.$$

For the expectation value, we obtain

$$\langle A \rangle = \sum_\nu p_\nu \xi_\nu^\dagger A \xi_\nu. \quad (4.12)$$

For mixed states, we can now modify the density matrix as

$$\mathbf{D}_M \stackrel{\text{def}}{=} \sum_\nu p_\nu \xi_\nu \xi_\nu^\dagger. \quad (4.13)$$

This provides a straightforward extension of the state concept, since we can now also describe systems whose state is not known in all detail (e.g., as a state vector in a Hilbert space). For this modified density matrix \mathbf{D}_M , we again have

$$\begin{aligned} \underline{\underline{\text{trace}(\mathbf{D}_M A)}} &= \sum_\nu \sum_\mu e_\nu^\top p_\mu \xi_\mu \xi_\mu^\dagger A e_\nu = \sum_\nu \sum_\mu p_\mu \xi_\mu^\dagger A e_\nu e_\nu^\top \xi_\mu = \\ &= \sum_\mu p_\mu \xi_\mu^\dagger A \xi_\mu = \underline{\underline{\langle A \rangle}}. \end{aligned} \quad (4.14)$$

However, we now have

$$\mathbf{D}_M^2 \neq \mathbf{D}_M, \quad (4.15)$$

since

$$\mathbf{D}_M^2 = \sum_\nu p_\nu \xi_\nu \xi_\nu^\dagger \sum_\mu p_\mu \xi_\mu \xi_\mu^\dagger = \sum_\nu p_\nu^2 \xi_\nu \xi_\nu^\dagger. \quad (4.16)$$

If you compare (4.13) with (4.16), the two are equal only if one $p_\nu = 1$, and all others are equal to zero. This is only the case, however, if the state is a *pure state*. For a *mixed state*, equation (4.15) is always true.

⁴The same holds if we only have incomplete information on the system, for example if the particle number is very large and when we can only make probability statements.

For the trace of \mathbf{D} , we obtain

$$\text{trace}(\mathbf{D}) = \sum_{\nu} \mathbf{e}_{\nu}^{\top} \boldsymbol{\xi}_{\nu} \boldsymbol{\xi}_{\nu}^{\dagger} \mathbf{e}_{\nu} = \sum_{\nu} \boldsymbol{\xi}_{\nu}^{\dagger} \mathbf{e}_{\nu} \mathbf{e}_{\nu}^{\top} \boldsymbol{\xi}_{\nu} = \sum_{\nu} \boldsymbol{\xi}_{\nu}^{\dagger} \boldsymbol{\xi}_{\nu} = 1, \quad (4.17)$$

and since $\mathbf{D}^2 = \mathbf{D}$, we also get

$$\text{trace}(\mathbf{D}^2) = 1. \quad (4.18)$$

For the mixed state is obtained, however,

$$\begin{aligned} \text{trace}(\mathbf{D}_M^2) &= \sum_{\nu} \mathbf{e}_{\nu}^{\top} \left(\sum_{\mu} p_{\mu} \boldsymbol{\xi}_{\mu} \boldsymbol{\xi}_{\mu}^{\dagger} \sum_{\kappa} p_{\kappa} \boldsymbol{\xi}_{\kappa} \boldsymbol{\xi}_{\kappa}^{\top} \right) \mathbf{e}_{\nu} = \\ &= \sum_{\nu} \mathbf{e}_{\nu}^{\top} \left(\sum_{\mu} p_{\mu}^2 \boldsymbol{\xi}_{\mu} \boldsymbol{\xi}_{\mu}^{\dagger} \right) \mathbf{e}_{\nu} = \sum_{\nu} p_{\nu}^2 < 1. \end{aligned} \quad (4.19)$$

Note that $\text{trace}(\mathbf{D}_M) = 1$ holds as well. With the trace of the squared density matrix, we now have a tool to determine whether the system is in a pure or a mixed state!

4.6.3 Examples

1. For the state vector

$$\boldsymbol{\xi} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

one obtains

$$\mathbf{D} = \boldsymbol{\xi} \boldsymbol{\xi}^{\dagger} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

In fact

$$\mathbf{D}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{D}.$$

2. For

$$\boldsymbol{\xi} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$$

one gets

$$\mathbf{D} = \boldsymbol{\xi} \boldsymbol{\xi}^{\dagger} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \frac{1}{\sqrt{2}} (1 \ -i) = \begin{pmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{i}{2} & \frac{1}{2} \end{pmatrix}.$$

Again we have $\mathbf{D}^2 = \mathbf{D}$, and for the trace we get $\text{trace}(\mathbf{D}) = 1$.

3. In the case of the Stern-Gerlach apparatus, we assigned the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to the two possible states. If the two vectors occur with the probabilities p_1 and p_2 , we obtain the density matrix

$$\mathbf{D}_M = p_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ 0) + p_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0 \ 1) = \begin{pmatrix} p_1 & 0 \\ 0 & p_2 \end{pmatrix}.$$

In this case, we get

$$\mathbf{D}_M^2 = \begin{pmatrix} p_1^2 & 0 \\ 0 & p_2^2 \end{pmatrix},$$

i.e., $\mathbf{D}_M^2 \neq \mathbf{D}_M$ and $\text{trace}(\mathbf{D}_M) = 1$. However $\text{trace}(\mathbf{D}_M^2) < 1$, since $p_1 < 1$, $p_2 < 1$ and $p_1 + p_2 = 1$ yields $p_1^2 + p_2^2 = 1 - p_1 p_2 < 1$.

4.7 Time Evolution of the Expectation Value

Although the state matrix \mathbf{D} is constant, the expectation value $\langle A \rangle = \langle \xi^\dagger A \xi \rangle = \text{trace}(\mathbf{D}A)$ of an observable A varies over time. This can be readily seen with the help of the Heisenberg equation of motion (if A does not explicitly depend on time):

$$\frac{d}{dt} A = \frac{i}{\hbar} [\mathbf{H} A] = \frac{i}{\hbar} (\mathbf{H} A - A \mathbf{H}), \quad (4.20)$$

and thus

$$\frac{d}{dt} \langle A \rangle = \frac{i}{\hbar} \langle (\mathbf{H} A - A \mathbf{H}) \rangle,$$

or, respectively,

$$\frac{d}{dt} \langle A \rangle = \text{trace}(\mathbf{D} \dot{A}).$$

Let us now sum up our findings. A single particle is described by a quantum state ξ . The only thing that you can measure, however, is the expectation value $\langle A \rangle = \langle \xi^\dagger A \xi \rangle$ of that state. In contrast, the density matrix \mathbf{D} describes everything that there is to know about the state of an ensemble of particles. If we consider a single particle of this ensemble, we know that it is in the state \mathbf{D} and that the expectation value is $\langle A \rangle = \text{trace}(\mathbf{D}A)$.

4.8 Heisenberg's Uncertainty Principle

In addition to the average value $\langle A \rangle$ (also called the *expectation value*), the *variance* is a good indicator for how strongly the measured values fluctuate. The variance is defined as the average of the squared deviations from the mean value $\langle A \rangle$:

$$(\Delta A)^2 \stackrel{\text{def}}{=} \langle (A - \langle A \rangle I)^2 \rangle = \langle A^2 - 2A \langle A \rangle + \langle A \rangle^2 I \rangle = \langle A^2 \rangle - \langle A \rangle^2. \quad (4.21)$$

Let us now assume that we would like to measure the two physical parameters a and b , represented by two matrices A and B . Let ξ be a normalized state vector. The expectation values of a and b are then

$$\langle A \rangle = \xi^\dagger A \xi$$

and

$$\langle B \rangle = \xi^\dagger B \xi.$$

The average fluctuation squares are given by

$$(\Delta A)^2 = \xi^\dagger (A - \langle A \rangle I)^2 \xi$$

and

$$(\Delta B)^2 = \xi^\dagger (B - \langle B \rangle I)^2 \xi.$$

We define a complex matrix

$$M \stackrel{\text{def}}{=} (A - \langle A \rangle I) + i \alpha (B - \langle B \rangle I),$$

where $\alpha > 0$ and real. Due to $(M\xi)^\dagger (M\xi) \geq 0$, we get

$$\begin{aligned} (M\xi)^\dagger (M\xi) &= \xi^\dagger M^\dagger M \xi = \\ &= \xi^\dagger [(A - \langle A \rangle I)^2 + \alpha^2 (B - \langle B \rangle I)^2 + i \alpha (AB - BA)] \xi = \\ &= (\Delta A)^2 + \alpha^2 (\Delta B)^2 + i \alpha \xi^\dagger (AB - BA) \xi \geq 0. \end{aligned}$$

It follows⁵

$$\alpha^{-1} (\Delta A)^2 + \alpha (\Delta B)^2 \geq -i \xi^\dagger [A, B] \xi = -i \langle [A, B] \rangle. \quad (4.22)$$

Varying α for fixed ΔA and ΔB , we find that the left side of this inequality has its

⁵With $\langle [A, B] \rangle \stackrel{\text{def}}{=} \xi^\dagger [A, B] \xi$.

minimum when α satisfies the equation

$$-\alpha^{-2}(\Delta A)^2 + (\Delta B)^2 = 0,$$

i.e., $\alpha = \Delta A / \Delta B$. For this value of α , the inequality (4.22) reads

$$2\Delta A \Delta B \geq -i \langle [A, B] \rangle,$$

that is,

$$\boxed{\Delta A \Delta B \geq -i \frac{1}{2} \langle [A, B] \rangle} \quad (4.23)$$

$$\boxed{.}$$

For $A = X$ and $B = P$, we get in particular

$$\langle [X, P] \rangle = \langle i\hbar I \rangle = i\hbar,$$

and therefore

$$\boxed{\Delta X \Delta P \geq \frac{\hbar}{2}} \quad (4.24)$$

$$\boxed{.}$$

This is Heisenberg's famous *uncertainty relation*! It shows that the uncertainty ΔP of the momentum must increase to the extent that the uncertainty ΔX of the coordinate decreases, and vice versa. Note, however, that the inequality only holds if $AB \neq BA$ (which is the case here).

The situation changes for systems with multiple degrees of freedom. The equations (3.23)–(3.26) state that X_k does not depend on X_i for $k \neq i$ and therefore $X_k P_i = P_i X_k$, $X_k X_i = X_i X_k$ and $P_k P_i = P_i P_k$. Therefore we get for these observables

$$\Delta X_k \Delta P_i \geq 0 (k \neq i),$$

$$\Delta X_k \Delta X_i \geq 0 (k \neq i)$$

and

$$\Delta P_k \Delta P_i \geq 0 (k \neq i).$$

In other words, we can simultaneously determine both observables of *different particles* with arbitrary precision!

4.9 Problems

- 4.1 Uncertainty Principle:** The speed of an electron is $1000 \frac{m}{s}$, and it is measured with an accuracy of 0.1%. With which accuracy can you measure the position of the electron?
- 4.2 Projection Matrices:** Under which condition is the product matrix $\mathbf{P} = \mathbf{P}_1 \cdot \mathbf{P}_2$ of two projection matrices \mathbf{P}_1 and \mathbf{P}_2 also a projection matrix?
- 4.3 Density Matrix:** What are the density matrices for the two pure spin 1/2 states, if the system is in the state a) $e_{3,1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and b) $e_{2,1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$?
- 4.4 Projection Matrix:** What are the eigenvalues and eigenvectors of a measurement device that is represented by the projection matrix $\mathbf{M} = e_1 e_1^\top + e_2 e_2^\top$, where e_1 and e_2 are orthogonal vectors?

Chapter 5

The Harmonic Oscillator

Abstract As a first application of the matrix method, the quantum mechanical behavior of the harmonic oscillator is discussed in detail.

5.1 Physics of the Harmonic Oscillator

A typical harmonic oscillator consists of a mass attached to a spring. The oscillator has always been a fruitful dynamic model for an atom or molecule in quantum physics. For example, it is often assumed that the electrons in an atom are elastically coupled (one-dimensional elastic oscillators). Such a one-dimensional system can be mathematically described by

$$m\ddot{x} = -kx,$$

where m is the mass, x is the displacement from the equilibrium position, and k is the spring constant.¹ This can also be written as

$$\ddot{x} + \omega_0^2 x = 0, \quad (5.1)$$

with the angular frequency

$$\omega_0 \stackrel{\text{def}}{=} \sqrt{k/m}.$$

The Hamiltonian function for this simple system is

$$H = \frac{m}{2}\dot{x}^2 + \frac{m}{2}\omega_0^2 x^2,$$

where the first term is the kinetic energy and the second term is the potential energy. If we introduce the momentum $p = m\dot{x}$, it becomes

¹We assume that the restoring force f is proportional to the deflection; that is, $f = -kx$.

$$H = \frac{1}{2m} p^2 + \frac{m}{2} \omega_0^2 x^2. \quad (5.2)$$

Heisenberg [13], and later also Born and Jordan in a more sustainable mathematical framework [5], used this one-dimensional harmonic oscillator as a first case study for quantum mechanics. They tried to determine for this model on which energy levels the electrons can move. For the corresponding quantum system, we use the expression

$$\mathbf{H} = \frac{1}{2m} \mathbf{P}^2 + \frac{m}{2} \omega_0^2 \mathbf{X}^2 \quad (5.3)$$

as the matrix for the total energy. This equation can be slightly modified into

$$\mathbf{H} = \hbar \omega_0 \left(\underbrace{\frac{1}{2m\omega_0 \hbar} \mathbf{P}^2}_{\tilde{\mathbf{P}}^2} + \underbrace{\frac{m\omega_0}{2\hbar} \mathbf{X}^2}_{\tilde{\mathbf{X}}^2} \right). \quad (5.4)$$

Therefore it is better to use scaled location and impulse operators:

$$\tilde{\mathbf{X}} \stackrel{\text{def}}{=} \sqrt{\frac{m\omega_0}{2\hbar}} \mathbf{X}, \quad \text{respectively} \quad \mathbf{X} = \sqrt{\frac{2\hbar}{m\omega_0}} \tilde{\mathbf{X}},$$

and

$$\tilde{\mathbf{P}} \stackrel{\text{def}}{=} \sqrt{\frac{1}{2m\omega_0 \hbar}} \mathbf{P}, \quad \text{respectively} \quad \mathbf{P} = \sqrt{2m\omega_0 \hbar} \tilde{\mathbf{P}}.$$

From basic algebra, we know that $(a + ib)(a - ib) = a^2 + b^2$. For noncommuting matrices \mathbf{A} and \mathbf{B} , however, we obtain

$$(\mathbf{A} + i\mathbf{B})(\mathbf{A} - i\mathbf{B}) = \mathbf{A}^2 + \mathbf{B}^2 + i(\mathbf{B}\mathbf{A} - \mathbf{A}\mathbf{B}).$$

Therefore

$$(\tilde{\mathbf{X}} + i\tilde{\mathbf{P}})(\tilde{\mathbf{X}} - i\tilde{\mathbf{P}}) = \tilde{\mathbf{X}}^2 + \tilde{\mathbf{P}}^2 - i(\tilde{\mathbf{X}}\tilde{\mathbf{P}} - \tilde{\mathbf{P}}\tilde{\mathbf{X}}). \quad (5.5)$$

Let us now introduce new terms for the matrices in parentheses on the left-hand side of the equation:

$$\mathbf{A} \stackrel{\text{def}}{=} \tilde{\mathbf{X}} + i\tilde{\mathbf{P}}. \quad (5.6)$$

Because \mathbf{X} and \mathbf{P} are Hermitian matrices (and therefore also $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{P}}$), we get

$$\mathbf{A}^\dagger = \tilde{\mathbf{X}} - i\tilde{\mathbf{P}}. \quad (5.7)$$

A and A^\dagger do not commute, inasmuch as

$$\begin{aligned} \underline{[A, A^\dagger]} &= AA^\dagger - A^\dagger A = (\tilde{X} + i\tilde{\mathbf{P}})(\tilde{X} - i\tilde{\mathbf{P}}) - (\tilde{X} - i\tilde{\mathbf{P}})(\tilde{X} + i\tilde{\mathbf{P}}) = \\ &= -\frac{i}{2\hbar} \underbrace{(\mathbf{X}\mathbf{P} - \mathbf{P}\mathbf{X})}_{i\hbar\mathbf{I}} + \frac{i}{2\hbar} \underbrace{(\mathbf{P}\mathbf{X} - \mathbf{X}\mathbf{P})}_{-i\hbar\mathbf{I}} = \underline{\mathbf{I}}. \end{aligned} \quad (5.8)$$

Note that we used the fact that the matrices \mathbf{X} and \mathbf{P} obey the commutation relation

$$\mathbf{X}\mathbf{P} - \mathbf{P}\mathbf{X} = i\hbar\mathbf{I}. \quad (5.9)$$

Also, we can calculate

$$A^\dagger A = \tilde{X}^2 + \tilde{\mathbf{P}}^2 - \frac{1}{2}\mathbf{I},$$

and therefore

$$\tilde{X}^2 + \tilde{\mathbf{P}}^2 = A^\dagger A + \frac{1}{2}\mathbf{I}.$$

With these findings, we can modify (5.4) into

$$\mathbf{H} = \hbar\omega_0 \left(A^\dagger A + \frac{1}{2}\mathbf{I} \right). \quad (5.10)$$

The following commutation relations hold:

$$[\mathbf{H}, A] = [\hbar\omega_0 A^\dagger A, A] = \hbar\omega_0 [A^\dagger, A]A = -\hbar\omega_0 A, \quad (5.11)$$

and

$$[\mathbf{H}, A^\dagger] = [\hbar\omega_0 A^\dagger A, A^\dagger] = \hbar\omega_0 A^\dagger [A, A^\dagger] = \hbar\omega_0 A^\dagger. \quad (5.12)$$

In the following, we are looking for the eigenvalues of the energy matrix \mathbf{H} , because they represent (in a simple atomic model) the energy levels for the electrons and define the radiation that they absorb or release when jumping from one level to another. Let us suppose that we found an eigenvector \mathbf{e} of \mathbf{H} . Then

$$\mathbf{H}\mathbf{e} = \lambda \mathbf{e}. \quad (5.13)$$

Now multiplying (5.11) by \mathbf{e} from the right yields

$$\mathbf{H}\mathbf{A}\mathbf{e} - \mathbf{A}\mathbf{H}\mathbf{e} = -\hbar\omega_0 \mathbf{A}\mathbf{e}.$$

With (5.13), we have

$$\underline{\underline{\mathbf{H}\mathbf{A}\mathbf{e}}} = (\lambda - \hbar\omega_0) \underline{\underline{\mathbf{A}\mathbf{e}}}. \quad (5.14)$$

In other words: if \mathbf{e} is an eigenvector of \mathbf{H} with eigenvalue λ , $A\mathbf{e}$ is also an eigenvector of \mathbf{H} , but with the eigenvalue $\lambda - \hbar\omega_0$. The eigenvalue has thus been reduced by the energy $\hbar\omega_0 = h\nu_0$.

Let us now multiply (5.11) by $A\mathbf{e}$ from the right. Using (5.14), we get

$$\mathbf{H}A^2\mathbf{e} - \underbrace{\mathbf{H}A\mathbf{A}\mathbf{e}}_{(\lambda - \hbar\omega_0)\mathbf{A}^2\mathbf{e}} = -\hbar\omega_0 A^2\mathbf{e},$$

which can be rewritten as

$$\underline{\underline{\mathbf{H}A^2\mathbf{e}}} = (\lambda - 2\hbar\omega_0)\mathbf{A}^2\mathbf{e}. \quad (5.15)$$

In other words, the eigenvalue was once again reduced by the energy $\hbar\omega_0$. This process can be continued for arbitrary powers of A . However, the eigenvalue of the energy \mathbf{H} can never become negative! After all, if we multiply \mathbf{H} in the form (5.10) from left and right with an eigenvector \mathbf{e} , we obtain

$$\mathbf{e}^\top \mathbf{H} \mathbf{e} = \lambda \mathbf{e}^\top \mathbf{e} = \hbar\omega_0 \left(\mathbf{e}^\top A^\dagger A \mathbf{e} + \frac{1}{2} \mathbf{e}^\top \mathbf{e} \right) = \underbrace{(\mathbf{A}\mathbf{e} \cdot \mathbf{A}\mathbf{e})}_{\geq 0} + \frac{1}{2} \underbrace{\mathbf{e}^\top \mathbf{e}}_{\geq 0} \geq 0;$$

that is, the eigenvalues λ of \mathbf{H} are all positive. For this reason, the above-explained reduction of energy has to come to an end at some point. Let this point be the basic eigenvector \mathbf{e}_0 . Then we must have $A\mathbf{e}_0 = \mathbf{0}$, because we could otherwise continue with another reduction loop. For the eigenvalue equation of the Hamiltonian, we obtain

$$\underline{\underline{\mathbf{H}\mathbf{e}_0}} = \left(\hbar\omega_0 A^\dagger A + \frac{1}{2} \hbar\omega_0 \right) \mathbf{e}_0 = \frac{1}{2} \hbar\omega_0 \mathbf{e}_0. \quad (5.16)$$

Multiplying (5.12) with \mathbf{e}_0 from the right, we obtain

$$\mathbf{H}A^\dagger \mathbf{e}_0 - A^\dagger \mathbf{H} \mathbf{e}_0 = \hbar\omega_0 A^\dagger \mathbf{e}_0,$$

and therefore

$$\underline{\underline{\mathbf{H}A^\dagger \mathbf{e}_0}} = \hbar\omega_0 \left(1 + \frac{1}{2} \right) A^\dagger \mathbf{e}_0. \quad (5.17)$$

We see that the eigenvalue is now *increased* by the energy $\hbar\omega_0$. For that reason, the matrix A^\dagger is also called the *creation operator*. If we multiply (5.12) once again with the new eigenvector $A^\dagger \mathbf{e}_0$ from the right, we get

$$\mathbf{H}(A^\dagger)^2 \mathbf{e}_0 - A^\dagger \mathbf{H} A^\dagger \mathbf{e}_0 = \hbar\omega_0 (A^\dagger)^2 \mathbf{e}_0,$$

and therefore

$$\underline{\underline{H(A^\dagger)^2 e_0 = \hbar\omega_0 \left(2 + \frac{1}{2}\right) (A^\dagger)^2 e_0.}}$$
(5.18)

As a general result, we find

$$e_n \stackrel{\text{def}}{=} (A^\dagger)^n e_0,$$

where n can be an arbitrary large positive integer. We then get

$$\underline{\underline{H e_j = \underbrace{\hbar\omega_0 \left(j + \frac{1}{2}\right)}_{\lambda_j = E_j} e_j, \quad j = 0, 1, 2, \dots,}}$$
(5.19)

which implies

$$\boxed{\lambda_j = E_j = \hbar\omega_0 \left(j + \frac{1}{2}\right), \quad j = 0, 1, 2, \dots}$$
(5.20)

In other words, you can pump as much energy into the system as you like. However, the lowest energy level is $E_0 = \frac{\hbar\omega_0}{2}$ (and not 0 as in classical theory). We can collect all energy levels in the diagonal matrix

$$\boxed{E = \hbar\omega_0 \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \dots \\ 0 & \frac{3}{2} & 0 & 0 & \dots \\ 0 & 0 & \frac{5}{2} & 0 & \dots \\ 0 & 0 & 0 & \frac{7}{2} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.}$$
(5.21)

The energy spectrum consists of equidistant energy values with the distance $\hbar\omega_0$. When jumping from one level to the next, the harmonic oscillator can therefore absorb or emit only integral multiples of $\hbar\omega_0$ (Fig. 5.1).

What is the shape of the matrices X and P for the harmonic oscillator? If we adopt Heisenberg's assumption that only transitions between neighboring states are possible (i.e., only $x(k, k+1) \neq 0$ and $x(k+1, k) \neq 0$), one obtains the matrix

$$X = \begin{pmatrix} 0 & x(0, 1) & 0 & 0 & 0 & \dots \\ x(1, 0) & 0 & x(1, 2) & 0 & 0 & \dots \\ 0 & x(2, 1) & 0 & x(2, 3) & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$
(5.22)

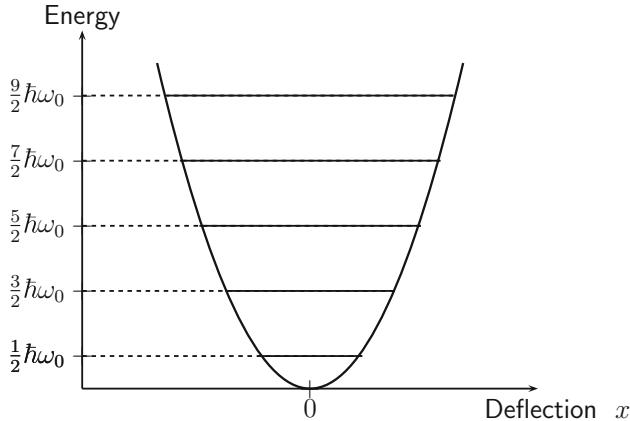


Fig. 5.1 The parabolic potential energy and the equidistant energy levels of the harmonic oscillator (to be continued upwards to infinity)

Because of

$$p(k, \ell) = m\dot{x}(k, \ell) = m i \omega(k, \ell)x(k, \ell), \quad (5.23)$$

we can calculate the respective matrix for \mathbf{P} :

$$\mathbf{P} = \begin{pmatrix} 0 & p(0, 1) & 0 & 0 & 0 & \dots \\ p(1, 0) & 0 & p(1, 2) & 0 & 0 & \dots \\ 0 & p(2, 1) & 0 & p(2, 3) & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \quad (5.24)$$

Let us now derive the shape of \mathbf{X} and \mathbf{P} in a more sophisticated way. From (5.10), we get

$$\begin{aligned} \mathbf{H}\mathbf{e}_n &= \hbar\omega_0 \mathbf{A}^\dagger \mathbf{A} \mathbf{e}_n + \frac{\hbar\omega_0}{2} \mathbf{e}_n = \\ &= \hbar\omega_0 \left(n + \frac{1}{2} \right) \mathbf{e}_n, \end{aligned}$$

and therefore with (5.19)

$$\mathbf{A}^\dagger \mathbf{A} \mathbf{e}_n = n \mathbf{e}_n. \quad (5.25)$$

The matrix $\mathbf{A}^\dagger \mathbf{A}$ is obviously an Hermitian matrix; that is, its eigenvalues are all real and its eigenvectors form an orthogonal basis. Moreover, one can descend in the series of eigenvectors \mathbf{e}_n by multiplying from the left with \mathbf{A} :

$$\mathbf{A}\mathbf{e}_n = \alpha_n \mathbf{e}_{n-1}. \quad (5.26)$$

We multiply this result by the complex conjugate from the left, which yields

$$\mathbf{e}_n^\dagger \mathbf{A}^\dagger \mathbf{A} \mathbf{e}_n = \mathbf{e}_n^\dagger \mathbf{e}_n |\alpha_n|^2 \stackrel{!}{=} |\alpha_n|^2, \quad (5.27)$$

in order for the length of the eigenvectors to be 1. With (5.25), we get from (5.27)

$$n \mathbf{e}_n^\dagger \mathbf{e}_n = n = |\alpha_n|^2. \quad (5.28)$$

Choosing α_n as a real number, we get

$$\alpha_n = \sqrt{n}.$$

Equation (5.26) now reads

$$\mathbf{A}\mathbf{e}_n = \sqrt{n} \mathbf{e}_{n-1}. \quad (5.29)$$

Using (5.25), we find

$$\mathbf{A}^\dagger \mathbf{A} \mathbf{e}_n = n \mathbf{e}_n = \mathbf{A}^\dagger (\sqrt{n} \mathbf{e}_{n-1});$$

in other words,

$$\mathbf{A}^\dagger \mathbf{e}_{n-1} = \sqrt{n} \mathbf{e}_n,$$

or (shifting to the index n rather than $n - 1$)

$$\mathbf{A}^\dagger \mathbf{e}_n = \sqrt{n+1} \mathbf{e}_{n+1}. \quad (5.30)$$

This calculation can be repeatedly applied, and we get

$$\mathbf{e}_n = \frac{1}{\sqrt{n}} \mathbf{A}^\dagger \mathbf{e}_{n-1} = \frac{\mathbf{A}^\dagger}{\sqrt{n}} \cdot \frac{\mathbf{A}^\dagger}{\sqrt{n-1}} \mathbf{e}_{n-2} = \dots$$

As a final result, we find

$$\mathbf{e}_n = \underbrace{\frac{1}{\sqrt{n!}}}_{\mathbf{e}_0} \underbrace{(\mathbf{A}^\dagger)^n}_{\mathbf{e}_0} \mathbf{e}_0. \quad (5.31)$$

The \mathbf{e}_i are eigenvectors of an Hermitian matrix, therefore they are mutually orthogonal; that is,

$$\mathbf{e}_i^\dagger \mathbf{e}_j = \delta_{ij}.$$

We therefore obtain the elements $A_{i,k}$ ($i, k = 0, 1, 2, 3, \dots$) of the infinitely large matrix \mathbf{A} from

$$A_{i,k} = \mathbf{e}_i^\dagger \mathbf{A} \mathbf{e}_k = (\mathbf{A}^\dagger \mathbf{e}_i)^\dagger \mathbf{e}_k \stackrel{(5.30)}{=} (\sqrt{i+1} \mathbf{e}_{i+1})^\dagger \mathbf{e}_k = \sqrt{i+1} \delta_{i+1,k};$$

that is,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{2} & 0 & \\ 0 & 0 & 0 & \sqrt{3} & \ddots \\ \vdots & & \ddots & \ddots & \end{pmatrix}. \quad (5.32)$$

Accordingly, the elements $A_{i,k}^\dagger$ of the infinitely large matrix \mathbf{A}^\dagger are obtained by evaluating

$$A_{i,k}^\dagger = \mathbf{e}_i^\dagger \mathbf{A}^\dagger \mathbf{e}_k = (\mathbf{A} \mathbf{e}_i)^\dagger \mathbf{e}_k \stackrel{(5.29)}{=} \sqrt{i} \mathbf{e}_{i-1}^\dagger \mathbf{e}_k = \sqrt{i} \delta_{i-1,k},$$

$$\mathbf{A}^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \\ 0 & \sqrt{2} & 0 & 0 & \\ 0 & 0 & \sqrt{3} & 0 & \ddots \\ \vdots & & \ddots & \ddots & \ddots \end{pmatrix}. \quad (5.33)$$

From (5.6) and (5.7) follows²

$$\tilde{\mathbf{X}} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^\dagger) \quad \text{and} \quad \tilde{\mathbf{P}} = \frac{i}{2}(\mathbf{A}^\dagger - \mathbf{A}),$$

and thus finally

$$\mathbf{X} = \sqrt{\frac{2\hbar}{m\omega_0}} \tilde{\mathbf{X}} = \sqrt{\frac{\hbar}{2m\omega_0}} (\mathbf{A} + \mathbf{A}^\dagger), \quad (5.34)$$

$$\mathbf{X} = \sqrt{\frac{\hbar}{2m\omega_0}} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 0 & \sqrt{2} & 0 & 0 & \cdots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 & \cdots \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} & \cdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (5.35)$$

²Note that we get in fact

$$\mathbf{A}\mathbf{A}^\dagger - \mathbf{A}^\dagger\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 2 & 0 & \cdots \\ 0 & 0 & 3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \mathbf{I}.$$

$$\boxed{X^2 = \frac{\hbar}{2m\omega_0} \begin{pmatrix} 1 & 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & 3 & 0 & \sqrt{2 \cdot 3} & 0 & \dots \\ \sqrt{2} & 0 & 5 & 0 & \sqrt{3 \cdot 4} & \dots \\ 0 & \sqrt{2 \cdot 3} & 0 & 7 & 0 & \dots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}} \quad (5.36)$$

For \mathbf{P} , we get similar results:

$$\boxed{\mathbf{P} = \sqrt{2\hbar m\omega_0} \tilde{\mathbf{P}} = \sqrt{\frac{\hbar m\omega_0}{2}} (i\mathbf{A}^\dagger - i\mathbf{A})}$$

$$\boxed{\mathbf{P} = i\sqrt{\frac{\hbar m\omega_0}{2}} \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & \dots \\ 1 & 0 & -\sqrt{2} & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & -\sqrt{3} & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & -\sqrt{4} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}} \quad (5.37)$$

$$\boxed{\mathbf{P}^2 = \frac{\hbar m\omega_0}{2} \begin{pmatrix} 1 & 0 & -\sqrt{2} & 0 & 0 & \dots \\ 0 & 3 & 0 & -\sqrt{2 \cdot 3} & 0 & \dots \\ -\sqrt{2} & 0 & 5 & 0 & -\sqrt{3 \cdot 4} & \dots \\ 0 & -\sqrt{2 \cdot 3} & 0 & 7 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}} \quad (5.38)$$

Note that their form is indeed as predicted in (5.22) and (5.24), which had originally been derived by Heisenberg.

Next, we are looking for the frequency matrix Ω . We have

$$\dot{X}(j, k) = i \cdot \omega(j, k) X(j, k),$$

and we can calculate the elements in the secondary diagonals (on both sides of the main diagonal) from the formula

$$P(j, k) = m \cdot i \cdot \omega(j, k) \cdot X(j, k); \quad (5.39)$$

that is,

$$\omega(j, k) = \frac{P(j, k)}{m \cdot i \cdot X(j, k)}. \quad (5.40)$$

This yields the value $\mp\omega_0$ for the secondary diagonals. In total, the frequency matrix looks like

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & -\omega_0 & 0 & 0 & 0 & \dots \\ \omega_0 & 0 & -\omega_0 & 0 & 0 & \dots \\ 0 & \omega_0 & 0 & -\omega_0 & 0 & \dots \\ 0 & 0 & \omega_0 & 0 & -\omega_0 & \dots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \quad (5.41)$$

According to (5.19), we have

$$E_n = \hbar\omega_0 \left(n + \frac{1}{2} \right) \quad \text{for } n = 0, 1, 2, \dots \quad (5.42)$$

Also, we know

$$\omega(n, m) = 2\pi\nu(n, m) = \frac{1}{\hbar}(E_n - E_m).$$

Together with (5.42), this yields

$$\omega(n, m) = (n - m) \cdot \omega_0 \quad \text{for } |n - m| = 1,$$

that is, a perfect agreement with the values in (5.41)!

5.2 Expectation Values and Variances

Let a harmonic oscillator be in the eigenstate \mathbf{e}_j with the eigenvalue $(j + 1/2)\hbar\omega_0$. If we choose the energy matrix to be diagonal, the eigenvector components must be $\mathbf{e}_j(k) = \delta_{jk}$; that is, the j th component is equal to one and the others are zero. The expectation value of any observable A in the considered eigenstate is then given by

$$\langle A \rangle = \mathbf{e}_j^\top \mathbf{A} \mathbf{e}_j = a_{jj}.$$

For example, the expectation value of the observable X is equal to zero, because in (5.35) we find $a_{jj} = 0$ for all j . Similarly, the mean value of \mathbf{P} is equal to zero. On the other hand, if we calculate the mean value of X^2 and \mathbf{P}^2 , it follows from (5.36) that

$$(X^2)_{jj} = \frac{\hbar}{2m\omega_0}(2j - 1),$$

and from (5.38) that

$$(\mathbf{P}^2)_{jj} = \frac{\hbar m \omega_0}{2}(2j - 1).$$

The mean values are all zero, thus this value also represents the mean square deviation of the position, namely

$$(\Delta x)^2 = \frac{\hbar}{2m\omega_0}(2j - 1).$$

Similarly, we get for the mean square deviation of the momentum

$$(\Delta p)^2 = \frac{\hbar m\omega_0}{2}(2j - 1).$$

It follows that

$$(\Delta x \Delta p)^2 = \hbar^2 \left(j - \frac{1}{2} \right)^2,$$

thus Heisenberg's uncertainty relation is fulfilled:

$$\underline{\underline{\Delta x \Delta p \geq \frac{1}{2} \hbar}}.$$

In the ground state, neither Δx nor Δp is equal to zero. Instead, $\Delta x \Delta p = \frac{1}{2} \hbar > 0$. In this state, the system has the “zero-point energy” $E_0 = \frac{\hbar\omega_0}{2}$, that can never leave the oscillator.

5.3 Problems

- 5.1 Commutation Relation:** Show that the standard commutation relations for \mathbf{P} and \mathbf{X} apply for the harmonic oscillator.
- 5.2 Dimension of A :** What is the physical dimension of $A = \tilde{\mathbf{X}} + i\tilde{\mathbf{P}}$, as defined in (5.6)?
- 5.3 Observables:** Is the matrix $\mathbf{A} = \tilde{\mathbf{X}} + i\tilde{\mathbf{P}}$ as defined in (5.6) an observable?
- 5.4 Hamiltonian:** According to (5.10) we have $\mathbf{H} = \hbar\omega_0(N + \frac{1}{2}\mathbf{I})$ with $N \stackrel{\text{def}}{=} \mathbf{A}^\dagger \mathbf{A}$. Is N Hermitian?
- 5.5 Commutator:** Show that for $A = \tilde{\mathbf{X}} + i\tilde{\mathbf{P}}$, as defined in (5.6), $[A^n, N] = nA^n$ holds.
- 5.6 Form of N :** What is the form of N , and what are the eigenvalues?
- 5.7 Form of X , X^2 , and X^3 :** What do the matrix elements of X , X^2 , and X^3 look like? Calculate them by means of the eigenvalue equations of the matrices \mathbf{A}^\dagger and \mathbf{A} .

Chapter 6

Angular Momentum

Abstract The general procedure for the *one-dimensional* harmonic oscillator is now extended to *three-dimensional* systems. In such systems, the three-dimensional angular momentum comes into play. It has an important role in the treatment of atoms and quantum mechanical problems with rotational symmetry. The generalization to three dimensions reached a first climax in Born's, Heisenberg's, and Jordan's famous “Three men work” (Born et al., ZS f. Physik, 1925, [6]). Green explains in [12]:

Even before the discovery of matrix mechanics, Bohr realized that atomic spectra could only be explained if the angular momentum was limited to certain values, which are integral multiples of Planck's constant \hbar . But a full knowledge of the behaviour of the angular momentum of atomic systems was obtained only when matrix mechanics was developed.

Most of the matrices in this chapter are finite-dimensional $N \times N$ -matrices with $N \in \mathbb{N} = \{0, 1, 2, 3, \dots\}$.

6.1 The Matrix Vector of the Angular Momentum

If a mass point revolves around a fixed axis, it has an angular momentum which is defined as

$$\boldsymbol{\ell} \stackrel{\text{def}}{=} \mathbf{r} \times \mathbf{p}, \quad (6.1)$$

where

$$\mathbf{r} \stackrel{\text{def}}{=} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (6.2)$$

is the distance vector of the mass point from the axis of rotation and

$$\mathbf{p} \stackrel{\text{def}}{=} m \dot{\mathbf{r}} \quad (6.3)$$

is the momentum of the particle. Componentwise, the equation for the angular momentum (6.1) looks like

$$\ell_1 = x_2 p_3 - x_3 p_2,$$

$$\ell_2 = x_3 p_1 - x_1 p_3,$$

$$\ell_3 = x_1 p_2 - x_2 p_1.$$

In quantum mechanics, the components of the distance vector are defined by three matrices X_1 , X_2 , and X_3 , which are summarized in the matrix vector

$$\mathfrak{R} \stackrel{\text{def}}{=} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}. \quad (6.4)$$

Accordingly, the matrix vector of the momentum is defined as

$$\mathfrak{P} = \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} m \dot{X}_1 \\ m \dot{X}_2 \\ m \dot{X}_3 \end{pmatrix}, \quad (6.5)$$

or, with the matrix vector for the velocity

$$\mathfrak{V} \stackrel{\text{def}}{=} \begin{pmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \dot{X}_3 \end{pmatrix}, \quad (6.6)$$

also as

$$\mathfrak{P} = m \mathfrak{V}. \quad (6.7)$$

Remember that for the matrices X_i and P_i the commutation relations according to (3.23) to (3.26) hold.

By defining the vector product for matrix vectors as

$$\mathfrak{A} \times \mathfrak{B} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \times \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} A_2 B_3 - A_3 B_2 \\ A_3 B_1 - A_1 B_3 \\ A_1 B_2 - A_2 B_1 \end{pmatrix},$$

we can show that

$$\mathfrak{A} \times \mathfrak{B} = -\mathfrak{B} \times \mathfrak{A},$$

but only if A_i commutes with B_j for $i \neq j$. Because X_i commutes with P_j for $i \neq j$, we obtain the definition:

Definition (*Matrix vector of angular momentum*)

$$\mathfrak{L} = \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} \stackrel{\text{def}}{=} \mathfrak{R} \times \mathfrak{P} = -\mathfrak{P} \times \mathfrak{R}. \quad (6.8)$$

Note, however, that the components of angular momentum do *not* commute! For example, we have

$$\begin{aligned} [\mathbf{L}_1, \mathbf{L}_2] &= \mathbf{L}_1 \mathbf{L}_2 - \mathbf{L}_2 \mathbf{L}_1 = \\ &= (X_2 P_3 - X_3 P_2)(X_3 P_1 - X_1 P_3) - (X_3 P_1 - X_1 P_3)(X_2 P_3 - X_3 P_2) \\ &= X_2 P_3 X_3 P_1 + X_3 P_2 X_1 P_3 - X_3 P_1 X_2 P_3 - X_1 P_3 X_3 P_2 \\ &= X_1 P_2 (X_3 P_3 - P_3 X_3) + X_2 P_1 (P_3 X_3 - X_3 P_3) \\ &= (X_1 P_2 - X_2 P_1) i \hbar = i \hbar \mathbf{L}_3. \end{aligned}$$

Similarly, we obtain

$$[\mathbf{L}_2, \mathbf{L}_3] = i \hbar \mathbf{L}_1$$

and

$$[\mathbf{L}_3, \mathbf{L}_1] = i \hbar \mathbf{L}_2.$$

As a consequence, the observables that are assigned to the components of the angular momentum \mathfrak{L} can *not* be measured at the same time.

For “normal” vectors, a vector multiplied by itself is always zero:

$$\mathbf{a} \times \mathbf{a} = \mathbf{0}.$$

This no longer applies to the matrix vector product in quantum mechanics. It follows from the above that

$$\boxed{\mathfrak{L} \times \mathfrak{L} = i \hbar \mathfrak{L} \neq \mathbf{0}}. \quad (6.9)$$

For the two mutually attracting masses m_e and m_n of an electron and a nucleus in an atom, the following equations of motion apply.

$$m_e \frac{d^2}{dt^2} \mathbf{x}_e = \mathbf{f}, \quad m_n \frac{d^2}{dt^2} \mathbf{x}_n = -\mathbf{f}.$$

By dividing the equations of motion by the masses and subtracting them, we obtain for the relative distance

$$\mathbf{x} \stackrel{\text{def}}{=} \mathbf{x}_e - \mathbf{x}_n \quad (6.10)$$

an equation of motion that looks like

$$\frac{d^2}{dt^2}(\mathbf{x}_e - \mathbf{x}_n) = \left(\frac{1}{m_e} + \frac{1}{m_n} \right) \mathbf{f} = \frac{1}{m} \mathbf{f},$$

where the *reduced mass* m is defined by

$$m \stackrel{\text{def}}{=} \frac{m_e m_n}{m_e + m_n}.$$

With the relative velocity $\mathbf{v} \stackrel{\text{def}}{=} \dot{\mathbf{x}}$ and the relative momentum

$$\mathbf{p} \stackrel{\text{def}}{=} m\mathbf{v},$$

we finally get

$$m \frac{d^2}{dt^2} \mathbf{x} = \mathbf{f} = m\dot{\mathbf{v}} = \dot{\mathbf{p}}.$$

On the other hand, it follows from (6.10) by differentiation with respect to time that

$$\mathbf{v} = \mathbf{v}_e - \mathbf{v}_n = \frac{1}{m_e} \mathbf{p}_e - \frac{1}{m_n} \mathbf{p}_n = \frac{1}{m} \mathbf{p},$$

which provides an alternative formulation for the relative momentum:

$$\mathbf{p} = m \left(\frac{1}{m_e} \mathbf{p}_e - \frac{1}{m_n} \mathbf{p}_n \right). \quad (6.11)$$

In quantum mechanics, the relative position and the relative momentum of the electron and the nucleus in an atom are represented by the matrices:

$$\mathbf{X}_j = \mathbf{X}_{ej} - \mathbf{X}_{nj}, \quad j = 1, 2, 3.$$

The velocities and momenta of the electron and the nucleus are represented by the matrices \mathbf{V}_{ej} and \mathbf{V}_{nj} as well as

$$\mathbf{V}_{ej} = \frac{1}{m_e} \mathbf{P}_{ej} \quad \text{and} \quad \mathbf{V}_{nj} = \frac{1}{m_n} \mathbf{P}_{nj}.$$

The relative momentum is defined as

$$\mathbf{P}_j = m \left(\frac{1}{m_e} \mathbf{P}_{ej} - \frac{1}{m_n} \mathbf{P}_{nj} \right) = \frac{1}{m_n + m_e} (m_n \mathbf{P}_{ej} - m_e \mathbf{P}_{nj}).$$

The matrices X_{ej} and \mathbf{P}_{ej} for the electron commute with the matrices X_{nj} and \mathbf{P}_{nj} for the nucleus. This leads to the following commutation relations for the relative position matrices and momentum matrices:

$$\begin{aligned} X_j \mathbf{P}_j - \mathbf{P}_j X_j &= \frac{m_n}{m_n + m_e} (X_{ej} \mathbf{P}_{ej} - \mathbf{P}_{ej} X_{ej}) + \frac{m_e}{m_n + m_e} (X_{nj} \mathbf{P}_{nj} - \mathbf{P}_{nj} X_{nj}) = \\ &= \left(\frac{m_n}{m_n + m_e} + \frac{m_e}{m_n + m_e} \right) i\hbar \mathbf{I} = i\hbar \mathbf{I}. \end{aligned}$$

Thus, the relative matrices satisfy the same conditions as in the case of a single mass.

If we define the scalar product of two matrix vectors as

$$(\mathfrak{A} \cdot \mathfrak{B}) \stackrel{\text{def}}{=} A_1 B_1 + A_2 B_2 + A_3 B_3,$$

then we get

$$(\mathcal{L} \cdot \mathcal{L}) \stackrel{\text{def}}{=} L^2 = L_1^2 + L_2^2 + L_3^2.$$

6.2 Eigenvalues and Eigenvectors of L^2 and L_3

6.2.1 Commutativity of L^2 and L_3

First, we show that for the matrix L^2 and all matrices L_i , $i = 1, 2, 3$ the following relation holds,

$$L^2 L_i - L_i L^2 = \mathbf{0};$$

that is, L^2 commutes with all matrices L_i , and therefore L^2 and L_i can be measured simultaneously (which is not the case for the L_i among themselves). We prove this result for the two matrices L^2 and L_3 . With the abbreviation

$$[A, B] \stackrel{\text{def}}{=} AB - BA,$$

we get

$$\begin{aligned} [L_3, L_1^2] &= \underbrace{L_3 L_1 L_1 - L_1 L_3 L_1}_{[L_3, L_1] L_1} + \underbrace{L_1 L_3 L_1 - L_1 L_1 L_3}_{L_1 [L_3, L_1]} = \\ &= i\hbar(L_2 L_1 + L_1 L_2). \end{aligned} \tag{6.12}$$

Similarly, we find that

$$[L_3, L_2^2] = -i\hbar(L_1 L_2 + L_2 L_1). \tag{6.13}$$

Furthermore, it is clear that

$$[\mathbf{L}_3, \mathbf{L}_3^2] = \mathbf{0}. \quad (6.14)$$

Adding (6.12), (6.13), and (6.14) gives the desired result. The general rule for the commutativity is

$$\mathbf{L}_i \mathbf{L}^2 = \mathbf{L}^2 \mathbf{L}_i, \quad i = 1, 2, 3;$$

that is,

$$[\mathbf{L}_i, \mathbf{L}^2] = \mathbf{0}.$$

6.2.2 Eigenvalues and Eigenvectors

We now want to determine the eigenvalues and eigenvectors of the two matrices \mathbf{L}^2 and \mathbf{L}_3 . \mathbf{L}_3 is preferred over \mathbf{L}_1 and \mathbf{L}_2 , inasmuch as we later observe the effect of a magnetic field on atoms, where the field is aligned along the x_3 — or z -direction.

Generally speaking, two commuting matrices \mathbf{A} and \mathbf{B} have the same set of eigenvectors. To see this, let \mathbf{e}_A be an eigenvector of \mathbf{A} to the eigenvalue λ_A ; that is, $\mathbf{A}\mathbf{e}_A = \lambda_A \mathbf{e}_A$. Because \mathbf{A} and \mathbf{B} commute, we have

$$\mathbf{A}\mathbf{B}\mathbf{e}_A = \mathbf{B}\mathbf{A}\mathbf{e}_A = \mathbf{B}(\lambda_A \mathbf{e}_A) = \lambda_A \mathbf{B}\mathbf{e}_A.$$

Hence $\mathbf{B}\mathbf{e}_A$ is an eigenvector of \mathbf{A} with the same eigenvalue λ_A , therefore $\mathbf{B}\mathbf{e}_A$ must be a multiple of the eigenvector \mathbf{e}_A :

$$\mathbf{B}\mathbf{e}_A = \lambda_B \mathbf{e}_A.$$

As claimed, the matrices \mathbf{A} and \mathbf{B} have the same eigenvectors.

The two matrices \mathbf{L}_3 and \mathbf{L}^2 commute, therefore they have the same set of eigenvectors. Let us single out one of them, \mathbf{e} . In addition, we extract the factor \hbar^2 from the eigenvalue, such that λ is a purely numerical value¹

$$\mathbf{L}^2 \mathbf{e} = \hbar^2 \lambda \mathbf{e}, \quad (6.15)$$

and

$$\mathbf{L}_3 \mathbf{e} = \hbar \mu \mathbf{e}. \quad (6.16)$$

Similar to the harmonic oscillator, we can now define the so-called *creation* and *annihilation operators* that allow us to jump from a known eigenvector to the next:

¹Planck's constant \hbar has the dimension *energy · time*. The angular momentum has the dimension *length · mass · length · time*⁻¹. The energy has the dimension *mass · length² · time*⁻². Therefore, \hbar has the dimension *mass · length² · time*⁻¹, that is, the same dimension as the angular momentum. The squared angular momentum then has the same dimension as \hbar^2 .

$$\mathbf{L}_+ \stackrel{\text{def}}{=} \mathbf{L}_1 + i \mathbf{L}_2, \quad (6.17)$$

and

$$\mathbf{L}_- \stackrel{\text{def}}{=} \mathbf{L}_1 - i \mathbf{L}_2. \quad (6.18)$$

The matrices \mathbf{L}_\pm commute with the matrix \mathbf{L}^2 , because

$$\begin{aligned} \mathbf{L}^2 \mathbf{L}_\pm - \mathbf{L}_\pm \mathbf{L}^2 &= \mathbf{L}^2 \mathbf{L}_1 \pm i \mathbf{L}^2 \mathbf{L}_2 - \mathbf{L}_1 \mathbf{L}^2 \mp i \mathbf{L}_2 \mathbf{L}^2 = \\ &= \underbrace{[\mathbf{L}^2, \mathbf{L}_1]}_0 + i \underbrace{[\mathbf{L}^2, \mathbf{L}_2]}_0 = \mathbf{0}. \end{aligned}$$

On the other hand, we have

$$[\mathbf{L}_3, \mathbf{L}_\pm] = \pm \hbar \mathbf{L}_\pm, \quad (6.19)$$

because

$$\begin{aligned} [\mathbf{L}_3, \mathbf{L}_\pm] &= [\mathbf{L}_3, \mathbf{L}_1] \pm i [\mathbf{L}_3, \mathbf{L}_2] = i \hbar \mathbf{L}_2 \mp i^2 \hbar \mathbf{L}_1 = \\ &= \hbar (\pm \mathbf{L}_1 + i \mathbf{L}_2) = \pm \hbar \mathbf{L}_\pm. \end{aligned}$$

It also follows that

$$[\mathbf{L}_+, \mathbf{L}_-] = 2\hbar \mathbf{L}_3. \quad (6.20)$$

Solving (6.20) for $\mathbf{L}_+ \mathbf{L}_-$ yields

$$\mathbf{L}_+ \mathbf{L}_- = \mathbf{L}^2 - \mathbf{L}_3 (\mathbf{L}_3 - \hbar \mathbf{I}) \quad (6.21)$$

and

$$\mathbf{L}_- \mathbf{L}_+ = \mathbf{L}^2 - \mathbf{L}_3 (\mathbf{L}_3 + \hbar \mathbf{I}). \quad (6.22)$$

6.2.3 Maximum and Minimum Eigenvalues

Instead of (6.15), we can also write

$$(\mathbf{L}_1^2 + \mathbf{L}_2^2) \mathbf{e} + \hbar^2 \mu^2 \mathbf{e} = \hbar^2 \lambda \mathbf{e}.$$

If we multiply this with \mathbf{e}^\top from the left, we get for the normalized eigenvectors (i.e., $|\mathbf{e}| = 1$)

$$\mathbf{e}^\top (\mathbf{L}_1^2 + \mathbf{L}_2^2) \mathbf{e} = \hbar^2 (\lambda - \mu^2). \quad (6.23)$$

Because the matrices \mathbf{L}_1 and \mathbf{L}_2 are Hermitian matrices, the eigenvalues are all real. Therefore, the left side of (6.23) must be positive: $(\lambda - \mu^2) \geq 0$, or $\lambda \geq \mu^2$. Let us now multiply (6.15) with the matrix \mathbf{L}_\pm from the left. Because \mathbf{L}_\pm commutes with \mathbf{L}^2 , we obtain

$$\mathbf{L}_\pm \mathbf{L}^2 \mathbf{e} = \underline{\mathbf{L}^2 \mathbf{L}_\pm \mathbf{e}} = \hbar^2 \lambda \underline{\mathbf{L}_\pm \mathbf{e}}; \quad (6.24)$$

that is, $\mathbf{L}_\pm \mathbf{e}$ is also an eigenvector of \mathbf{L}^2 with the same eigenvalue $\hbar^2 \lambda$. Next, we obtain with (10.21) for $\mathbf{L}_3 \mathbf{L}_\pm \mathbf{e}$

$$\begin{aligned} \underline{\underline{\mathbf{L}_3 \mathbf{L}_\pm \mathbf{e}}} &= ([\mathbf{L}_3, \mathbf{L}_\pm] + \mathbf{L}_\pm \mathbf{L}_3) \mathbf{e} = \\ &= (\pm \hbar \mathbf{L}_\pm + \mathbf{L}_\pm \mathbf{L}_3) \mathbf{e} = \underline{\hbar(1 \pm \mu) \mathbf{L}_\pm \mathbf{e}}. \end{aligned} \quad (6.25)$$

In other words, $\mathbf{L}_\pm \mathbf{e}$ is also an eigenvector of \mathbf{L}_3 . However, the associated eigenvalue changes to $\hbar(1 \pm \mu)$. As in the case of the harmonic oscillator, a multiple application of \mathbf{L}_+ to the eigenvalue equation subsequently increases the eigenvalue of \mathbf{L}_3 . Of course, this has to stop at some point, inasmuch as otherwise the condition $\lambda \geq \mu^2$ would be violated. Let \bar{m} be the maximum value of λ that does not violate $\lambda \geq \mu^2$. With the definition $\mathbf{e}_m \stackrel{\text{def}}{=} \mathbf{L}_+^m \mathbf{e}$, we then get

$$\mathbf{L}_3 \mathbf{e}_m = \hbar \bar{m} \mathbf{e}_m. \quad (6.26)$$

Because \mathbf{L}_+ and \mathbf{L}_3 commute, it follows that

$$\mathbf{L}_3 (\mathbf{L}_+ \mathbf{e}_m) = \hbar(1 + \bar{m}) (\mathbf{L}_+ \mathbf{e}_m).$$

But an eigenvalue $\hbar(1 + \bar{m})$ is not possible, because \bar{m} was by definition the largest value of μ ! Therefore $(\mathbf{L}_+ \mathbf{e}_m) = \mathbf{0}$ must apply. Under consideration of

$$\mathbf{L}_- \mathbf{L}_+ = \mathbf{L}^2 - \mathbf{L}_3^2 - \hbar \mathbf{L}_3,$$

this equation multiplied by the matrix \mathbf{L}_- from the left finally gives

$$\mathbf{L}_- (\mathbf{L}_+ \mathbf{e}_m) = \mathbf{0},$$

or

$$(\mathbf{L}^2 - \mathbf{L}_3^2 - \hbar \mathbf{L}_3) \mathbf{e}_m = \mathbf{0}.$$

Therefore, we get

$$\hbar^2 (\lambda - \bar{m}^2 - \bar{m}) \mathbf{e}_m = \mathbf{0}.$$

Inasmuch as the eigenvector \mathbf{e}_m cannot be a zero vector, the value in brackets must be zero; that is,

$$\lambda = \bar{m}(\bar{m} + 1). \quad (6.27)$$

If we now multiply the second eigenvalue equation (6.26) with the matrix \mathbf{L}_- from the left, we obtain

$$\mathbf{L}_- \mathbf{L}_3 \mathbf{e}_m = \hbar \bar{m} \mathbf{L}_- \mathbf{e}_m,$$

or with (6.25)

$$\mathbf{L}_3 (\mathbf{L}_- \mathbf{e}_m) = \hbar(\bar{m} - 1) (\mathbf{L}_- \mathbf{e}_m).$$

For the largest eigenvalue \bar{m} , we get

$$\mathbf{L}^2 \mathbf{e}_m = \hbar^2 \lambda \mathbf{e}_m,$$

or

$$\mathbf{L}^2 (\mathbf{L}_- \mathbf{e}_m) = \hbar^2 \lambda (\mathbf{L}_- \mathbf{e}_m).$$

In other words, $\mathbf{L}_- \mathbf{e}_m$ is an eigenvector of \mathbf{L}^2 and \mathbf{L}_3 with the eigenvalues $\hbar^2 \lambda$ and $\hbar(\bar{m} - 1)$, respectively. Similarly, n -times multiplication by \mathbf{L}_- yields

$$\mathbf{L}_3 (\mathbf{L}_-^n \mathbf{e}_m) = \hbar(\bar{m} - n) (\mathbf{L}_-^n \mathbf{e}_m).$$

However, there must be a lowest value for m , because otherwise $\lambda \geq m^2$ would be violated. The minimum value of m without violation of $\lambda \geq m^2$ is $\hbar(\bar{m} - n)$. Then

$$\mathbf{L}_3 (\mathbf{L}_-^{n+1} \mathbf{e}_m) = \hbar(\bar{m} - (n + 1)) (\mathbf{L}_-^{n+1} \mathbf{e}_m) = \mathbf{0}.$$

Because $\hbar(\bar{m} - (n + 1)) \neq 0$, $(\mathbf{L}_-^{n+1} \mathbf{e}_m) = \mathbf{0}$ must hold; that is, $(\mathbf{L}_- \mathbf{L}_-^n \mathbf{e}_m) = \mathbf{0}$ and therefore also $(\mathbf{L}_+ \mathbf{L}_- \mathbf{L}_-^n \mathbf{e}_m) = \mathbf{0}$. Replacing \mathbf{L}_+ by \mathbf{L}_- , we obtain

$$(\mathbf{L}^2 - \mathbf{L}_3^2 + \hbar \mathbf{L}_3) \mathbf{L}_-^n \mathbf{e}_m = \mathbf{0},$$

or

$$\hbar^2 (\lambda - (\bar{m} - n)^2 + (\bar{m} - n)) \mathbf{L}_-^n \mathbf{e}_m = \mathbf{0}.$$

Because $\mathbf{L}_-^n \mathbf{e}_m \neq \mathbf{0}$, we get the condition

$$\lambda - (\bar{m} - n)^2 + (\bar{m} - n) = 0.$$

With the value of $\lambda = \bar{m}(\bar{m} + 1)$ according to (6.27), we get

$$\bar{m}(\bar{m} + 1) - (\bar{m} - n)^2 + (\bar{m} - n) = (n + 1)(2\bar{m} - n) = 0.$$

Due to $(n + 1) \neq 0$, $2\bar{m} - n = 0$ (i.e., $\bar{m} = n/2$) must hold, where n is the number of steps from the maximum eigenvalue $\hbar^2 \bar{m}$ to the minimal eigenvalue $\hbar^2(\bar{m} - n)$. This number n is always an integer (including zero). If we set ℓ for $n/2$, $\hbar\ell$ and $-\hbar\ell$ are the

maximum and minimum eigenvalues of \mathbf{L}_3 . For a given value of ℓ , the eigenvalues of \mathbf{L}_3 are therefore equal to $\hbar m$, where $m = \ell, \ell - 1, \dots, -\ell$. In particular, there are $2\ell + 1$ eigenvalues of \mathbf{L}_3 . Because n is always an integer, the possible values of ℓ are $0, 1/2, 1, 3/2, \dots$. Note that these fractional quantum numbers automatically occurred in our treatment of angular momentum.

If we denote $e_{\ell m}$ the common eigenvector of \mathbf{L}^2 and \mathbf{L}_3 with eigenvalues $\hbar^2 \ell(\ell + 1)$ and $\hbar m$, respectively, we finally obtain the following theorem.

Theorem *For the matrices \mathbf{L}^2 and \mathbf{L}_3 , the eigenvalue equations are*

$$\mathbf{L}^2 e_{\ell m} = \hbar^2 \ell(\ell + 1) e_{\ell m} \quad (6.28)$$

and

$$\mathbf{L}_3 e_{\ell m} = \hbar m e_{\ell m}, \quad (6.29)$$

where the allowed quantum numbers are the Angular momentum quantum number,

$$\ell = 0, 1/2, 1, 3/2, \dots,$$

and the Magnetic quantum number,

$$m = -\ell, -\ell + 1, \dots, \ell - 1, \ell.$$

For a given ℓ , the matrix \mathbf{L}_3 has a spectrum with $\ell + 1$ eigenvalues

$$-\hbar \ell, \hbar(-\ell + 1), \dots, \hbar(\ell - 1), \hbar \ell.$$

The matrix is a square $(2\ell + 1) \times (2\ell + 1)$ -matrix. Therefore in the Heisenberg matrix mechanics also *finitely large* matrices occur in a Hilbert space!

6.2.4 Orientation of the Angular Momentum Vectors

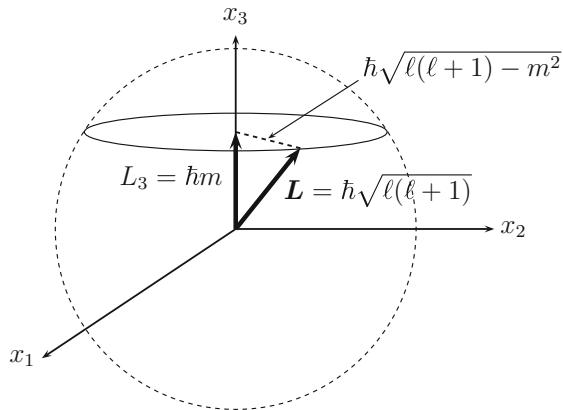
From the above theorem, we obtain for $|\mathbf{L}|$ and the z-component of angular momentum \mathbf{L}

$$|\mathbf{L}| = \hbar \sqrt{\ell(\ell + 1)}, \quad (6.30)$$

$$L_3 = \hbar m. \quad (6.31)$$

Because it is not possible in quantum mechanics to determine the components (L_1, L_2) in addition to the length $|\mathbf{L}|$ and the z-component L_3 at the same time,

Fig. 6.1 The angular momentum \mathbf{L} has a component along the x_3 axis with the allowed values $L_3 = \hbar m$, $-\ell \leq m \leq \ell$, and undetermined components L_1 and L_2 in the x_1 - x_2 plane



the angular momentum vector \mathbf{L} can never be exactly parallel to the z -axis of the coordinate system. In such a case, the x - and the y -components would be exactly zero and thus accurately determined. The only statement that can be made about the x - and y -components is that they make up a circular path together, because

$$L_1^2 + L_2^2 = |\mathbf{L}|^2 - L_3^2 = \hbar^2[\ell(\ell+1) - m^2], \quad (6.32)$$

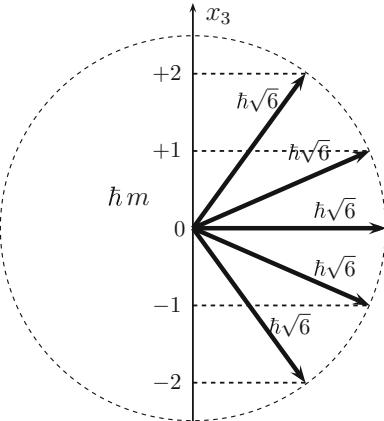
where the right side of this equation is constant for given values of the quantum numbers ℓ and m . The circle has the radius $\hbar\sqrt{\ell(\ell+1) - m^2}$ (see Fig. 6.1).

Our system with the angular momentum $|\mathbf{L}| = \hbar\sqrt{\ell(\ell+1)}$ has a well-defined component along the axis x_3 , with allowed values $L_3 = \hbar m$, $(-\ell \leq m \leq \ell)$, and it has indefinite components L_1 and L_2 in the x_1 - x_2 plane. The angular momentum is therefore quantized, and ℓ is called the *angular momentum quantum number*. The quantum number m specifies the x_3 -component of the angular momentum. Because m is bounded by the value $2\ell + 1$, the x_3 -component of the angular momentum is limited to $2\ell + 1$ discrete values for a given ℓ . This limitation of the angular momentum is also called *space quantization*. This name results from the vector representation of the angular momentum, where the angular momentum is defined by a vector of length $\hbar\sqrt{\ell(\ell+1)}$ and a direction that is determined by its x_3 -component of length $\hbar m$; see Fig. 6.2. It is not possible to determine the x_1 - and x_2 -components of the angular momentum from the quantum numbers ℓ and m .

6.2.5 The Matrices L^2 and L_3

We can use the eigenvectors as a complete set of orthonormal vectors for constructing a matrix that represents the angular momentum. Because the two matrices L^2 and L_3 commute, they are both *diagonal matrices* in this representation. In contrast, the

Fig. 6.2 The five (i.e., $2\ell + 1$) permitted directions of the angular momentum for $\ell = 2$. The length of the vector is $\hbar\sqrt{\ell(\ell + 1)} = \hbar\sqrt{6}$



two matrices L_1 and L_2 do not commute with the matrix L_3 , therefore they will not be diagonal. If we multiply the eigenvalue equations as specified in the above theorem with the transposed normalized eigenvector $e_{\ell m}^\top$ from the left, we obtain for the diagonal elements

$$e_{\ell m}^\top L^2 e_{\ell m} = L_{mm}^2 = \hbar^2 \ell(\ell + 1) \quad (6.33)$$

and

$$e_{\ell m}^\top L_3 e_{\ell m} = L_{3,mm} = \hbar m. \quad (6.34)$$

We now characterize matrices for the value ℓ by a superscript number, for example, $L_3^{(\ell)}$. The matrices $L^{2(\ell)}$ then have the simple diagonal form

$$\underline{\underline{L^{2(\ell)}}} = \hbar^2 \ell(\ell + 1) I. \quad (6.35)$$

For $\ell = \frac{1}{2}$, for example, we obtain the diagonal matrices

$$L^{2(1/2)} = \hbar^2 \begin{pmatrix} 3/4 & 0 \\ 0 & 3/4 \end{pmatrix}, \quad (6.36)$$

$$L_3^{(1/2)} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (6.37)$$

For $\ell = 1$, the diagonal matrices are

$$L^{2(1)} = \hbar^2 \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad L_3^{(1)} = \hbar \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

6.2.6 The Matrices \mathbf{L}_+ , \mathbf{L}_- , \mathbf{L}_1 and \mathbf{L}_2

We know

$$\mathbf{L}_3 \mathbf{L}_+ \mathbf{e}_{\ell m} = \hbar(m+1) \mathbf{L}_+ \mathbf{e}_{\ell m}. \quad (6.38)$$

For the eigenvalue $\hbar(m+1)$, this eigenvalue equation can also be written as

$$\mathbf{L}_3 \mathbf{e}_{\ell, m+1} = \hbar(m+1) \mathbf{e}_{\ell, m+1}. \quad (6.39)$$

Because the eigenvalues of \mathbf{L}_3 in the (6.38) and (6.39) are the same, the eigenvectors can only differ by a multiplicative factor α_m :

$$\mathbf{L}_+ \mathbf{e}_{\ell, m} = \alpha_m \mathbf{e}_{\ell, m+1}. \quad (6.40)$$

Similarly, we obtain

$$\mathbf{L}_- \mathbf{e}_{\ell m} = \beta_m \mathbf{e}_{\ell, m-1}, \quad (6.41)$$

where

$$\alpha_m = \mathbf{e}_{\ell, m+1}^\top \mathbf{L}_+ \mathbf{e}_{\ell m} \quad (6.42)$$

or

$$\alpha_m^* = \mathbf{e}_{\ell, m}^\top \mathbf{L}_- \mathbf{e}_{\ell, m+1}, \quad (6.43)$$

and

$$\beta_m = \mathbf{e}_{\ell, m-1}^\top \mathbf{L}_- \mathbf{e}_{\ell m} \quad (6.44)$$

or

$$\beta_{m+1} = \mathbf{e}_{\ell, m}^\top \mathbf{L}_- \mathbf{e}_{\ell, m+1}. \quad (6.45)$$

A comparison of (6.43) and (6.45) provides

$$\alpha_m^* = \beta_{m+1}. \quad (6.46)$$

Multiplication of (6.40) by the matrix \mathbf{L}_- from the left yields

$$\mathbf{L}_- \mathbf{L}_+ \mathbf{e}_{\ell m} = \alpha_m \mathbf{L}_- \mathbf{e}_{\ell, m+1}. \quad (6.47)$$

Replacing $\mathbf{L}_- \mathbf{L}_+$ by $\mathbf{L}^2 - \mathbf{L}_3^2 - \hbar \mathbf{L}_3$ and using (6.41) leads to

$$(\mathbf{L}^2 - \mathbf{L}_3^2 - \hbar \mathbf{L}_3) \mathbf{e}_{\ell m} = \alpha_m \beta_{m+1} \mathbf{e}_{\ell m}$$

or

$$(\ell(\ell+1) - m^2 - m) \hbar^2 \mathbf{e}_{\ell m} = |\alpha_m|^2 \mathbf{e}_{\ell m}.$$

This implies

$$\alpha_m = [\ell(\ell + 1) - m(m + 1)]^{1/2} \hbar. \quad (6.48)$$

With this α_m , we obtain

$$\mathbf{L}_+ \mathbf{e}_{\ell m} = [\ell(\ell + 1) - m(m + 1)]^{1/2} \hbar \mathbf{e}_{\ell, m+1}, \quad (6.49)$$

or

$$\mathbf{e}_{\ell' m'}^\top \mathbf{L}_+ \mathbf{e}_{\ell m} = [\ell(\ell + 1) - m(m + 1)]^{1/2} \hbar \delta_{\ell \ell'} \delta_{m', m+1}. \quad (6.50)$$

Similarly, we obtain

$$\mathbf{e}_{\ell' m'}^\top \mathbf{L}_- \mathbf{e}_{\ell m} = [\ell(\ell + 1) - m(m - 1)]^{1/2} \hbar \delta_{\ell \ell'} \delta_{m', m-1}. \quad (6.51)$$

The equations (6.50) and (6.51) provide all the elements of the matrices \mathbf{L}_+ and \mathbf{L}_- . The Kronecker-Delta functions indicate that all nonvanishing matrix elements occur in blocks along the diagonal $\ell' = \ell$. The block matrices belonging to the values $\ell = 0, \frac{1}{2}$, and 1 are shown below. The matrices \mathbf{L}_1 and \mathbf{L}_2 can now be derived from the following relations.

$$\mathbf{L}_1 = \frac{1}{2} (\mathbf{L}_+ + \mathbf{L}_-) \quad \text{and} \quad \mathbf{L}_2 = \frac{1}{2i} (\mathbf{L}_+ - \mathbf{L}_-). \quad (6.52)$$

For $\ell = 0$, we get

$$\mathbf{L}_+ = \mathbf{L}_- = \mathbf{L}_1 = \mathbf{L}_2 = \mathbf{0}. \quad (6.53)$$

For $\ell = 1/2$, we get

$$\begin{aligned} \mathbf{L}_+^{(1/2)} &= \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{L}_-^{(1/2)} = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \\ \mathbf{L}_1^{(1/2)} &= \frac{1}{2} \hbar \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{L}_2^{(1/2)} = \frac{1}{2} \hbar \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \end{aligned} \quad (6.54)$$

Finally, for $j = 1$ we have

$$\begin{aligned} \mathbf{L}_+^{(1)} &= \hbar \begin{pmatrix} 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{L}_-^{(1)} = \hbar \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix}, \\ \mathbf{L}_1^{(1)} &= \frac{1}{\sqrt{2}} \hbar \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{L}_2^{(1)} = \frac{1}{\sqrt{2}} \hbar \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \end{aligned} \quad (6.55)$$

Note that all matrices \mathbf{L}_1 , \mathbf{L}_2 , and \mathbf{L}_3 are Hermitian matrices; that is, $\mathbf{L}_i^\dagger = \bar{\mathbf{L}}_i^\top = \mathbf{L}_i$ for all i . Also,

$$\begin{aligned}
 & (\mathbf{L}_1^{(\ell)})^2 + (\mathbf{L}_2^{(\ell)})^2 + (\mathbf{L}_3^{(\ell)})^2 = \\
 &= \frac{1}{2} \hbar^2 \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} + \frac{1}{2} \hbar^2 \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix} + \hbar^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \\
 &= \hbar^2 \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} = \mathbf{L}^{2(\ell)}.
 \end{aligned}$$

A given ℓ -value defines a $(2\ell + 1) \times (2\ell + 1)$ matrix. This dimension reflects that the space is spanned by the $2\ell + 1$ different m -values. If we allow all possible ℓ -values, we obtain an infinitely large matrix consisting of $(2\ell + 1) \times (2\ell + 1)$ -blocks along the main diagonal. Their general form can be written as a direct sum ($\alpha = 1, 2, 3, +$ or $-$):

$$L_\alpha = \bigoplus_{\ell=0}^{\infty} L_\alpha^{(\ell)} = \begin{pmatrix} L_\alpha^{(0)} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & L_\alpha^{(1)} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & L_\alpha^{(2)} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

For integer values of ℓ , we obtain matrices such as

and

With these matrices, we obtain

6.3 Problems

6.1 Commutation Relation: What is $\mathfrak{X} \cdot \mathfrak{P} - \mathfrak{P} \cdot \mathfrak{X}$ for the harmonic oscillator?

6.2 Matrix Vectors: Prove the identity

$$(\mathfrak{A} \times \mathfrak{B}) \cdot \mathfrak{C} = \mathfrak{A} \cdot (\mathfrak{B} \times \mathfrak{C}).$$

6.3 Matrices L_+ and L_- : Are L_+ and L_- Hermitian?

6.4 Eigenvector for L_+ and L_- : Let $e(j, m)$ be normalized, common eigenvectors of L^2 and L_3 . Show that

$$L_+ e(j, m) = \hbar \sqrt{(j-m)(j+m+1)} e(j, m+1),$$

and

$$L_- e(j, m) = \hbar \sqrt{(j+m)(j-m+1)} e(j, m-1).$$

Chapter 7

Wolfgang Pauli and the Hydrogen Atom

Abstract It was up to the physicist Wolfgang Pauli to apply the new matrix quantum mechanics to the hydrogen atom successfully.

7.1 Basic Matrices and Matrix Vectors

In his work, Pauli [18] first introduced some matrices that we already know from the angular momentum. This is not surprising, however, because the hydrogen atom essentially consists of a nucleus and one electron orbiting around it. At least, that was the general image in the older atomic theory.

The Cartesian coordinates are again represented by the matrices X_1 , X_2 , X_3 and combined into the matrix vector

$$\mathfrak{R} \stackrel{\text{def}}{=} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix},$$

which satisfies the relation

$$R^2 = (\mathfrak{R} \cdot \mathfrak{R}) = X_1^2 + X_2^2 + X_3^2. \quad (7.1)$$

Next, Pauli defined the momentum matrices $P_1 \stackrel{\text{def}}{=} m\dot{X}_1$, $P_2 \stackrel{\text{def}}{=} m\dot{X}_2$, and $P_3 \stackrel{\text{def}}{=} m\dot{X}_3$, which are summarized in the matrix vector

$$\mathfrak{P} \stackrel{\text{def}}{=} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = m \frac{d}{dt} \mathfrak{R}.$$

Also, he derived the relations¹ from equations (3.23) to (3.26)

$$\mathbf{P}_\rho \mathbf{P}_\sigma - \mathbf{P}_\sigma \mathbf{P}_\rho = \mathbf{0}, \quad \mathbf{X}_\rho \mathbf{X}_\sigma - \mathbf{X}_\sigma \mathbf{X}_\rho = \mathbf{0}, \quad (7.2)$$

$$\mathbf{P}_\rho \mathbf{X}_\sigma - \mathbf{X}_\sigma \mathbf{P}_\rho = \begin{cases} \mathbf{0} & \text{for } \rho \neq \sigma \\ \frac{\hbar}{2\pi i} \mathbf{I} & \text{for } \rho = \sigma. \end{cases} \quad (7.3)$$

He further assumed that for an arbitrary function \mathbf{F} of $\mathbf{R}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, this relation is valid for all i (see (3.19) and (3.27)):

$$\mathbf{P}_i \mathbf{F} - \mathbf{F} \mathbf{P}_i = \frac{\hbar}{2\pi i} \frac{\partial \mathbf{F}}{\partial \mathbf{X}_i} \quad (7.4)$$

In particular, it is true for $\mathbf{F} = \mathbf{R}$:

$$\mathbf{P}_i \mathbf{R} - \mathbf{R} \mathbf{P}_i = \frac{\hbar}{2\pi i} \mathbf{X}_i \mathbf{R}^{-1}. \quad (7.5)$$

With these relations, together with the energy equation for a single particle,

$$E = \frac{1}{2m} \mathfrak{P}^2 + \mathbf{F}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3), \quad (7.6)$$

where E is a diagonal matrix, as well as the Heisenberg equation for any quantity Φ ,

$$\frac{\hbar}{2\pi i} \dot{\Phi} = \mathbf{E} \Phi - \Phi \mathbf{E},$$

he obtained the equations of motion for all i as

$$\frac{d\mathbf{P}_i}{dt} = -\frac{\partial E}{\partial \mathbf{X}_i}. \quad (7.7)$$

¹Pauli summarizes these relationships as

$$\mathbf{R} \mathfrak{R} = \mathfrak{R} \mathbf{R}$$

and

$$\mathfrak{P} \mathbf{R} - \mathbf{R} \mathfrak{P} = \frac{\hbar}{2\pi i} \mathfrak{R}/\mathbf{R},$$

which of course is not formally correct for both row-vector matrices and for column-vector matrices. Rather, one must write, for example, for column-vector matrices \mathfrak{R} and \mathfrak{P} :

$$(\mathbf{I}_3 \otimes \mathbf{R}) \mathfrak{R} = \mathfrak{R} \mathbf{R}$$

and

$$\mathfrak{P} \mathbf{R} - (\mathbf{I}_3 \otimes \mathbf{R}) \mathfrak{P} = \frac{\hbar}{2\pi i} \mathfrak{R} \mathbf{R}^{-1},$$

where \otimes is the *Kronecker product*.

Hereby, Pauli postulated the existence of a matrix \mathbf{R} that satisfies the relations (7.1) and (7.5). Next, he introduced the vector matrix \mathfrak{L} corresponding to the angular momentum of the particle with respect to the origin. Similar to (6.8), with $m\mathfrak{V}$ instead of \mathfrak{P} , we get the following definition.

Definition The angular momentum of a particle is

$$\mathfrak{L} \stackrel{\text{def}}{=} m(\mathfrak{R} \times \mathfrak{V}) = -m(\mathfrak{V} \times \mathfrak{R}). \quad (7.8)$$

7.2 Introduction of the Matrix Vector \mathfrak{A}

For the further calculations, we need the so-called *Laplace–Runge–Lenz vector*. This vector was introduced by Lenz² into quantum theory. An atom contains an electron with mass m and charge e . The electron is attracted by the fixed nucleus with the charge $+Ze$ due to the Coulomb force. The matrix \mathbf{E} is set as

$$\mathbf{E} = \frac{1}{2m} \mathfrak{P}^2 - Ze^2 \mathbf{R}^{-1}. \quad (7.9)$$

Note that \mathbf{E} is supposed to be diagonal. The equations of motion corresponding to (7.7) are

$$\dot{\mathfrak{P}} = m\ddot{\mathfrak{R}} = -Ze^2 \mathfrak{R} \mathbf{R}^{-1}. \quad (7.10)$$

We can now define the Lenz-Matrix-Vector.

Definition The Lenz-Matrix-Vector is given by

$$\mathfrak{A} \stackrel{\text{def}}{=} \frac{1}{Ze^2 m} \frac{1}{2} (\mathfrak{L} \times \mathfrak{P} - \mathfrak{P} \times \mathfrak{L}) + \mathfrak{R} \mathbf{R}^{-1}. \quad (7.11)$$

Note that the symmetrized vector matrix difference $\frac{1}{2}(\mathfrak{L} \times \mathfrak{P} - \mathfrak{P} \times \mathfrak{L})$ was used for the vector product $\ell \times p$ in the Lenz-Vector, in order for the vector matrix components A_i to be anti-symmetrical.

In a somewhat lengthy, but trivial calculation (see Problem 7.2), Pauli also showed that

²See Appendix D.

$$\frac{d}{dt}(\mathfrak{R}\mathbf{R}^{-1}) = \frac{1}{2m} [\mathfrak{L} \times (\mathfrak{R}\mathbf{R}^{-3}) - (\mathfrak{R}\mathbf{R}^{-3}) \times \mathfrak{L}]. \quad (7.12)$$

Using this relation, one can show (see Problem 7.3) that the time derivative of \mathfrak{A} is equal to zero. In other words, \mathfrak{A} is a constant matrix vector over time.

Summing up, we now have three time-constant matrix vectors or matrices \mathfrak{A} , \mathfrak{L} , and \mathbf{E} that describe the system (i.e., the atom) completely. We do not need the coordinate matrices X_i and \mathbf{R} anymore. Let us summarize the governing equations for \mathfrak{A} , \mathfrak{L} , and \mathbf{E} .

$$\mathfrak{L} \times \mathfrak{L} = i\hbar\mathfrak{L}, \quad (7.13)$$

$$[A_i, \mathbf{L}_i] = \mathbf{0}, \quad (7.14)$$

$$[A_i, \mathbf{L}_j] = i\hbar\varepsilon_{ijk}A_k, \text{ for } i \neq j \neq k, \quad (7.15)$$

$$\mathfrak{A} \cdot \mathfrak{L} = \mathfrak{L} \cdot \mathfrak{A} = \mathbf{0}, \quad (7.16)$$

$$\mathfrak{A} \times \mathfrak{A} = \frac{\hbar}{2\pi i} \frac{2}{mZ^2e^4} \mathfrak{L}\mathbf{E}, \quad (7.17)$$

$$\mathfrak{A}^2 = \frac{2}{mZ^2e^4} \mathbf{E} \left(\mathfrak{L}^2 + \frac{\hbar^2}{4\pi^2} \mathbf{I} \right) + \mathbf{I}. \quad (7.18)$$

Equation (7.13) is identical to equation (6.9). Equations (7.14) and (7.15) are analogous to (7.2) and (7.3) in form. Equation (7.18) is analogous to the classical equation (D.5). These equations are proofed in Appendix H.

Because the matrices \mathbf{L}_i and A_i commute with the matrix \mathbf{E} , the quantities that are represented by them can be measured at a given energy value. Once the atom has a specific energy, the variables that correspond to the matrices \mathbf{L}_i and A_i describe the simplified system completely. Inasmuch as the matrices X_i and \mathbf{P}_i do not commute with \mathbf{E} , they cannot be measured for a given energy value.

Let ϵ be this fixed energy value; that is, $\mathbf{E} = \epsilon \cdot \mathbf{I}$. We introduce yet another matrix vector

$$\mathfrak{K} \stackrel{\text{def}}{=} \sqrt{-\frac{mZ^2e^4}{2\epsilon}} \mathfrak{A}.$$

Because of (7.16), we certainly also have

$$\mathfrak{L} \cdot \mathfrak{K} = \mathbf{0}.$$

Furthermore, (7.18) implies

$$\mathfrak{K}^2 = \left(\mathfrak{L}^2 + \frac{\hbar^2}{\pi^2} \mathbf{I} \right) + \frac{mZ^2e^4}{2\epsilon} \mathbf{I}.$$

Relation (7.17) now reads

$$\mathfrak{K} \times \mathfrak{K} = \frac{i\hbar}{2\pi} \mathfrak{L}.$$

We further define the matrix vectors

$$\mathfrak{M} \stackrel{\text{def}}{=} \frac{1}{2\hbar} (\mathfrak{L} + \mathfrak{K})$$

and

$$\mathfrak{N} \stackrel{\text{def}}{=} \frac{1}{2\hbar} (\mathfrak{L} - \mathfrak{K}).$$

Based on the relations (7.13) to (7.18), it is easy to see that the following relations hold.

$$\mathfrak{M}^2 - \mathfrak{N}^2 = \frac{1}{\hbar^2} (\mathfrak{L} \cdot \mathfrak{K}) = \mathbf{0},$$

$$2(\mathfrak{M}^2 + \mathfrak{N}^2) = \frac{1}{\hbar^2} (\mathfrak{L}^2 + \mathfrak{K}^2) = - \left(1 + \frac{mZ^2e^4}{2\hbar^2\epsilon} \right) \mathbf{I},$$

$$\mathfrak{M} \times \mathfrak{M} = i\mathfrak{M},$$

and

$$\mathfrak{N} \times \mathfrak{N} = i\mathfrak{N}.$$

Therefore, we get

$$\mathfrak{M}^2 = \mathfrak{N}^2$$

and

$$4\mathfrak{M}^2 = - \left(1 + \frac{mZ^2e^4}{2\hbar^2\epsilon} \right) \mathbf{I}. \quad (7.19)$$

For negative values of ϵ (as assumed by Niels Bohr in his atom model), the square root $\sqrt{-\frac{mZ^2e^4}{2\epsilon}}$ is real. In this case, the matrices \mathbf{K}_i , \mathbf{M}_i and N_i represent real entities. The commutation relations for the \mathbf{M}_i are the same as for the matrices \mathbf{L}_i representing the angular momentum (except a factor \hbar). For this reason, the matrix

$$\mathfrak{M}^2 = \mathbf{M}_1^2 + \mathbf{M}_2^2 + \mathbf{M}_3^2$$

can only have the eigenvalues

$$\ell(\ell + 1), \ell = 1, \frac{3}{2}, 2, \dots$$

From (7.19), it follows that

$$4\ell(\ell + 1) = -1 - \frac{mZ^2e^4}{2\hbar^2\epsilon};$$

that is,

$$-\frac{mZ^2e^4}{2\hbar^2\epsilon} = 4\ell(\ell + 1) + 1 = (2\ell + 1)^2.$$

This yields for the energy value

$$\epsilon = -\frac{m(Ze^2)^2}{2\hbar^2(2\ell + 1)^2}.$$

If we introduce the quantum number

$$n \stackrel{\text{def}}{=} 2\ell + 1,$$

then n must be one of the numbers

$$1, 2, 3, 4, \dots,$$

and the possible energy values are (with the usual notation E instead of ϵ)

$$E_n = -\frac{m(Ze^2)^2}{2\hbar^2n^2}.$$

(7.20)

Did you notice that these are the exact same values as predicted in Bohr's atom model (1.29)? We solved the eigenvalue problem of the hydrogen atom! For $Z = 1$, (7.20) defines its energy spectrum. n is called the *principal quantum number*. In addition to the principal quantum number, there is still the *angular momentum quantum number* ℓ and the *magnetic quantum number* m (see the chapter on angular momentum). Note that this equation was derived without describing the hydrogen atom in detail, such as electrons orbiting around the nucleus.

7.3 The Hydrogen Spectrum

In Sect. 2.1 we have already provided some information on the hydrogen spectrum. Back then, however, the relationships were more or less an educated guess; see Balmer. Based on the results of Pauli's calculations, we can now give a more profound summary.

For an arbitrary nuclear charge (with Z protons), we get an energy

$$E_n = -\frac{me^4}{8\varepsilon_0^2 h^2} \frac{Z^2}{n^2} = -13,6 \frac{Z^2}{n^2} \text{ eV}, \quad (7.21)$$

where ε_0 is the so-called *permittivity of free space*. For the energy difference between the n_1 th and the n_2 th state, we obtain

$$\Delta E = E_{n_2} - E_{n_1} = \frac{me^4 Z^2}{8\varepsilon_0^2 h^2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right). \quad (7.22)$$

For $n_2 > n_1$, this energy difference is positive; that is, the total energy of the system is increased by external energy input. Otherwise energy is emitted. This so-called *Rydberg formula* was found by Johannes Rydberg³ in 1888. Rydberg had no knowledge of any atomic model, but found this formula based on the observed line spectra only. A few years earlier, Balmer⁴ had found the famous Balmer formula for the case $n_1 = 2$ in the hydrogen atom ($Z = 1$),

$$\lambda = A \left(\frac{n^2}{n^2 - 4} \right) = A \left(\frac{n^2}{n^2 - 2^2} \right),$$

which covers the visible region of the spectrum lines. The empirical constant is $A = 364.56 \text{ nm} = 3645.6 \times 10^{-10} \text{ m}$. For the explanation of the spectra, we are interested in the frequency. According to Einstein and Planck, it is $E = h\nu$. For a jump from the n_1 th to the n_2 th state ($n_1 > n_2$), the frequency of the emitted radiation is therefore

$$\nu = \frac{c}{\lambda} = \frac{me^4}{8\varepsilon_0^2 h^3} \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right) = cR \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right), \quad (7.23)$$

with the *Rydberg constant*

$$R = \frac{me^4}{8\varepsilon_0^2 h^2} = 10\,973\,731.568\,539\,(55) \text{ m}^{-1}.$$

³Johannes Robert Rydberg, 1854–1919, Swedish physicist.

⁴Johann Jakob Balmer, 1825–1898, Swiss mathematics teacher.

If we set n_1 to infinity in (7.22), we obtain the energy that is needed to move an electron from infinity to the state n_2 , so the total energy of the state n_2 .

7.4 Problems

7.1 Time Derivative of \mathfrak{P} : Under which conditions is (7.10) equal to $\dot{\mathfrak{P}} = -Ze^2\mathfrak{R}\mathbf{R}^{-1}$?

7.2 Time Invariant Matrix: Show the relation (7.12),

$$\frac{d}{dt}(\mathfrak{R}\mathbf{R}^{-1}) = \frac{1}{2m} [\mathfrak{L} \times (\mathfrak{R}\mathbf{R}^{-3}) - (\mathfrak{R}\mathbf{R}^{-3}) \times \mathfrak{L}].$$

7.3 Constancy of \mathfrak{A} : Show that the Lenz-Matrix-Vector \mathfrak{A} is constant.

Chapter 8

Spin

Abstract We introduce the spin based on symmetry considerations. Its effect is described by spinors and Pauli matrices. Also, spin-orbit coupling is investigated.

8.1 Magnetic Fields and Light

In the nineteenth century, the influence of magnetic fields on light was studied extensively. Based on the idea of classical physics that light is an electromagnetic wave caused by atomic vibrations, in 1892 Lorentz¹ theoretically derived that the spectral lines are split threefold once the radiating atom is placed in a magnetic field. In 1896, the Dutch physicist Zeeman² confirmed this splitting in an experiment. Subsequent measurements of the splitting showed that it even complies with Lorentz's formula if it is assumed that not the entire atom, but only the much lighter electron vibrates. Already back then, electrons were assumed to be part of the atom. Thanks to this so-called *normal Zeeman effect*, the electron hypothesis was considerably strengthened.

In addition to the normal Zeeman effect, however, a great number of observations were made where more than three lines appear. This so-called *anomalous Zeeman effect* was an inexplicable phenomenon both for classical physics and Bohr's atom model and stimulated further theoretical investigation.

8.2 Derivation of the Zeeman Effect (Without Spin)

According to classical physics, an electron with the mass m and the electric charge e that moves on a circular path with the angular momentum \mathbf{L} has a *magnetic moment*

$$\mathbf{p}_m = -\frac{e}{2m} \mathbf{L}.$$

¹Hendrik Antoon Lorentz, 1853–1928, Dutch mathematician and physicist, Nobel Prize 1902.

²Pieter Zeeman, 1865–1943, Dutch physicist, Nobel Prize 1902.

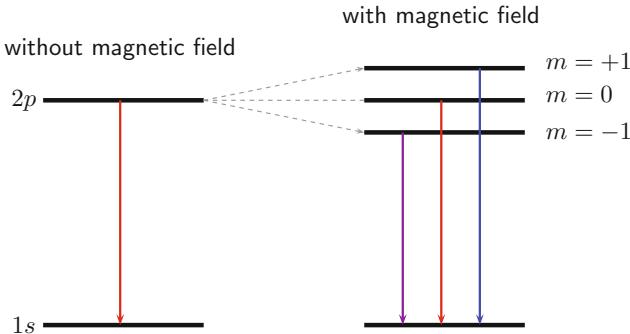


Fig. 8.1 Splitting of spectral lines due to the Zeeman effect. Photons with three different energies can be emitted

The factor $\frac{e}{2m}$ is called the *gyromagnetic ratio*. Under the influence of a magnetic field \mathbf{b} , there is an additional contribution to the potential energy of the electron in the Hamiltonian function to

$$E_m = -\mathbf{p}_m \cdot \mathbf{b} = \frac{e}{2m}(\mathbf{L}_3 \cdot \mathbf{b}),$$

where we assumed the magnetic field to be parallel to the x_3 -axis, having the magnitude b . If we transfer this relationship into quantum mechanics, using the eigenvalue $\hbar m$ of \mathbf{L}_3 , we get

$$E_m = \frac{e}{2m} \cdot b \cdot \hbar m = \mu_{Bohr} \cdot b \cdot m,$$

with the *Bohr magneton*

$$\mu_{Bohr} \stackrel{\text{def}}{=} \frac{e\hbar}{2m} = 9.2732 \cdot 10^{-24} \frac{J}{T}.$$

Overall, we now get the energy

$$E_{n,m} \stackrel{\text{def}}{=} E_n + \mu_{Bohr} b m.$$

The so-called Zeeman splitting then yields an energy difference for $\Delta m = \pm 1$ (Fig. 8.1):

$$\Delta E = E_{n,m+1} - E_{n,m} = \pm \mu_{Bohr} b.$$

For the so-called one-electron atoms [21] (i.e., one electron on the outermost electron shell), the standard terminology for the energy levels $\ell = 0, 1, 2, 3$ is s , p , d , and f . For the hydrogen atom and other one-electron atoms, these letters are preceded by a number that indicates the total energy level. For example, the lowest energy level in the hydrogen atom is $1s$, the next $2s$ and $2p$, then $3s$, $3p$ and $3d$, and

so on. For each n , the value of ℓ ranges from 0 to $n - 1$, and for each ℓ we have $2\ell + 1$ values of m . The total number of states with an energy value of E_n is therefore

$$\sum_{\ell=0}^{n-1} (2\ell + 1) = 2 \frac{n(n-1)}{2} + n = n^2.$$

The alkali metals lithium, sodium, potassium, and so on are one-electron atoms that consist of $Z - 1$ inner electrons and one external electron. The transitions of the latter between the energy levels are responsible for the spectral lines. If no external fields are present, the energy levels depend on the angular momentum quantum number ℓ as well as the principal quantum number n . Due to the spherical symmetry, they do not depend on the x_3 -component $\hbar m$ of the angular momentum. In other words, there is one energy level for each n , ℓ , and m . However, studies of the spectra revealed that (except for the s -level) all levels occur twice. For example, the so-called D -line of sodium, which is generated by the transition $3p \rightarrow 3s$, occurs twice with the wavelengths 589.6 and 589.0 nm. It is for this very reason that Pauli predicted a fourth quantum number for electrons in atoms that has two values (except for the s -level). In 1925, Uhlenbeck³ and Goudsmit⁴ proposed that the doubling of the energy level is created by an internal angular momentum of the electron. The component in the direction of the angular momentum L of the electron orbit around the nucleus can only assume two values, and its interaction with the weak magnetic field formed by the electrons that orbit around the nucleus, all except the s category, splits into two levels. All components of the inner angular momentum S can have $2\ell + 1$ values, such that the size s corresponding to ℓ assumes the unusual value $\frac{1}{2}$. This inner angular momentum was called the *spin* of the electron. How is a spin of size $\hbar/2$ generated? For an angular momentum of this amount and the classical electron radius of $r_E = \alpha\hbar/m_e c$, a peripheral speed would be required that is more than 70 times the speed of light! From a classical perspective, the spin can therefore not be explained as an angular momentum. Let us approach the solution to this question via symmetry considerations.

8.3 Symmetry Considerations

Historically, classical mechanics supplied quantum mechanics with observable quantities and their properties [21]. In some cases, this contradicts Heisenberg's idea to use only *measurable*, that is, observable, quantities, for example, when analyzing the motion of an electron around a nucleus (which cannot be observed). However, many of the entities required in quantum mechanics are introduced solely by *symmetry considerations*.

³George Uhlenbeck, Dutch/American physicist, 1900–1988.

⁴Samuel Goudsmit, Dutch/American physicist, 1902–1978.

The *symmetry principle* states that a law of nature does not change, for example, if we change our viewpoint. In other words, if a law is valid here (locus \mathbf{x}), then the law is also valid there (at the location $\mathbf{x} + \mathbf{a}$).

The unity operator \mathbf{I} represents a trivial symmetry that does not alter the state. Next, there is a special class of symmetries that can be represented by linear unitary operators \mathbf{U} that are arbitrarily close to \mathbf{I} . Such operators can be written as a *Taylor series*

$$\mathbf{U}(\epsilon) \stackrel{\text{def}}{=} \mathbf{I} + \epsilon \cdot \frac{d\mathbf{U}}{d\epsilon} \Big|_{\epsilon=0} + \mathcal{O}(\epsilon^2), \quad (8.1)$$

where ϵ is arbitrarily small. In order for \mathbf{U} to be unitary (i.e., $\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$), we must require

$$\mathbf{U}\mathbf{U}^\dagger = \left(\mathbf{I} + \epsilon \cdot \frac{d\mathbf{U}}{d\epsilon} \Big|_{\epsilon=0} + \mathcal{O}(\epsilon^2) \right) \left(\mathbf{I} + \epsilon \cdot \frac{d\mathbf{U}^\dagger}{d\epsilon} \Big|_{\epsilon=0} + \mathcal{O}(\epsilon^2) \right) \stackrel{!}{=} \mathbf{I}.$$

This leads to

$$\underbrace{\mathbf{I} + \epsilon \left[\frac{d\mathbf{U}}{d\epsilon} + \frac{d\mathbf{U}^\dagger}{d\epsilon} \right] \Big|_{\epsilon=0}}_{\stackrel{!}{=}\mathbf{0}} + \mathcal{O}(\epsilon^2) \stackrel{!}{=} \mathbf{I}.$$

Let us now introduce the notation

$$\frac{d\mathbf{U}}{d\epsilon} = i \cdot \mathbf{G}, \quad (8.2)$$

with the (symmetric) *generator*

$$\mathbf{G} = \mathbf{G}^\dagger. \quad (8.3)$$

Then we get indeed

$$\left[\frac{d\mathbf{U}}{d\epsilon} + \frac{d\mathbf{U}^\dagger}{d\epsilon} \right] = i \cdot \mathbf{G} - i \cdot \mathbf{G} = \mathbf{0}.$$

With the definition $\epsilon \stackrel{\text{def}}{=} \theta/N$, (8.1) now reads

$$\mathbf{U}(\theta/N) = \mathbf{I} + i \cdot \theta/N \cdot \mathbf{G} + \mathcal{O}(\epsilon^2).$$

Now let θ be fixed and N get large. Then we can obtain the transformation according to θ by applying N times the infinitesimal transformation according to θ/N :

$$\mathbf{U}(\theta) = (\mathbf{I} + i \cdot \theta/N \cdot \mathbf{G})^N.$$

For $N \rightarrow \infty$, this becomes

$$\underline{\underline{U}}(\theta) = \lim_{N \rightarrow \infty} (\mathbf{I} + i \cdot \theta / N \cdot \mathbf{G})^N = \underline{\underline{e}}^{i \theta \mathbf{G}}. \quad (8.4)$$

Note that we left the higher-order terms $\mathcal{O}(\epsilon^2)$ aside in this rather heuristic derivation. However, it can be shown in a mathematically rigorous way that this is indeed a feasible definition for *any* size of θ ! In particular, the first terms of the series representation

$$e^X = \sum_{m=0}^{\infty} \frac{X^m}{m!},$$

with $X = i\varepsilon\mathbf{G}$, indeed yield the desired result (8.1) for small ϵ .

It is well known from mathematics that a similarity transformation leaves eigenvalues and eigenvectors unchanged. Now we want to show that there is a unitary transformation such that a coordinate shift occurs according to

$$\mathbf{x} \Rightarrow \mathbf{x} + \mathbf{a}, \text{ for all } \mathbf{x},$$

where \mathbf{a} is an arbitrary three-dimensional vector. In the case of the Heisenberg matrices X_j of the matrix vector

$$\mathfrak{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix},$$

the equivalent condition is

$$X_j \Rightarrow X_j + a_j \mathbf{I}, j = 1, 2, 3,$$

or

$$\mathfrak{X} \Rightarrow \mathfrak{X} + \mathbf{a} \otimes \mathbf{I}.$$

In other words, we are looking for a unitary matrix $\mathbf{U}(\mathbf{a})$ such that

$$\mathbf{U}(\mathbf{a}) X_j \mathbf{U}(\mathbf{a})^\dagger = X_j + a_j \mathbf{I}. \quad (8.5)$$

For infinitesimally small a_j , \mathbf{U} must have a shape like that in (8.1). Therefore, it seems useful to start with the ansatz

$$\mathbf{U}(\mathbf{a}) = \mathbf{I} + \frac{i}{\hbar} (a_1 \mathbf{P}_1 + a_2 \mathbf{P}_2 + a_3 \mathbf{P}_3) + \mathcal{O}(\mathbf{a}^2). \quad (8.6)$$

The dimension of \hbar is *length*² · *mass/time*. It turns out later that \mathbf{P} represents the momentum (with dimension *length* · *mass/time*) and that the a_i have the dimension *length*. It is therefore convenient to add the \hbar in order to render \mathbf{U} a dimensionless transformation matrix. For an infinitesimal three-vector \mathbf{a} the condition (8.5) requires that

$$i[\mathbf{P}_j \cdot \mathbf{a}, X_j]/\hbar = a_j \mathbf{I}.$$

\mathbf{P}_j commutes with itself, and of course also with any function of \mathbf{P}_j . Therefore, we get

$$\mathbf{U}(\mathbf{a}) \mathbf{P}_j \mathbf{U}^\dagger(\mathbf{a}) = \mathbf{P}_j \mathbf{U}(\mathbf{a}) \mathbf{U}^\dagger(\mathbf{a}) = \mathbf{P}_j. \quad (8.7)$$

In other words, \mathbf{P}_j remains unchanged.

Does this transformation also shift the position vector $\mathbf{x}' = \mathbf{x} + \mathbf{a}$; that is, $\mathfrak{X}' = \mathfrak{X} + \mathbf{a} \otimes \mathbf{I}$? We have

$$X'_j(\mathbf{a}) = \mathbf{U}(\mathbf{a}) X_j \mathbf{U}^\dagger(\mathbf{a}).$$

Let us now form the derivative with respect to a_k :

$$\frac{\partial}{\partial a_k} X'_j(\mathbf{a}) = \left[\frac{\partial}{\partial a_k} \mathbf{U}(\mathbf{a}) \right] X_j \mathbf{U}^\dagger(\mathbf{a}) + \mathbf{U}(\mathbf{a}) X_j \frac{\partial}{\partial a_k} \mathbf{U}^\dagger(\mathbf{a}). \quad (8.8)$$

Differentiating (8.4) by a_k and exploiting the fact that all three \mathbf{P}_j matrices commute with each other, one obtains

$$\frac{\partial}{\partial a_k} \mathbf{U}(\mathbf{a}) = \frac{i}{\hbar} \mathbf{U}(\mathbf{a}) \mathbf{P}_k, \quad \frac{\partial}{\partial a_k} \mathbf{U}^\dagger(\mathbf{a}) = -\frac{i}{\hbar} \mathbf{P}_k \mathbf{U}^\dagger(\mathbf{a}).$$

Putting this result in (8.8) yields

$$\frac{\partial}{\partial a_k} X'_j(\mathbf{a}) = -\frac{i}{\hbar} \mathbf{U}(\mathbf{a}) \underbrace{[X_j, \mathbf{P}_k]}_{i \hbar \delta_{jk} \mathbf{I}} \mathbf{U}^\dagger(\mathbf{a}) = \delta_{jk} \mathbf{I}, \quad (8.9)$$

where we used the known fact

$$X_j \mathbf{P}_k - \mathbf{P}_k X_j = [X_j, \mathbf{P}_k] = i \hbar \delta_{jk} \mathbf{I}.$$

We can now integrate (8.9) to get

$$X'_j(\mathbf{a}) = X'_j(0) + a_j \mathbf{I}.$$

Because of $\mathbf{U}(0) = \mathbf{I}$, we have $X'_j(0) = X_j$. Therefore, we indeed find that

$$\mathfrak{X}'(\mathbf{a}) = \mathfrak{X} + \mathbf{a} \otimes \mathbf{I}.$$

The most exciting result of this whole consideration is that we found the previously called momentum matrix

$$\mathfrak{P} = \begin{pmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{P}_3 \end{pmatrix}$$

to be a generator of a symmetry operation, without having to use classical mechanics as a justification.

8.4 Symmetry and the Spin

Consider now an infinitesimal *rotation* of a physical system around the x_3 -axis by an angle $\delta\varphi$. The components of a vector \mathbf{v} then change according to

$$v'_1 = v_1 + \delta\varphi v_2; \quad v'_2 = v_2 - \delta\varphi v_1 \quad \text{und} \quad v'_3 = v_3.$$

In vector form, this reads

$$\mathbf{v}' = \mathbf{v} - \boldsymbol{\omega} \times \mathbf{v}, \quad (8.10)$$

where $\boldsymbol{\omega}$ is defined as

$$\boldsymbol{\omega} \stackrel{\text{def}}{=} \begin{pmatrix} 0 \\ 0 \\ \delta\varphi \end{pmatrix}.$$

If we want this infinitesimal rotation to be performed by means of a generator \mathbf{G} via a transformation $\mathbf{U}(\delta\varphi) = \mathbf{I} + i\mathbf{G}\delta\varphi + \mathcal{O}(\delta\varphi^2)$, we must start with

$$\mathbf{G} = \frac{1}{\hbar}(\boldsymbol{\omega} \cdot \mathfrak{J}) = \frac{1}{\hbar}(\omega_1 \mathbf{J}_1 + \omega_2 \mathbf{J}_2 + \omega_3 \mathbf{J}_3). \quad (8.11)$$

It will turn out that \mathfrak{J} is the *total angular momentum*, a 3-vector matrix, consisting of the three matrices \mathbf{J}_1 , \mathbf{J}_2 , and \mathbf{J}_3 . Applying the transformation to \mathbf{V}_j yields

$$\mathbf{V}_j' = (\mathbf{I} - i\mathbf{G})\mathbf{V}_j(\mathbf{I} + i\mathbf{G}) = \mathbf{V}_j + i(\mathbf{V}_j \mathbf{G} - \mathbf{G} \mathbf{V}_j) + \mathcal{O}(\delta\varphi^2), \quad (8.12)$$

or with (8.11)

$$\mathbf{V}_j' - \mathbf{V}_j = i(\mathbf{V}_j \mathbf{G} - \mathbf{G} \mathbf{V}_j) = i[\mathbf{V}_j, \mathbf{G}] = \frac{i}{\hbar}[\mathbf{V}_j, \boldsymbol{\omega} \cdot \mathfrak{J}]. \quad (8.13)$$

Multiplying (8.13) with a_1 , a_2 , and a_3 , and summing up the three equations, we get

$$\mathbf{a} \cdot (\mathfrak{V}' - \mathfrak{V}) = \frac{1}{i\hbar}[\mathbf{a} \cdot \mathfrak{V}, \boldsymbol{\omega} \cdot \mathfrak{J}], \quad (8.14)$$

where

$$\mathbf{a} \cdot \mathfrak{V} = a_1 V_1 + a_2 V_2 + a_3 V_3$$

and

$$\boldsymbol{\omega} \cdot \mathfrak{J} = \omega_1 \mathbf{J}_1 + \omega_2 \mathbf{J}_2 + \omega_3 \mathbf{J}_3.$$

From (8.10) follows

$$\frac{1}{i\hbar}[\mathbf{a} \cdot \mathfrak{V}, \boldsymbol{\omega} \cdot \mathfrak{J}] = \mathbf{a} \cdot (\boldsymbol{\omega} \times \mathfrak{V}). \quad (8.15)$$

In compliance with (see also Problem 6.2)

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c},$$

we get the relationship

$$\frac{1}{i\hbar}[\mathbf{a} \cdot \mathfrak{V}, \boldsymbol{\omega} \cdot \mathfrak{J}] = \mathbf{a} \cdot (\boldsymbol{\omega} \times \mathfrak{V}) = (\mathbf{a} \times \boldsymbol{\omega}) \cdot \mathfrak{V}. \quad (8.16)$$

If \mathbf{a} and $\boldsymbol{\omega}$ are parallel to each other, it follows that

$$[V_j, J_j] = \mathbf{0}, \quad j = 1, 2 \text{ or } 3. \quad (8.17)$$

If \mathbf{a} and $\boldsymbol{\omega}$ are perpendicular, however, it follows that

$$\frac{1}{i\hbar}[V_j, J_k] = \begin{cases} +V_\ell & \text{if } j, k, \ell \text{ are cyclic permutations of } 1, 2, 3, \\ -V_\ell & \text{if } j, k, \ell \text{ are anti-cyclic permutations.} \end{cases} \quad (8.18)$$

We can summarize this as⁵

$$[V_j, J_k] = i\hbar \sum_{\ell} \epsilon_{jkl} V_{\ell}. \quad (8.19)$$

The same is true for $V_j = J_j = L_j$, thus

$$[L_j, L_k] = i\hbar \sum_{\ell} \epsilon_{jkl} L_{\ell}. \quad (8.20)$$

Now we are ready to turn to the spin. For that purpose, let us define a vector matrix \mathfrak{S} according to

$$\mathfrak{S} \stackrel{\text{def}}{=} \mathfrak{J} - \mathfrak{L};$$

⁵The Levi-Civita symbol ϵ_{jkl} is defined as

$$\epsilon_{jkl} = \begin{cases} +1 & \text{if } (j, k, \ell) \text{ is } (1, 2, 3), (2, 3, 1) \text{ or } (3, 1, 2), \\ -1 & \text{if } (j, k, \ell) \text{ is } (3, 2, 1), (1, 3, 2) \text{ or } (2, 1, 3), \\ 0 & \text{if } j = k \text{ or } k = \ell \text{ or } \ell = j \end{cases}$$

that is,

$$\mathfrak{J} = \mathfrak{L} + \mathfrak{S}. \quad (8.21)$$

By subtracting (8.20) from (8.19), after substituting \mathbf{L}_j for \mathbf{V}_j in (8.19), we find

$$[S_i, \mathbf{L}_j] = \mathbf{0}. \quad (8.22)$$

Together with the other equations above, we obtain

$$[S_j, S_k] = i\hbar \sum_{\ell} \epsilon_{jkl} S_{\ell}. \quad (8.23)$$

In other words, \mathfrak{S} also behaves as an angular momentum and is called the *spin*. The spin angular momentum \mathfrak{S} of a closed system, therefore, is the proportion of the total angular momentum \mathfrak{J} that is *not* due to an orbital angular momentum \mathfrak{L} .

Goudsmith and Uhlenbeck suggested in 1925 to assign such an additional angular eigenmomentum called spin to the electron. It must have a half-integral angular momentum quantum number $s = \frac{1}{2}$ in order for the magnetic spin quantum number to be limited to two possible values $m_s = \pm \frac{1}{2}$. The spin accounts for a twofold or, together with an orbital angular momentum $\ell \geq 1$, a higher splitting of the spectrum.

8.5 Spin- $\frac{1}{2}$ Systems and Spinors

We can obtain the matrices representing S_1 , S_2 , and S_3 from (6.37) and (6.54):

$$S_1 \stackrel{\text{def}}{=} \mathbf{L}_1^{(1/2)} = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad S_2 \stackrel{\text{def}}{=} \mathbf{L}_2^{(1/2)} = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

and $S_3 \stackrel{\text{def}}{=} \mathbf{L}_3^{(1/2)} = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (8.24)$

Using the *Pauli matrices*⁶ σ_i , we can write

$$S_i = \frac{\hbar}{2} \sigma_i, \quad (8.25)$$

where the Pauli matrices have the form

$$\sigma_1 \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 \stackrel{\text{def}}{=} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_3 \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (8.26)$$

⁶We stick to the traditional notation for these matrices.

The eigenvalues of these matrices are ± 1 , and we have

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = I_2. \quad (8.27)$$

Pauli was the first to recognize the need for two-component state vectors for describing the observed properties of atomic spectra.

The spin-(1/2) matrices S_1 , S_2 , and S_3 have eigenvalues and two-dimensional, normalized eigenvectors as follows.

$$S_1 = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ has eigenvalues } \pm \frac{\hbar}{2} \text{ and eigenvectors } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$$

,

$$S_2 = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \text{ has eigenvalues } \pm \frac{\hbar}{2} \text{ and eigenvectors } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \end{pmatrix}$$

,

$$S_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ has eigenvalues } \pm \frac{\hbar}{2} \text{ and eigenvectors } \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

8.6 Adding Angular Momenta

8.6.1 Clebsch–Gordan Coefficients

We now consider a system with two sources of angular momenta, which we call

$$\mathbf{J}' = \begin{pmatrix} J'_1 \\ J'_2 \\ J'_3 \end{pmatrix} \text{ and } \mathbf{J}'' = \begin{pmatrix} J''_1 \\ J''_2 \\ J''_3 \end{pmatrix}.$$

The system may be a single particle with a spin and an angular momentum, or may consist of two particles with either spin or angular momentum. What are the commutator rules for the total angular momentum? As for all angular momenta, the following eigenvalue equations apply.

$$\mathbf{J}'^2 \mathbf{e}' = j'(j' + 1)\hbar^2 \mathbf{e}', \quad (8.28)$$

$$J'_3 \mathbf{e}' = m' \hbar \mathbf{e}', \quad (8.29)$$

and

$$\mathbf{J}''^2 \mathbf{e}'' = j''(j'' + 1)\hbar^2 \mathbf{e}'', \quad (8.30)$$

$$\mathbf{J}_3'' \mathbf{e}'' = m'' \hbar \mathbf{e}'', \quad (8.31)$$

where

$$m' = j', j' - 1, \dots, -j'; \quad \text{und} \quad m'' = j'', j'' - 1, \dots, -j''.$$

For independent angular momentum sources, we certainly have

$$[\mathbf{J}'_k, \mathbf{J}''_\ell] = \mathbf{0} \quad \text{for all } k, \ell \in \{1, 2, 3\}, \text{ and} \quad [\mathbf{J}'^2, \mathbf{J}''^2] = \mathbf{0}. \quad (8.32)$$

However,

$$[\mathbf{J}'_j, \mathbf{J}'_k] = i\hbar \epsilon_{jkl} \mathbf{J}'_\ell, \quad (8.33)$$

and

$$[\mathbf{J}''_j, \mathbf{J}''_k] = i\hbar \epsilon_{jkl} \mathbf{J}''_\ell. \quad (8.34)$$

Is the sum

$$\mathfrak{J} = \begin{pmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \\ \mathbf{J}_3 \end{pmatrix} \stackrel{\text{def}}{=} \mathfrak{J}' + \mathfrak{J}''$$

also an angular momentum that satisfies the usual commutation rules? In fact, we get

$$\begin{aligned} [\mathbf{J}_1, \mathbf{J}_2] &= [\mathbf{J}'_1 + \mathbf{J}''_1, \mathbf{J}'_2 + \mathbf{J}''_2] = \\ &= [\mathbf{J}'_1, \mathbf{J}'_2] + [\mathbf{J}''_1, \mathbf{J}''_2] + [\mathbf{J}'_1, \mathbf{J}''_2] + [\mathbf{J}''_1, \mathbf{J}'_2] = \\ &= i\hbar \mathbf{J}'_3 + i\hbar \mathbf{J}''_3 + \mathbf{0} + \mathbf{0} = \\ &= i\hbar \mathbf{J}_3. \end{aligned} \quad (8.35)$$

The same applies to all cyclic permutations of the indices, therefore $\mathfrak{J} = \mathfrak{J}' + \mathfrak{J}''$ is indeed an angular momentum. Its quantum numbers j and m_j can have the following values. Generally we know, that for given j ,

$$-j \leq m \leq j \quad \text{and} \quad m_{max} = j.$$

Because $m = m' + m''$, the maximum value of m for all j is $j' + j''$; that is, the maximum value of $m' + m''$. This must be the maximum value of j also, for otherwise there would be a larger value of $m' + m''$. Thus

$$j_{max} = j' + j''.$$

With the minimum value of $j_{min} = |j' - j''|$, given at the end of this subsection, we obtain

$$|j' - j''| \leq j \leq |j' + j''|,$$

$$m_j = -j, -j+1, \dots, j.$$

Summing up, \mathbf{J} is an angular momentum with the length $\sqrt{j(j+1)}\hbar$, where j can be an integer or half-integer, and the x_3 -component has the value $m_j\hbar$ (with $m_j = j, j-1, \dots, -j$).

Each angular momentum has its own eigenspace that is spanned by the eigenvectors \mathbf{e}'_i and \mathbf{e}''_i , respectively. In the basis of the eigenvectors \mathbf{e}'_i , \mathbf{J}'_j has a simple diagonal form. The same applies to \mathbf{J}''_j . The eigenvectors of the overall system can be assembled of eigenvectors for the subsystems (see Chap. 10) according to

$$\mathbf{e}_i = \mathbf{e}'_j \otimes \mathbf{e}''_k.$$

However, these vectors are generally not eigenvectors to

$$\mathbf{J}^2 = \mathbf{J}_1^2 + \mathbf{J}_2^2 + \mathbf{J}_3^2,$$

As a consequence, this matrix is not diagonal in this basis. Therefore, one better shifts from the complete set of commuting matrices $\mathbf{J}'^2, \mathbf{J}'_3, \mathbf{J}''^2, \mathbf{J}''_3$ with the eigenstates $\mathbf{e}'_j \otimes \mathbf{e}''_k$ to the complete set of commuting matrices $\mathbf{J}^2, \mathbf{J}_3, \mathbf{J}'^2, \mathbf{J}''^2$, with the eigenvectors $\mathbf{e}(j, m_j, j', j'')$. In this new basis, all four matrices $\mathbf{J}^2, \mathbf{J}_3, \mathbf{J}'^2$, and \mathbf{J}''^2 can be diagonalized simultaneously. The new eigenvectors satisfy the following eigenvalue equations:

$$\mathbf{J}^2 \mathbf{e}(j, m_j, j', j'') = \hbar^2 j(j+1) \mathbf{e}(j, m_j, j', j''),$$

$$\mathbf{J}_3 \mathbf{e}(j, m_j, j', j'') = \hbar m_j \mathbf{e}(j, m_j, j', j'').$$

How can we obtain the new basis vectors $\mathbf{e}(j, m_j, j', j'')$ if the basis vectors $\mathbf{e}'_j \otimes \mathbf{e}''_k$ are given? The elements of one basis must be a linear combination of the elements of the other basis. In the basis with the basis vectors $\mathbf{e}'_j \otimes \mathbf{e}''_k$, the new basis vector $\mathbf{e}(j, m_j, j', j'')$ looks like

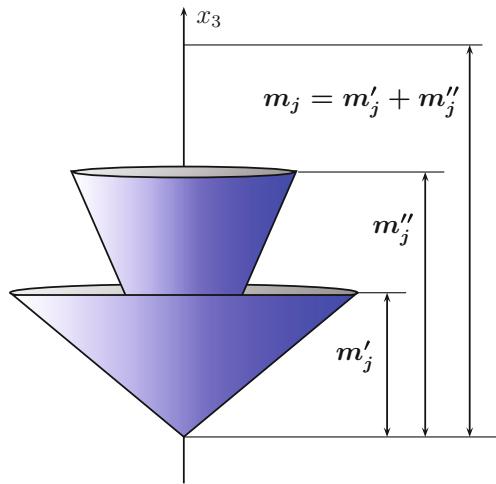
$$\mathbf{e}(j, m_j, j', j'') = \sum_{a,b} C(a, b; j, m_j, j', j'') \mathbf{e}'_a \otimes \mathbf{e}''_b. \quad (8.36)$$

The coefficients $C(a, b; j, m_j, j', j'')$ can be calculated by multiplying (8.36) from the left with the transposed basis vector $\mathbf{e}'_a \otimes \mathbf{e}''_b$ (remember that the basis vectors $\mathbf{e}'_a \otimes \mathbf{e}''_b$ are orthogonal). These coefficients are called *Clebsch*⁷-*Gordan*⁸-coefficients

⁷Rudolf Friedrich Alfred Clebsch, 1833–1872, German mathematician.

⁸Paul Albert Gordan, 1837–1912, German mathematician.

Fig. 8.2 The relationship
 $m_j = m'_j + m''_j$



and are defined by

$$C(a, b; j, m_j, j', j'') = (\mathbf{e}'_a \otimes \mathbf{e}''_b)^\dagger \mathbf{e}(j, m_j, j', j''). \quad (8.37)$$

Now which values j can exist for the system for a given j' and j'' ? Because \mathbf{J}^2 commutes with its own components, it commutes in particular with $\mathbf{J}_3 = \mathbf{J}'_3 + \mathbf{J}''_3$. This shows us that we can specify the value of m_j and j at the same time. The allowed values for m_j follow immediately from the relationship

$$\mathbf{J}_3 = \mathbf{J}'_3 + \mathbf{J}''_3;$$

that is (see Fig. 8.2),

$$m_j = m'_j + m''_j. \quad (8.38)$$

To determine the allowed values of j , we first note that the complete number of possible values for the uncoupled systems is

$$(2j' + 1)(2j'' + 1) = 4j'j'' + 2j' + 2j'' + 1.$$

There is only one state in which both components reach their maximum value, namely $m'_j = j'$ and $m''_j = j''$ (which entails $m_j = j' + j''$). However, the maximum value of m_j is j by definition, therefore the maximum value of j is $j = j' + j''$. Note that the Clebsch–Gordan coefficients only differ from zero if $|j' - j''| \leq j \leq j' + j''$ and $m_j = m' + m''$.

8.6.2 Clebsch–Gordan Coefficients on the Internet

You can find the following formula for the calculation of the Clebsch–Gordan coefficients on the Internet:

$$\begin{aligned} C(m', m''; j, m_j, j', j'') &= \\ &= \delta_{m_j, m'+m''} \left(\frac{(2j+1)(j'+j''-j)!(j'-j''+j)!(j+j''-j')!}{(j'+j''+j+1)!} \right)^{\frac{1}{2}} \times \\ &\sum_n \left(\frac{(-1)^n [(j'+m'')!(j'-m')!(j'+m'')!(j''-m'')!(j+m_j)!(j-m_j)!]^{\frac{1}{2}}}{n!(j'+j''-j-n)!(j'-m'-n)!(j''+m''-n)!(j-j''+m'+n)!(j-j'-m''+n)!} \right). \end{aligned}$$

The sum runs over all n such that the factorials are always well-defined and positive.
The formula is available at

<http://www.spektrum.de/lexikon/physik/clebsch-gordan-koeffizienten/2438>.

A downloadable PDF table of Clebsch–Gordan coefficients can be found under
<http://pdg.lbl.gov/2011/reviews/rpp2011-rev-clebsch-gordan-coefs.pdf>.

You can also download a calculator under <http://www.volya.net/index.php>.

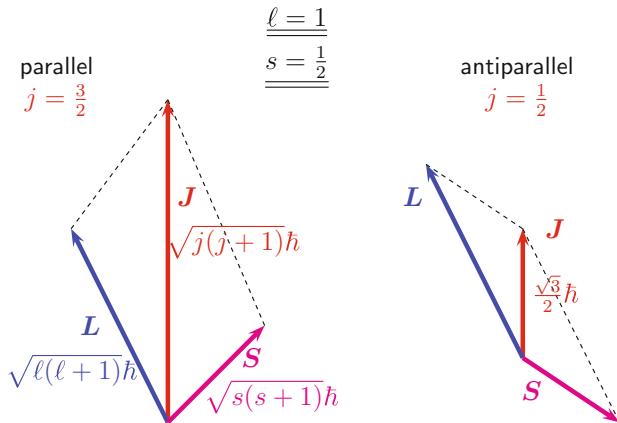
8.7 Spin-Orbit Coupling

This stands for the observation in atomic, nuclear, and elementary particle physics that a particle's direction of the spin with respect to its orbital angular momentum affects the energy of the particle. Good examples are electrons in the atomic shell, where the spin-orbit coupling leads to the splitting of the spectral lines and thus contributes to the fine structure of atomic spectra.

It is easiest to describe spin-orbit coupling of the electrons in a semi-classical model. Maxwell's theory and the special theory of relativity require that a magnetic field act on an electron when it circulates in the electric field of the atomic nucleus. After all, a circular motion of the nucleus is perceived in the electron's frame of reference. Due to the charge of the nucleus, this movement represents a circulating current that generates a magnetic field parallel to the angular momentum vector. Because the electron with its intrinsic angular momentum (spin) has a magnetic moment in its resting frame of reference as well, the spin direction parallel to the field yields a lower energy and vice versa. Hence, a single energy level is split into two levels, and we find two optical spectra that are slightly shifted compared to the original lines. The total angular momentum of the electron is a combination of the spin and the orbital part:

$$\mathbf{J} = \mathbf{L} + \mathbf{S}.$$

The usual quantum mechanical properties must be valid:

**Fig. 8.3** Spin-orbit coupling

$$\mathbf{J}^2 \mathbf{e}_{jm_j} = j(j+1)\hbar^2 \mathbf{e}_{jm_j},$$

$$\mathbf{J}_3 \mathbf{e}_{jm_j} = m_j \hbar \mathbf{e}_{jm_j},$$

$$-j \leq m_j \leq j,$$

where \mathbf{e}_{jm_j} is a common eigenvector of \mathbf{J}^2 and \mathbf{J}_3 .

As we already found out, in the hydrogen atom the total spin quantum number can have two values (depending on the coupling):

$$j = \ell + \frac{1}{2} (\text{parallel}).$$

$$j = \ell - \frac{1}{2} (\text{antiparallel}).$$

Figure 8.3 shows this result for $\ell = 1$ in vector form.

The following section gives a relatively simple description of the spin-orbit interaction for an electron that is bound to an atom. We use semi-classical electrodynamics and nonrelativistic quantum mechanics. The results agree quite well with the observations.

Energy of the Magnetic Dipole Moment

The energy of a magnetic dipole moment in a magnetic field is given by

$$\Delta E = -\mu \cdot \mathbf{b}$$

where μ is the magnetic dipole moment of the particle and \mathbf{b} is the magnetic induction of the magnetic field.

Magnetic Field

Let us first consider the magnetic field. An electron with mass m_e and charge e that moves with velocity \mathbf{v} in an electric field \mathbf{e} generates the magnetic field

$$\mathbf{b} = \frac{\mathbf{e} \times \mathbf{v}}{c^2}.$$

If the electric field is generated by the potential $U(r)$, we have

$$\mathbf{e} = -\frac{\mathbf{r}}{r} \frac{\partial U(r)}{\partial r}.$$

The unit vector \mathbf{r}/r implies a radial direction of the electric field. With this, we get

$$\mathbf{b} = -\frac{1}{rc^2} \frac{\partial U(r)}{\partial r} \mathbf{r} \times \mathbf{v}.$$

Therefore, if \mathbf{r} and \mathbf{v} are in the x_1, x_2 plane, \mathbf{b} points in x_3 -direction. Now remember the angular momentum of a particle,

$$\boldsymbol{\ell} = \mathbf{r} \times \mathbf{p},$$

and the relation $\mathbf{p} = m_e \mathbf{v}$. Then

$$\mathbf{b} = \frac{-1}{m_e c^2 r} \frac{\partial U(r)}{\partial r} \boldsymbol{\ell}. \quad (8.39)$$

In other words, the magnetic field \mathbf{b} is parallel to the orbital angular momentum $\boldsymbol{\ell}$ of the particle.

Magnetic Dipole Moment of the Electron

The magnetic dipole moment of the electron is given by

$$\boldsymbol{\mu} = -\frac{g_s \mu_B}{\hbar} \mathbf{s} = -\frac{g_s e}{2m_e} \mathbf{s}, \quad (8.40)$$

where \mathbf{s} is the spin angular momentum vector,

$$\mu_B \stackrel{\text{def}}{=} \frac{e\hbar}{2m_e} = 9.274 \cdot 10^{-24} \frac{J}{T}$$

is the Bohr magneton, and $g_s = 2,002319304 \approx 2$ is the electron spin g-factor⁹. We see that the magnetic dipole moment is anti-parallel to the spin angular momentum.

⁹The g-factor (Lande factor) is the ratio of the measured magnetic moment to the magnetic moment, which we would expect from classical physics. It comes into play when considering quantum electrodynamics (Dirac).

The potential due to the spin orbit interaction consists of two parts. The Larmor¹⁰-part has to do with the interaction of the magnetic moment of the electron with the magnetic field of the nucleus in the coordinate system of the electron. The second contribution is related to the so-called Thomas¹¹ precession.

In a homogeneous magnetic field with the magnetic induction density \mathbf{b} , a particle with a magnetic dipole moment $\boldsymbol{\mu}$ experiences the angular momentum $\mathbf{m} = \boldsymbol{\mu} \times \mathbf{b}$. The work required for the rotation of a magnetic dipole by an angle $d\varphi$ is given by

$$dW = -|\mathbf{m}| \cdot d\varphi = -|\boldsymbol{\mu}| \cdot |\mathbf{b}| \sin \varphi \cdot d\varphi,$$

where φ is the angle between $\boldsymbol{\mu}$ and \mathbf{b} (the minus sign is because φ decreases by the action of the angular momentum \mathbf{m}). This work equals the decrease of potential energy E_{pot} of the system; that is,

$$dE_{pot} = -dW = |\boldsymbol{\mu}| \cdot |\mathbf{b}| \sin \varphi \cdot d\varphi.$$

Integrating the equation, we get

$$E_{pot} = -|\boldsymbol{\mu}| \cdot |\mathbf{b}| \cos \varphi = -\boldsymbol{\mu} \cdot \mathbf{b}.$$

This contribution is also called *Larmor* interaction energy. Substituting the expressions for the magnetic moment and the magnetic field leads to

$$\Delta E_L = -\boldsymbol{\mu} \cdot \mathbf{b} = \frac{2\mu_B}{\hbar m_e c^2} \frac{1}{r} \frac{\partial U(r)}{\partial r} \boldsymbol{\ell} \cdot \mathbf{s}.$$

Next, we have to consider the *Thomas precession correction* that is caused by the curved path of the electron. In 1926, Llewellyn Thomas computed the splitting in the fine structure of the atomic spectrum in a relativistic framework. The Thomas precession rate, $\boldsymbol{\Omega}_T$, is related to the angular frequency of the orbital motion ω of a spinning particle as follows.

$$\boldsymbol{\Omega}_T = \omega(1 - \gamma),$$

where $\gamma = 1/\sqrt{1 - (v/c)^2}$ is the Lorentz factor of the moving particle. The Hamiltonian that causes the spin precession $\boldsymbol{\Omega}_T$ is given by¹²

$$\Delta E_T = \boldsymbol{\Omega}_T \cdot \mathbf{S}.$$

In the first order of $(v/c)^2$, this yields

¹⁰Joseph Larmor, 1857–1942, Irish physicist.

¹¹Llewellyn Thomas, 1903–1992, British physicist.

¹²Note that we now moved back from the matrix vectors to operators.

$$\Delta E_T = -\frac{\mu_B}{\hbar m_e c^2} \frac{1}{r} \frac{\partial U(r)}{\partial r} \mathbf{L} \cdot \mathbf{S}.$$

Total Interaction Energy

The total spin-orbit potential in an external electrostatic potential therefore has the form

$$\Delta E \equiv \Delta E_L + \Delta E_T = \frac{\mu_B}{\hbar m_e c^2} \frac{1}{r} \frac{\partial U(r)}{\partial r} (\mathbf{L} \cdot \mathbf{S}).$$

Note that the net effect of the Thomas precession is a reduction of the Larmor interaction energy by a factor $\frac{1}{2}$, which became famous as the “Thomas half”.

Evaluation of the Energy Shift

With all the above approximations, we can now determine the energy shift in this model. In particular, we want to find a vector basis that diagonalizes both H_0 (the nonperturbed Hamilton function) and H . For this purpose, we first define the total angular momentum matrix (operator)

$$\mathbf{J} = \mathbf{L} + \mathbf{S}.$$

The scalar product of \mathbf{J} with itself yields

$$\mathbf{J}^2 = \mathbf{L}^2 + \mathbf{S}^2 + 2\mathbf{L} \cdot \mathbf{S},$$

and therefore

$$\mathbf{L} \cdot \mathbf{S} = \frac{1}{2}(\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2).$$

It is easy to show that the five matrices \mathbf{H}_0 , \mathbf{J}^2 , \mathbf{L}^2 , \mathbf{S}^2 , and \mathbf{J}_z all commute with each other as well as with \mathbf{H} . Therefore, the basis that we are looking for is also a simultaneous eigenbasis of these five matrices (i.e., a basis in which all five matrices are diagonal). Eigenvectors of this basis have five quantum numbers:

- n (the “principal quantum number”)
- j (the “total angular momentum quantum number”)
- ℓ (the “orbital angular momentum quantum number”)
- s (the “spin quantum number”)
- j_3 (the “ x_3 -component of total angular momentum”)

The expectation value of $\mathbf{L} \cdot \mathbf{S}$ is

$$\langle \mathbf{L} \cdot \mathbf{S} \rangle = \frac{1}{2}(\langle \mathbf{J}^2 \rangle - \langle \mathbf{L}^2 \rangle - \langle \mathbf{S}^2 \rangle) = \frac{\hbar^2}{2}(j(j+1) - \ell(\ell+1) - s(s+1)).$$

To calculate the final energy shift, we can say that

$$\Delta E = \frac{\beta}{2}(j(j+1) - \ell(\ell+1) - s(s+1)),$$

where

$$\beta = \frac{-\mu_B}{m_e ec^2} \left\langle \frac{1}{r} \frac{\partial U(r)}{\partial r} \right\rangle.$$

For a nucleus with charge Ze , the Coulomb potential is

$$U(r) = \frac{Ze}{4\pi\epsilon_0 r}$$

Therefore,

$$\frac{\partial U}{\partial r} = \frac{Ze}{4\pi\epsilon_0} \left(\frac{\partial 1/r}{\partial r} \right) = \frac{-Ze}{4\pi\epsilon_0 r^2}. \quad (8.41)$$

For hydrogen, the expectation value for $\frac{1}{r^3}$ is given by (see, e.g., [1] p. 251)

$$\left\langle \frac{1}{r^3} \right\rangle = \frac{2}{a^3 n^3 \ell(\ell+1)(2\ell+1)},$$

where $a = \hbar/Z\alpha m_e c$ is the Bohr radius divided by the nuclear charge Ze . For hydrogen, we therefore get the result

$$\beta(n, l) = \frac{\mu_0}{4\pi} g_s \mu_B^2 \frac{1}{n^3 a_0^3 \ell(\ell+1/2)(\ell+1)}.$$

For any hydrogenlike atom with Z protons, one can show that

$$\boxed{\beta(n, l) = Z^4 \frac{\mu_0}{4\pi} g_s \mu_B^2 \frac{1}{n^3 a_0^3 \ell(\ell+1)(2\ell+1)}}. \quad (8.42)$$

This leads to an energy shift for the individual energy levels as

$$\boxed{\Delta E = Z^4 \frac{\mu_0 g_s \mu_B^2}{8\pi a_0^3} \frac{j(j+1) - \ell(\ell+1) - s(s+1)}{n^3 \ell(\ell+1)(2\ell+1)}}. \quad (8.43)$$

In general, the magnitude of this energy shift is $10^{-4} eV$. Note that the splitting due to spin-orbit coupling increases with increasing atomic number (namely as Z^4). Inasmuch as spin-orbit coupling is possible only for $\ell \geq 1$, the presence of ℓ in the denominator causes no issues.

For an electron in the p -shell, for example, the difference between the energies $j = \frac{3}{2}$ and $j = \frac{1}{2}$ is only

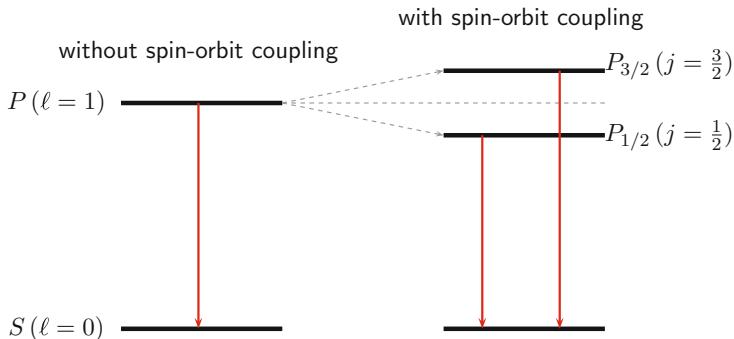


Fig. 8.4 Splitting of spectral lines due to spin-orbit coupling. Photons can be emitted at two different energies

$$Z^4 \frac{\mu_0 g_s \mu_B^2}{8\pi a_0^3 n^3},$$

which is negligible with increasing n . For an electron in the p -shell (i.e., for $\ell = 1$), j can only assume the two values $1 + \frac{1}{2} = \frac{3}{2}$ and $1 - \frac{1}{2} = \frac{1}{2}$. For $j = \frac{3}{2}$, we find $2j + 1 = 4$ degenerate states with an energy that is proportional to $j(j+1) - \ell(\ell+1) - s(s+1) = 1$. For $j = \frac{1}{2}$, there are $2j + 1 = 2$ degenerate states with an energy of $j(j+1) - \ell(\ell+1) - s(s+1) = -2$. Due to $4 \cdot 1 + 2 \cdot (-2) = 0$, the center of the distribution of energy levels is unchanged (see Fig. 8.4). Also, the center of energy levels that are split due to spin-orbit coupling is always the same as without splitting, because the interference comes from within the atom (and not from the environment).

8.8 Problems

- 8.1 Exponential Function of a Pauli Matrix:** Evaluate the exponential function of the Pauli matrix σ_1 .
- 8.2 Eigenvalues and Eigenvectors of the Pauli Matrices:** Evaluate the eigenvalues and eigenvectors of the Pauli matrices σ_1 , σ_2 and σ_3 .
- 8.3 Transformation of the Pauli Matrices to Diagonal Form:** How can you transform the Pauli matrices to diagonal form?
- 8.4 Clebsch–Gordan Coefficients:** What are the Clebsch–Gordan coefficients of two spin- $\frac{1}{2}$ particles?
- 8.5 Clebsch–Gordan Coefficients:** The orbital angular momentum \mathbf{L} and the spin \mathbf{S} of an electron are coupled to the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$. What are the coupled states and the Clebsch–Gordan coefficients?

Chapter 9

Atoms in Electromagnetic Fields

Abstract We examine once again how external magnetic and electric fields affect the energy levels and thus the spectra of atoms. Now, however, we take the spin into account. The Zeeman effect reflects the response to magnetic fields.

9.1 Normal Zeeman Effect

If we apply an external magnetic field to an atom, it will interact with the magnetic dipole moment of the atom. The atomic magnetic dipole moment contains contributions from both the orbital and the spin angular momenta of the electron. The normal Zeeman effect (without spin) has already been discussed in Sect. 8.2. Therefore, two pictures (namely Figs. 9.1 and 9.2) should suffice to remind you of the results.

9.2 Anomalous Zeeman Effect

9.2.1 Weak Field Limit

In fact, the anomalous Zeeman effect is much more common than the normal Zeeman effect. Here, however, the spectral lines are split into more than three lines in a very complicated way, often an even number (quartet, sextet, etc.). To explain this effect, we must consider the spin. This angular momentum s of the electron that amounts to $\frac{1}{2}\hbar$ cannot be explained by classical physics. Although it is only half as large as the unit \hbar of the orbital angular momentum, the spin carries the same strength in terms of the magnetic effect (1 Bohr magneton). In the anomalous Zeeman effect, we thus have to include both the orbital and the spin magnetism.

Let us now calculate the complete magnetic moment. The Hamiltonian for the interaction of the magnetic field \mathbf{b} with the orbital and spin angular momenta \mathbf{L} and \mathbf{S} is

$$\mathbf{H} = -\mu_L \cdot \mathbf{b} - \mu_S \cdot \mathbf{b} = -\gamma(\mathbf{L} + 2\mathbf{S}) \cdot \mathbf{b}, \quad (9.1)$$

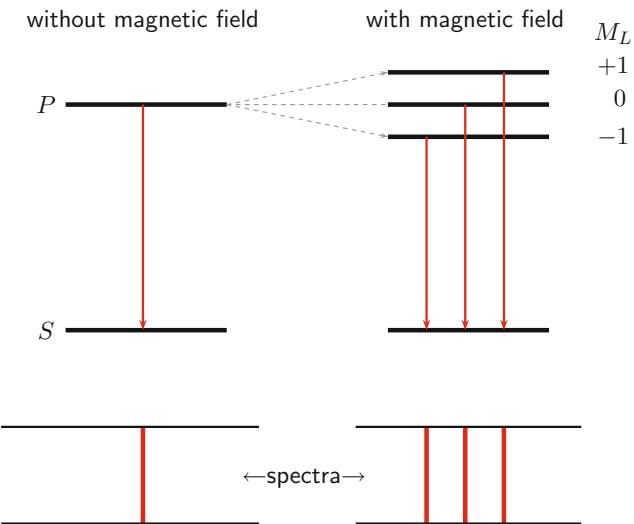


Fig. 9.1 Splitting of spectral lines in the normal Zeeman effect. Photons can be emitted with three different energies

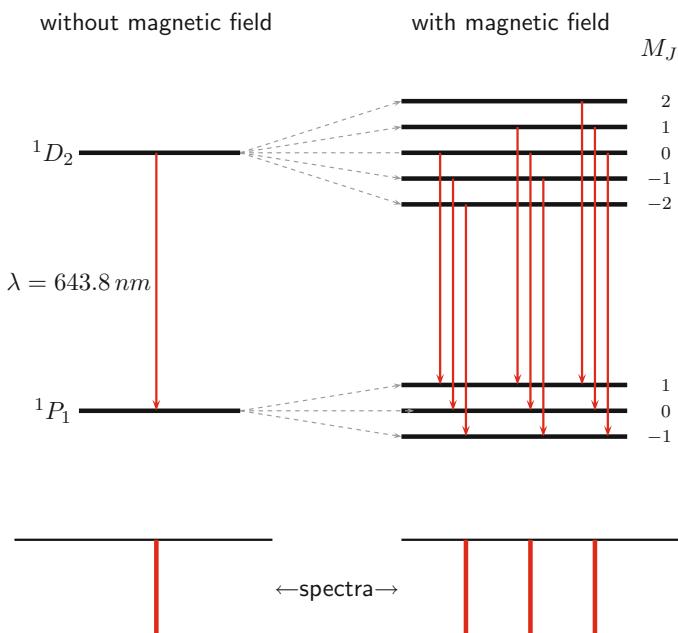
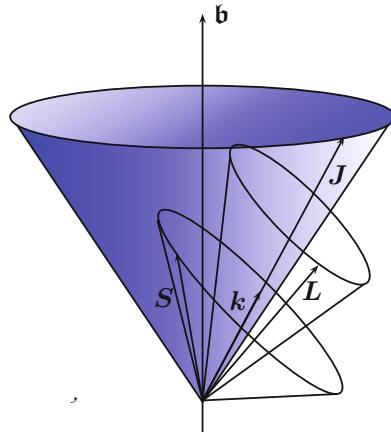


Fig. 9.2 Splitting of spectral lines due to the normal Zeeman effect in cadmium. There are three line triplets (with almost similar wavelengths within the triplet)

Fig. 9.3 Vector diagram for the calculation of the Landé g-factor



where we chose $\gamma = 2$ for the spin. What does this Hamiltonian look like in proportion to \mathbf{J} ? We write

$$\mathbf{H} = -g_J(L, S)\gamma \mathbf{J} \cdot \mathbf{b}, \quad (9.2)$$

where $g_J(L, S)$ is a constant that depends on \mathbf{L} , \mathbf{S} , and \mathbf{J} . Hereby, we assume that the Hamilton functions in (9.1) and (9.2) share at least their values on the diagonals.

In Fig. 9.3, three precession movements can be seen, namely \mathbf{L} around \mathbf{J} , \mathbf{S} around \mathbf{J} , and \mathbf{J} around \mathbf{b} . The effective magnetic moment can be determined by projecting \mathbf{L} onto \mathbf{J} and \mathbf{J} onto \mathbf{b} , and then the same for \mathbf{S} . Let \mathbf{k} be a unit vector pointing in the direction of \mathbf{J} , that is, $\mathbf{k} \stackrel{\text{def}}{=} \mathbf{J}/|\mathbf{J}|$. Then the projections are

$$\mathbf{L} \cdot \mathbf{b} \longrightarrow (\mathbf{L} \cdot \mathbf{k})(\mathbf{k} \cdot \mathbf{b}) = \frac{(\mathbf{L} \cdot \mathbf{J})(\mathbf{J} \cdot \mathbf{b})}{|\mathbf{J}|^2}, \quad (9.3)$$

$$\mathbf{S} \cdot \mathbf{b} \longrightarrow (\mathbf{S} \cdot \mathbf{k})(\mathbf{k} \cdot \mathbf{b}) = \frac{(\mathbf{S} \cdot \mathbf{J})(\mathbf{J} \cdot \mathbf{b})}{|\mathbf{J}|^2}.$$

Because $\mathbf{J} = \mathbf{L} + \mathbf{S}$, it follows that

$$2\mathbf{L} \cdot \mathbf{J} = \mathbf{J}^2 + \mathbf{L}^2 - \mathbf{S}^2 \quad \text{and} \quad 2\mathbf{S} \cdot \mathbf{J} = \mathbf{J}^2 + \mathbf{S}^2 - \mathbf{L}^2.$$

Substituting this in (9.1) and using the quantum mechanical variables (e.g., \mathbf{J}^2 is replaced by $J(J+1)\hbar^2$), we obtain

$$\begin{aligned} \mathbf{H} &= -\gamma(\mathbf{L} + 2\mathbf{S}) \cdot \mathbf{b} = \\ &= -\gamma \left(1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \right) \mathbf{J} \cdot \mathbf{b}. \end{aligned} \quad (9.4)$$

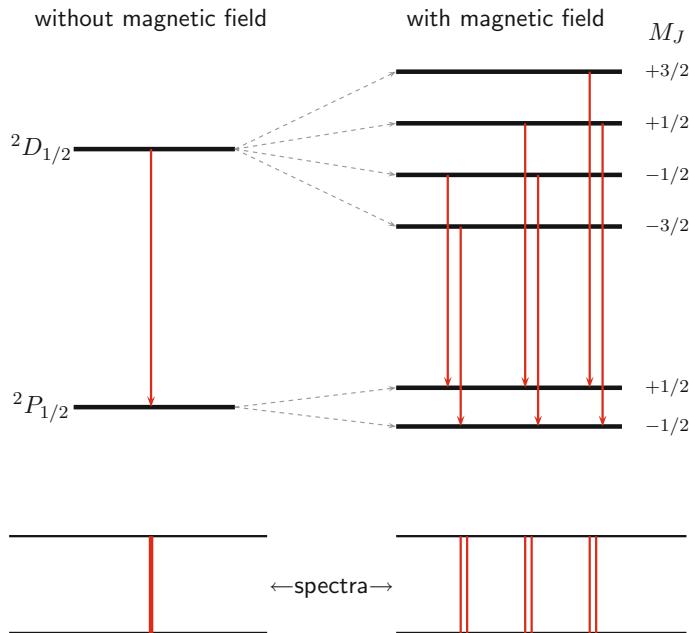


Fig. 9.4 Splitting of spectral lines due to the anomalous Zeeman effect. The various g -values make the spectrum more complicated than in the normal Zeeman effect (Fig. 9.2)

Compared with (9.2), we get the *Landé g-factor*

$$g_J(L, S) \stackrel{\text{def}}{=} 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \quad (9.5)$$

If $S = 0$, then $g_J(L, S) = 1$, because J is then equal to L . In this case, the magnetic moment is independent of L , and we get the normal Zeeman effect as before; that is, all lines are shifted by the same amount. For $S \neq 0$, the value $g_J(L, S)$ depends on the values of L and S ; that is, different lines are shifted by different amounts.

Figure 9.4 shows the anomalous Zeeman effect where the magnetic field is applied to the transition $^2D_{3/2} \rightarrow ^2P_{1/2}$. The Landé g -factor is computed for each level separately, and the level energy is then shifted proportional to its g -value. Hereby, the selection rule¹ $\Delta M_J = 0, \pm 1$ decides which transitions are allowed. For the level of $^2D_{3/2}$, we get $L = 2$, $S = \frac{1}{2}$, and $J = \frac{3}{2}$. This implies $g_{3/2}(2, \frac{1}{2}) = \frac{4}{5}$. For the lower level $^2P_{1/2}$, we find $g_{1/2}(1, \frac{1}{2}) = \frac{2}{3}$. The shift therefore amounts to $\frac{4}{5}\mu_B b$ in $^2D_{3/2}$, and $\frac{2}{3}\mu_B b$ in $^2P_{1/2}$, respectively. The six allowed transitions, namely three doublets, are shown in Fig. 9.4.

¹In physics and chemistry, a selection rule constrains the possible transitions of a system from one quantum state to another.

9.2.2 Strong Magnetic Field

In strong magnetic fields ($B > 1$ Tesla), the coupling of the magnetic moments to the applied field can be stronger than the spin-orbit coupling. In such a case, the total spin S and the total orbital angular momentum L do not couple to J , but precess independently around the axis of the applied magnetic field.

When the magnetic field increases, deviations from the equidistance of the splitting occur in the anomalous Zeeman effect. In fact, some of the individual lines approach each other such that the result resembles the normal Zeeman effect, with only triple splitting. This Paschen-Back effect case is called,^{2,3}

9.3 Problem

9.1 Zeeman effect: Into how many lines does an energy line for $\ell = 2$ split up once a magnetic field is applied?

²Friedrich Paschen, 1865–1947, German physicist.

³Ernst Emil Alexander Back, 1881–1959, German physicist.

Chapter 10

Many Particle Systems

Abstract In this chapter, systems of distinguishable and indistinguishable particles are discussed in detail. Besides, the new concept of the entangled state is introduced. Also, the Pauli principle for the occupation of atomic shells is derived, and its application is discussed with respect to an explanation of the atomic structure.

10.1 Composed Systems

10.1.1 Systems with Two Distinguishable Particles

Let us consider a composite system consisting of two particles which do not interact with each other. The matrix \mathbf{A} describes an observable of the particle 1, and \mathbf{B} an observable of the particle 2. Any behaviour of a particle can therefore be described without reference to the other particle.

We can now prepare a state of the particle 1 in which the observable belonging to \mathbf{A} has a unique value with probability 1. The corresponding state vector is an eigenvector of \mathbf{A} . A similar state vector exists for \mathbf{B} . The probability for the result of a simultaneous measurement of such observables that only relate to one or the other particle must be the product of the individual probabilities. For an overall representation of the system, this strongly suggests that all possible products of the components of the two state vectors occur in a common representation of both particles. Let the two state vectors be \mathbf{a} and \mathbf{b} . We then obtain a vector containing all possible combinations of $a_i b_j$ by using the *Kronecker product* of the two vectors:

$$\mathbf{a} \otimes \mathbf{b} = \begin{pmatrix} a_1 \mathbf{b} \\ a_2 \mathbf{b} \\ a_3 \mathbf{b} \\ \vdots \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \\ a_1 b_3 \\ \vdots \end{pmatrix}.$$

If \mathbf{A} is a $N \times N$ matrix, and \mathbf{B} a $M \times M$ matrix, the state vector \mathbf{a} belongs to the N -dimensional Hilbert space \mathcal{H}_1 , and \mathbf{b} is an element of the M -dimensional Hilbert

space \mathcal{H}_2 . The Kronecker-product $\mathbf{a} \otimes \mathbf{b}$ then belongs to the $N \times M$ -dimensional Hilbert space $\mathcal{H}_{1,2}$, which is spanned by the Kronecker products $\mathbf{u}_A \otimes \mathbf{u}_B$ of the N eigenvectors $\mathbf{u}_A(i)$ of \mathbf{A} and the M eigenvectors $\mathbf{u}_B(j)$ of \mathbf{B} . As an abbreviation for the newly established Hilbert space, we write

$$\mathcal{H}_{1,2} = \mathcal{H}_1 \otimes \mathcal{H}_2.$$

If the set $\{\mathbf{u}_A\}$ of vectors defines an orthonormal basis for \mathcal{H}_1 , and the set $\{\mathbf{u}_B\}$ of vectors is an orthonormal basis for \mathcal{H}_2 , then the set of pairs $\{(\mathbf{u}_A, \mathbf{u}_B)\}$ forms an orthonormal basis for the product space $\mathcal{H}_{1,2}$. For the scalar product on $\mathcal{H}_{1,2}$, we find

$$\begin{aligned} (\mathbf{u}_A(i) \otimes \mathbf{u}_B(m))^\dagger (\mathbf{u}_A(j) \otimes \mathbf{u}_B(n)) &= (\mathbf{u}_A(i)^\dagger \mathbf{u}_A(j)) \otimes (\mathbf{u}_B(m)^\dagger \mathbf{u}_B(n)) = \\ &= (\mathbf{u}_A(i)^\dagger \mathbf{u}_A(j)) \cdot (\mathbf{u}_B(m)^\dagger \mathbf{u}_B(n)). \end{aligned} \quad (10.1)$$

Note that this is a normal product of two numbers, since the Kronecker product of two scalars on the right side of the equation is just an ordinary multiplication.

The expectation value $\langle A \rangle$ for the observable A in a system that is in the state ξ is (as has been defined in (4.5))

$$\langle A \rangle = \xi^\dagger A \xi.$$

For a two-particle system, we can define $\mathbf{A}_{1,2} = \mathbf{A} \otimes \mathbf{B}$ for any two observables \mathbf{A} and \mathbf{B} . If the subsystems are in the states ξ_1 and ξ_2 , respectively (the overall system is therefore in the state $\xi_1 \otimes \xi_2$), the expectation value is given by

$$\begin{aligned} \langle \mathbf{A}_{1,2} \rangle &= (\xi_1 \otimes \xi_2)^\dagger (\mathbf{A} \otimes \mathbf{B}) (\xi_1 \otimes \xi_2) = (\xi_1^\dagger \mathbf{A} \xi_1) \otimes (\xi_2^\dagger \mathbf{B} \xi_2) = \\ &= (\xi_1^\dagger \mathbf{A} \xi_1) (\xi_2^\dagger \mathbf{B} \xi_2) = \langle \mathbf{A} \rangle \langle \mathbf{B} \rangle. \end{aligned} \quad (10.2)$$

In other words, the expectation value of the composite system equals the product of the expectation values of the observables in the subsystems.

One important note has to be made on the matrices in $\mathcal{H}_{1,2}$. In general, we have

$$(\mathbf{A} \otimes \mathbf{B})(\xi_1 \otimes \xi_2) = (\mathbf{A} \xi_1) \otimes (\mathbf{B} \xi_2). \quad (10.3)$$

For $\mathbf{B} = \mathbf{I}_M$, (10.3) would read

$$(\mathbf{A} \otimes \mathbf{I}_M)(\xi_1 \otimes \xi_2) = (\mathbf{A} \xi_1) \otimes \xi_2.$$

If we want to apply the matrices \mathbf{A} or \mathbf{B} (which only act on \mathcal{H}_1 and \mathcal{H}_2) in the composite system, we have to extend their effect to $\mathcal{H}_{1,2}$ by setting

$$\mathbf{A}_{1,2} \stackrel{\text{def}}{=} \mathbf{A} \otimes \mathbf{I}_M \quad \text{and} \quad \mathbf{B}_{1,2} \stackrel{\text{def}}{=} \mathbf{I}_N \otimes \mathbf{B}. \quad (10.4)$$

With this definition, we get indeed

$$\mathbf{A}_{1,2}(\xi_1 \otimes \xi_2) = (\mathbf{A} \xi_1) \otimes \xi_2,$$

or

$$\mathbf{B}_{1,2}\mathbf{A}_{1,2}(\xi_1 \otimes \xi_2) = (\mathbf{A} \xi_1) \otimes (\mathbf{B} \xi_2).$$

One more interesting fact is provided by the following theorem.

Theorem *If \mathbf{A} has the eigenvalues a_i and \mathbf{B} has the eigenvalues b_j , then the so-called Kronecker sum*

$$\mathbf{A} \oplus \mathbf{B} \stackrel{\text{def}}{=} (\mathbf{A} \otimes \mathbf{I}_M) + (\mathbf{I}_N \otimes \mathbf{B}) \in \mathbb{R}^{N \cdot M \times N \cdot M} \quad (10.5)$$

has the eigenvalues

$$a_i + b_j.$$

If \mathbf{a}_i is an eigenvector of \mathbf{A} and \mathbf{b}_j is an eigenvector of \mathbf{B} , then

$$\mathbf{a}_i \otimes \mathbf{b}_j$$

is an eigenvector of $\mathbf{A} \oplus \mathbf{B}$.

Proof $[(\mathbf{A} \otimes \mathbf{I}_M) + (\mathbf{I}_N \otimes \mathbf{B})](\mathbf{a}_i \otimes \mathbf{b}_j) = (\mathbf{A}\mathbf{a}_i \otimes \mathbf{b}_j) + (\mathbf{a}_i \otimes \mathbf{B}\mathbf{b}_j) =$
 $= (\mathbf{a}_i \mathbf{a}_i \otimes \mathbf{b}_j) + (\mathbf{a}_i \otimes \mathbf{b}_j \mathbf{b}_j) =$
 $= \underbrace{(\mathbf{a}_i \otimes 1)}_{a_i}(\mathbf{a}_i \otimes \mathbf{b}_j) + \underbrace{(1 \otimes \mathbf{b}_j)}_{b_j}(\mathbf{a}_i \otimes \mathbf{b}_j) = (a_i + b_j)(\mathbf{a}_i \otimes \mathbf{b}_j)$ **q.e.d.**

10.1.2 Systems with N Distinguishable Subsystems

For more than two subsystems, for example N , the Hilbert space $\mathcal{H}^{(N)}$ is the product space of N single Hilbert spaces:

$$\mathcal{H}^{(N)} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \otimes \cdots \otimes \mathcal{H}_N.$$

A system state is then given by

$$\xi^{(N)} = \xi_1 \otimes \xi_2 \otimes \xi_3 \otimes \cdots \otimes \xi_N.$$

Instead of the matrices \mathbf{A}_i that only act on \mathcal{H}_i , the corresponding matrices $\mathbf{A}_i^{(N)}$ that act on $\mathcal{H}^{(N)}$ have to be used:

$$\mathbf{A}_i^{(N)} = \mathbf{I}_{(1)} \otimes \cdots \otimes \mathbf{I}_{(i-1)} \otimes \mathbf{A}_i \otimes \mathbf{I}_{(i+1)} \otimes \cdots \otimes \mathbf{I}_{(N)}.$$

Then we get in fact

$$\mathbf{A}_i^{(N)} \boldsymbol{\xi}^{(N)} = \boldsymbol{\xi}_1 \otimes \cdots \otimes \boldsymbol{\xi}_{i-1} \otimes \mathbf{A}_i \boldsymbol{\xi}_i \otimes \boldsymbol{\xi}_{i+1} \otimes \cdots \otimes \boldsymbol{\xi}_N,$$

i.e. the operator only acts on $\boldsymbol{\xi}_i$ in the subspace \mathcal{H}_i . Consequently, the expectation value in the composite system is defined by

$$\langle \mathbf{A}^{(N)} \rangle = \langle \mathbf{A}_1 \rangle \langle \mathbf{A}_2 \rangle \cdots \langle \mathbf{A}_N \rangle.$$

In other words: The expectation value for a combination of individual observables is equal to the product of the expectation values for each subsystem observable.

Like in the case of two particles, the extended Kronecker sum

$$\bigoplus_{i=1}^N \mathbf{A}_i = \sum_{i=1}^N \mathbf{A}_i^{(N)}$$

has the eigenvalues

$$\sum_i a_i^{(N)}$$

if the \mathbf{A}_i have the eigenvalues a_i . The proof is very similar to the above theorem for only two matrices \mathbf{A} and \mathbf{B} .

10.1.3 Entangled Systems

Let us again start with two systems in the Hilbert spaces \mathcal{H}_A and \mathcal{H}_B , connected to $\mathcal{H}_A \otimes \mathcal{H}_B$. If the system A is in the state $\boldsymbol{\xi}_A$ and the System B in state $\boldsymbol{\xi}_B$, the state of the composite system is $\boldsymbol{\xi}_A \otimes \boldsymbol{\xi}_B$. Let $\{\mathbf{u}_A(i)\}$ and $\{\mathbf{u}_B(j)\}$ be a orthonormal basis of \mathcal{H}_A and \mathcal{H}_B , respectively. Then the composite state can be rewritten as

$$\boldsymbol{\xi}_A \otimes \boldsymbol{\xi}_B = \left(\sum_i a_i \mathbf{u}_A(i) \right) \otimes \left(\sum_j b_j \mathbf{u}_B(j) \right).$$

States of the composite system that can be represented in such a product form are called *separable states*. Not all possible states in $\mathcal{H}_A \otimes \mathcal{H}_B$ are separable states, however! The most general state in the composite system has the representation

$$\xi_{AB} = \sum_{i,j} c_{ij} (\mathbf{u}_A(i) \otimes \mathbf{u}_B(j)).$$

Obviously this state is only separable if $c_{ij} = a_i \cdot b_j$, yielding $\xi_A = \sum_i a_i \mathbf{u}_A(i)$ and $\xi_B = \sum_j b_j \mathbf{u}_B(j)$. The state is inseparable if any one of the c_{ij} cannot be factored into a product $a_i \cdot b_j$. If a state is not separable, we call it an *entangled state*. E.g., if $\mathbf{u}_A(1)$ and $\mathbf{u}_A(2)$ are two basis vectors in \mathcal{H}_A and $\mathbf{u}_B(1)$ and $\mathbf{u}_B(2)$ two basis vectors in \mathcal{H}_B , the following state is an entangled state:

$$\xi_{AB} = [\mathbf{u}_A(1) \otimes \mathbf{u}_B(2) - \mathbf{u}_A(2) \otimes \mathbf{u}_B(1)] / \sqrt{2}. \quad (10.6)$$

To see this, we write down the general condition for a state in $\mathcal{H}_A \otimes \mathcal{H}_B$ (which is $2 \times 2 = 4$ -dimensional) to be a separable state:

$$\begin{aligned} \xi_{AB} &= c_{11}(\mathbf{u}_A(1) \otimes \mathbf{u}_B(1)) + c_{12}(\mathbf{u}_A(1) \otimes \mathbf{u}_B(2)) + \\ &\quad + c_{21}(\mathbf{u}_A(2) \otimes \mathbf{u}_B(1)) + c_{22}(\mathbf{u}_A(2) \otimes \mathbf{u}_B(2)) \stackrel{!}{=} \\ &= a_1 b_1 (\mathbf{u}_A(1) \otimes \mathbf{u}_B(1)) + a_1 b_2 (\mathbf{u}_A(1) \otimes \mathbf{u}_B(2)) + \\ &\quad + a_2 b_1 (\mathbf{u}_A(2) \otimes \mathbf{u}_B(1)) + a_2 b_2 (\mathbf{u}_A(2) \otimes \mathbf{u}_B(2)). \end{aligned} \quad (10.7)$$

This yields a necessary condition for the state to be separable:

$$c_{11} \cdot c_{22} = c_{12} \cdot c_{21}. \quad (10.8)$$

In the case of (10.6), however, we find

$$c_{12} \cdot c_{21} = -1/2,$$

but

$$c_{11} \cdot c_{22} = 0 \neq -1/2.$$

The state ξ_{AB} in (10.6) is therefore not separable, i.e., the state is entangled. Entangled states are unique to quantum mechanics. They do not exist in classical mechanics and have no classical analogue! In quantum mechanics, however, they are fundamental in modern applications such as *quantum communication* and *quantum computing*.

10.2 Indistinguishable Subsystems

The most prominent examples for indistinguishable, i.e. identical particles are electrons in the atomic shell, or the protons and neutrons in the atomic nucleus. When the structure of the electron shells in an atom and the resulting setup of the periodic table of the elements was derived in the Bohr–Sommerfeld quantum theory, it had just been assumed that the energy levels of an atom cannot be occupied by an arbitrary

number of electrons. Today we know that this is a consequence of the *Pauli principle*, according to which two electrons can never simultaneously occupy one and the same state. As we shall see soon, this principle is a direct consequence of the fundamental indistinguishability of the electrons!

Identical particles have identical particle properties (such as mass, spin, or charge). Of course, the state of each particle may be different and may also change over time. Still, an individual assignment such as

$$\text{Particle } i \Leftrightarrow \text{State } \xi_i$$

is not possible! Instead, only a general assignment

$$\{\text{Set of all particles, } i = 1, \dots, N\} \Leftrightarrow N\text{-particle state } \xi^{(N)}.$$

Here, the overall state is again a composed state, i.e.

$$\xi^{(N)} = \xi_1 \otimes \xi_2 \otimes \cdots \otimes \xi_N. \quad (10.9)$$

However, all expectation values are only allowed to refer to the *totality* of the particles. In other words, they must look like

$$\xi^{(N)\dagger} A_N \xi^{(N)}. \quad (10.10)$$

Moreover, this result must not depend on the order of the sub-states ξ_i in (10.9). This gives us the condition

$$\begin{aligned} & (\xi_1 \otimes \cdots \otimes \xi_i \otimes \cdots \otimes \xi_j \otimes \cdots \otimes \xi_N)^\dagger A_N (\xi_1 \otimes \cdots \otimes \xi_i \otimes \cdots \otimes \xi_j \otimes \cdots \otimes \xi_N) = \\ & (\xi_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes \xi_i \otimes \cdots \otimes \xi_N)^\dagger A_N (\xi_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes \xi_i \otimes \cdots \otimes \xi_N). \end{aligned} \quad (10.11)$$

This condition must hold for any permutation of the state vectors.

Permutations

In Appendix B, a *permutation relationship* is specified as

$$U_{s \times p}(\mathbf{B} \otimes \mathbf{A}) U_{q \times t} = \mathbf{A} \otimes \mathbf{B}, \quad \text{if } \mathbf{A} \in \mathbb{R}^{p \times q} \text{ and } \mathbf{B} \in \mathbb{R}^{s \times t}.$$

In the present case, we have to deal with the interchange of vectors, i.e.

$$U_{s \times p}(\mathbf{b} \otimes \mathbf{a}) U_{1 \times 1} = \underline{\underline{\mathbf{a} \otimes \mathbf{b}}} = \underline{\underline{U_{s \times p}(\mathbf{b} \otimes \mathbf{a})}}, \quad \text{if } \mathbf{a} \in \mathbb{R}^{p \times 1} \text{ and } \mathbf{b} \in \mathbb{R}^{s \times 1}, \quad (10.12)$$

since $U_{1 \times 1} = 1$.

We now derive as an example what the permutation matrix must look like for the case of three state vectors $\xi_1 \otimes \xi_2 \otimes \xi_3$, when the first and third state vector are interchanged. This is done in three steps. Since all subsystems are indistinguishable, we must assume that the state vectors all have the same dimension. In our example, the dimension is 2, i.e., $\xi_i \in \mathbb{R}^2$. In the first step, ξ_1 and ξ_2 are interchanged. This is achieved by multiplying $\xi_1 \otimes \xi_2 \otimes \xi_3$ with the *transposition matrix*¹ $T_{12} = (U_{2 \times 2} \otimes I_2)$, where $U_{2 \times 2} \in \mathbb{R}^{4 \times 4}$ has the form

$$U_{2 \times 2} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right),$$

i.e., it has only one 1 and otherwise only zeros in each row and in each column (like all permutation matrices). The transposition matrix $T_{12} \in \mathbb{R}^{8 \times 8}$ then has the form

$$T_{12} = U_{2 \times 2} \otimes I_2 = \left(\begin{array}{cc|cc} I_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_2 & \mathbf{0} \\ \hline \mathbf{0} & I_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_2 \end{array} \right).$$

In fact, we get

$$\begin{aligned} (U_{2 \times 2} \otimes I_2)(\xi_1 \otimes \xi_2 \otimes \xi_3) &= \left(\begin{array}{cc|cc} I_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_2 & \mathbf{0} \\ \hline \mathbf{0} & I_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_2 \end{array} \right) \left(\begin{array}{c} \xi_{11}\xi_{21}\xi_3 \\ \xi_{11}\xi_{22}\xi_3 \\ \xi_{12}\xi_{21}\xi_3 \\ \xi_{12}\xi_{22}\xi_3 \end{array} \right) = \\ &= \left(\begin{array}{c} \xi_{11}\xi_{21}\xi_3 \\ \xi_{12}\xi_{21}\xi_3 \\ \xi_{11}\xi_{22}\xi_3 \\ \xi_{12}\xi_{22}\xi_3 \end{array} \right) = \left(\begin{array}{c} \xi_{21}(\xi_1 \otimes \xi_3) \\ \xi_{22}(\xi_1 \otimes \xi_3) \end{array} \right) = \xi_2 \otimes \xi_1 \otimes \xi_3. \end{aligned}$$

In a next step, we multiply this intermediate result with $T_{23} = (I_2 \otimes U_{2 \times 2})$, which leads to

$$T_{23}(\xi_2 \otimes \xi_1 \otimes \xi_3) = \xi_2 \otimes \xi_3 \otimes \xi_1.$$

In a third step, we multiply the result with $T_{12} = (U_{2 \times 2} \otimes I_2)$ and finally obtain the desired result

$$T_{12}(\xi_2 \otimes \xi_3 \otimes \xi_1) = \xi_3 \otimes \xi_2 \otimes \xi_1.$$

¹If only two states are exchanged, the permutation is called *transposition matrix*.

In total, the permutation matrix $\mathbf{P}_{1,3}$ has the form

$$\mathbf{P}_{1,3} = \mathbf{T}_{12}\mathbf{T}_{23}\mathbf{T}_{12} = (\mathbf{U}_{2 \times 2} \otimes \mathbf{I}_2)(\mathbf{I}_2 \otimes \mathbf{U}_{2 \times 2})(\mathbf{U}_{2 \times 2} \otimes \mathbf{I}_2) =$$

$$= \left(\begin{array}{c|cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

You can easily check that

$$\mathbf{P}_{1,3}\mathbf{P}_{1,3} = \mathbf{I},$$

which yields

$$\mathbf{P}_{1,3}^{-1} = \mathbf{P}_{1,3}.$$

Of course this result does not come as a surprise! After all, applying the same permutation twice must lead back to the original state.

Transposition and Permutation

As just explained, a transposition matrix² \mathbf{T}_{ij} just swaps the i th and the j th state vector. In an N -particles system, there are $N(N - 1)/2$ different transpositions, and it must hold

$$\mathbf{T}_{ii} = \mathbf{I}, \quad \mathbf{T}_{ij} = \mathbf{T}_{ji}.$$

Each of the N different permutations can be obtained as a product of transpositions. This can be done in various ways. In any case, however, the sign of a permutation is always defined as $+1$ if an even number of transpositions is needed, and -1 for an odd number of transpositions.

Considering (10.10), (10.11) can be rewritten with permutation matrices as

$$\begin{aligned} \boldsymbol{\xi}^{(N)\dagger} \mathbf{A}_N \boldsymbol{\xi}^{(N)} &= (\mathbf{P}_{ij} \boldsymbol{\xi}^{(N)})^\dagger \mathbf{A}_N (\mathbf{P}_{ij} \boldsymbol{\xi}^{(N)}) = \\ &= \boldsymbol{\xi}^{(N)\dagger} \mathbf{P}_{ij}^\dagger \mathbf{A}_N \mathbf{P}_{ij} \boldsymbol{\xi}^{(N)}, \end{aligned} \quad (10.13)$$

which yields the matrix identity

$$\mathbf{A}_N = \mathbf{P}_{ij}^\dagger \mathbf{A}_N \mathbf{P}_{ij}. \quad (10.14)$$

²A transposition is an exchange of two elements in an ordered list with all other elements staying the same, i.e., a permutation of two elements. For example, the swapping of 2 and 5 in the list 123456 yields a transposition 153426 of the list. The *permutation symbol* $\epsilon_{ijk\dots}$ is defined as $(-1)^n$, where n is the number of transpositions that must be applied to build up the permutation.

Furthermore, the scalar product of the permuted state vector must satisfy the condition

$$(\mathbf{P}_{ij}\boldsymbol{\xi}^{(N)})^\dagger(\mathbf{P}_{ij}\boldsymbol{\xi}^{(N)}) = \boldsymbol{\xi}^{(N)\dagger}\boldsymbol{\xi}^{(N)}, \quad (10.15)$$

i.e.,

$$\mathbf{I}_N = \mathbf{P}_{ij}^\dagger \mathbf{P}_{ij}, \quad (10.16)$$

or

$$\mathbf{P}_{ij}^\dagger = \mathbf{P}_{ij}^{-1}. \quad (10.17)$$

In other words, the permutation matrix \mathbf{P}_{ij} is unitary. From (10.14), we then get

$$\mathbf{P}_{ij}\mathbf{A}_N = \mathbf{A}_N\mathbf{P}_{ij},$$

i.e., the \mathbf{P}_{ij} commute with \mathbf{A}_N :

$$\underline{\underline{[\mathbf{A}_N, \mathbf{P}_{ij}] = \mathbf{0}}}. \quad (10.18)$$

What are the eigenvalues of \mathbf{P}_{ij} ? Since \mathbf{P}_{ij} commutes with \mathbf{A}_N , the two matrices (operators) have the same eigenvectors. Naming the eigenvectors \mathbf{e}_{ij} and the eigenvalues λ_{ij} , we get

$$\mathbf{P}_{ij}\mathbf{e}_{ij} = \lambda_{ij}\mathbf{e}_{ij}. \quad (10.19)$$

Since $\mathbf{P}_{ij}^2 = \mathbf{I}$, we have $\lambda_{ij}^2 = 1$. Therefore, the eigenvalues can only be $+1$ or -1 .

10.2.1 Interchanging Two Particles

We consider the example of two spin-1/2 particles, e.g., two electrons. The spin state of an electron is described by a two-dimensional vector in \mathbb{C}^2 . The spin itself is three-dimensional and has three components in x_1 -, x_2 - and x_3 -direction. This means that the information about a three-dimensional real vector is included in a two-dimensional complex state vector. What are the eigenvectors? Using the Pauli matrices σ_i , each of the two particles can be represented according to (8.24) as

$$\mathbf{S}_i = \frac{\hbar}{2}\boldsymbol{\sigma}_i,$$

hence

$$\mathbf{S}_1 = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{S}_2 = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{S}_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

All three matrices have the eigenvalues $+\frac{\hbar}{2}$ and $-\frac{\hbar}{2}$. The eigenvectors, which are called *spinors* in this case, are given by

$$S_1 : \quad e_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}; \quad S_2 : \quad e_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \end{pmatrix};$$

$$S_3 : \quad e_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad e_- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The spins of the two electrons are described by the matrix vectors \mathfrak{S}' and \mathfrak{S}'' , where

$$\mathfrak{S}' = \begin{pmatrix} S'_1 \\ S'_2 \\ S'_3 \end{pmatrix} \quad \text{and} \quad \mathfrak{S}'' = \begin{pmatrix} S''_1 \\ S''_2 \\ S''_3 \end{pmatrix}.$$

For the two spin angular momentum, we have the usual relations

$$[S'_1, S'_2] = i\hbar S'_3 \quad \text{and} \quad [S''_1, S''_2] = i\hbar S''_3.$$

The two sets of matrices S'_i and S''_j , however, commute with each other, since the two particles are assumed to be independent of each other. This implies

$$[S'_i, S''_j] = \mathbf{0}.$$

According to (10.4) and (10.5), we define

$$S_3 = S'_3 \oplus S''_3 = (S'_3 \otimes I_2) + (I_2 \otimes S''_3). \quad (10.20)$$

With

$$S'_3 \otimes I_2 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and

$$I_2 \otimes S''_3 = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

we obtain

$$S_3 = S'_3 \oplus S''_3 = \hbar \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

The eigenvectors of S_3 are

$$e'_i \otimes e''_j,$$

where S'_3 has the eigenvectors e'_i , and S''_3 has the eigenvectors e''_j . The eigenvalues are then the sum of the corresponding eigenvalues, i.e.

$$\lambda'_i + \lambda''_j.$$

With (6.28) and (6.29), we get for the spinors e'_{\pm} of the first electron

$$S'^2 e'_{\pm} = \frac{1}{2} \left(\frac{1}{2} + 1 \right) \hbar^2 e'_{\pm} = \frac{3}{4} \hbar^2 e'_{\pm} \quad (10.21)$$

and

$$S'_3 e'_{\pm} = \pm \frac{\hbar}{2} e'_{\pm}. \quad (10.22)$$

The same holds for the spinors e''_{\pm} of the second electron. The matrix S_3 of the two-particle system then has the four eigenvectors

$$\begin{aligned} e'_+ \otimes e''_+ &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, & e'_+ \otimes e''_- &= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \\ e'_- \otimes e''_+ &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, & e'_- \otimes e''_- &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \end{aligned} \quad (10.23)$$

to which we find the four eigenvalues $+\hbar$, 0 , 0 and $-\hbar$, e.g. by having a look at the main diagonal of the diagonal matrix S_3 .

What subspaces of $\mathcal{H}_{1,2}$ do the eigenvectors span? Let the two electrons be in the states ξ_1 and ξ_2 , respectively, where each ξ_i completely describes the particle's coordinates: $\xi_i = x_i \otimes s_i$ (x_i is the spatial coordinate and s_i the spin). The permutation matrix $P_{1,2} = T_{12}$ is again defined such that the states ξ_1 and ξ_2 of the two particles are exchanged:

$$P_{1,2}(\xi_1 \otimes \xi_2) = \xi_2 \otimes \xi_1. \quad (10.24)$$

Of course, the permutation matrix $P_{1,2}$ has only the two eigenvalues $+1$ and -1 also in this case. The eigenstates with eigenvalue $+1$ are called *symmetric*, and the eigenstates with eigenvalue -1 are called *antisymmetric* states. The symmetric and the antisymmetric states each form a subspace of $\mathcal{H}_{1,2} = \mathcal{H}_1 \otimes \mathcal{H}_2$, since each linear combination of symmetric or antisymmetric states is again symmetric or antisymmetric. Let $\mathcal{H}_{1,2}^{(+)}$ be the symmetric subspace and $\mathcal{H}_{1,2}^{(-)}$ the antisymmetric subspace. It can be easily verified that a state of the form

$$(e_1 \otimes e_2)^{(+)}) \stackrel{\text{def}}{=} \alpha^+(e_1 \otimes e_2 + e_2 \otimes e_1) \quad (10.25)$$

is always symmetrical (with the normalization constant α^+). However, a compound state like

$$(\mathbf{e}_1 \otimes \mathbf{e}_2)^{(-)} \stackrel{\text{def}}{=} \alpha^- (\mathbf{e}_1 \otimes \mathbf{e}_2 - \mathbf{e}_2 \otimes \mathbf{e}_1) \quad (10.26)$$

is always antisymmetric, since in such case we get

$$\mathbf{P}_{12}(\mathbf{e}_1 \otimes \mathbf{e}_2)^{(-)} = -(\mathbf{e}_1 \otimes \mathbf{e}_2)^{(-)}.$$

If we want to generate members of the *symmetric subspace* $\mathcal{H}_{1,2}^{(+)}$ from members of the complete space $\mathcal{H}_{1,2} = \mathcal{H}_1 \otimes \mathcal{H}_2$, we can apply the operator

$$\mathbf{P}_S \stackrel{\text{def}}{=} \mathbf{I} + \mathbf{P}_{12}$$

on $\xi_1 \otimes \xi_2$. Accordingly, the elements of the *antisymmetric subspace* $\mathcal{H}_{1,2}^{(-)}$ are obtained by applying the operator

$$\mathbf{P}_A \stackrel{\text{def}}{=} \mathbf{I} - \mathbf{P}_{12}.$$

As a proof, let us apply \mathbf{P}_S or \mathbf{P}_A on an arbitrary vector $\alpha(\mathbf{e}_1 \otimes \mathbf{e}_2) + \beta(\mathbf{e}_2 \otimes \mathbf{e}_1)$. The resulting vector is

$$\mathbf{P}_S(\alpha(\mathbf{e}_1 \otimes \mathbf{e}_2) + \beta(\mathbf{e}_2 \otimes \mathbf{e}_1)) = \frac{\alpha + \beta}{2}(\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1),$$

or

$$\mathbf{P}_A(\alpha(\mathbf{e}_1 \otimes \mathbf{e}_2) + \beta(\mathbf{e}_2 \otimes \mathbf{e}_1)) = \frac{\alpha + \beta}{2}(\mathbf{e}_1 \otimes \mathbf{e}_2 - \mathbf{e}_2 \otimes \mathbf{e}_1).$$

Note that the following also applies:

$$\begin{aligned} & \frac{1}{2} \left(\frac{1}{\alpha^+} (\mathbf{e}_1 \otimes \mathbf{e}_2)^{(+)} + \frac{1}{\alpha^-} (\mathbf{e}_1 \otimes \mathbf{e}_2)^{(-)} \right) = \\ & = \frac{1}{2} [(\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1) + (\mathbf{e}_1 \otimes \mathbf{e}_2 - \mathbf{e}_2 \otimes \mathbf{e}_1)] = \mathbf{e}_1 \otimes \mathbf{e}_2. \end{aligned} \quad (10.27)$$

In other words, each state of the two-body system can be written as a linear combination of a symmetric and an antisymmetric state. Therefore, the two Hilbert spaces $\mathcal{H}_{1,2}^{(+)}$ and $\mathcal{H}_{1,2}^{(-)}$ span the entire Hilbert space $\mathcal{H}_{1,2}$.

Moreover, symmetric and antisymmetric states are orthogonal to each other, since their scalar product yields $\xi^{(+)} \in \mathcal{H}_{1,2}^{(+)}$ with $\xi^{(-)} \in \mathcal{H}_{1,2}^{(-)}$

$$\xi^{(+)\dagger} \xi^{(-)} = \xi^{(+)\dagger} \underbrace{\mathbf{P}_{12}^\dagger \mathbf{P}_{12}}_{\mathbf{I}} \xi^{(-)} = \xi^{(+)\dagger} (+1)(-1) \xi^{(-)} = -\xi^{(+)\dagger} \xi^{(-)}, \quad (10.28)$$

what can only be true if $\xi^{(+)\dagger} \xi^{(-)} = 0$, i.e., $\xi^{(+)}$ and $\xi^{(-)}$ are orthogonal to each other.

Fermions and Bosons

As we just discussed, a state with exchanged particles is physically indistinguishable from the initial state in a quantum mechanical two-body system of identical particles. After all, we can only measure the expectation value, e.g. $\langle X \rangle = \xi^\dagger X \xi$ for the observable X . If under the permutation of ξ the value changes to $-\xi$, then we still have

$$(-\xi^\dagger)X(-\xi) = \xi^\dagger X \xi = \langle X \rangle.$$

Now experience shows that this factor (+1 or -1) due to the permutation is always either +1 or -1 for particles of a given kind. All particles with *half-integer spin* ($\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$), also called *fermions*,³ always have a factor -1 under exchange. Meanwhile, all particles with *integer spin* (0, 1, 2, ...), also called *bosons*,⁴ always have a factor +1. Mathematically speaking, we get for the states of fermions

$$\xi_{Fermion} \in \mathcal{H}_{1,2}^{(-)},$$

and for bosons

$$\xi_{Boson} \in \mathcal{H}_{1,2}^{(+)}.$$

If we want to describe two indistinguishable particles, it is therefore not required to use the Hilbert space $\mathcal{H}_{1,2}$. Instead, only one of the subspaces $\mathcal{H}_{1,2}^{(-)}$ or $\mathcal{H}_{1,2}^{(+)}$ is enough, depending on whether the particles are fermions or bosons. The vector basis is then given according to (10.25) or (10.26):

$$(\mathbf{e}_1 \otimes \mathbf{e}_2)^{(\mp)} = \frac{1}{\sqrt{2}} (\mathbf{e}_1 \otimes \mathbf{e}_2 \mp \mathbf{e}_2 \otimes \mathbf{e}_1).$$

The atomic building blocks are all made out of fermions:

Leptons: electron e , electron-neutrino ν_e , myon-neutrino ν_μ , tauon-neutrino ν_τ , myon μ , tauon τ ;

Baryons: proton, neutron, Λ -baryon, Σ -baryon, Ω -baryon;

Quarks: up, charm, top (all $s = +\frac{1}{2}$), down, strange, bottom (all $s = -\frac{1}{2}$);

The *Helium-isotopes* 3He ($s = \frac{1}{2}$).

The following particles are bosons:

Mesons: pion, kaon, ρ -meson, ω -meson;

Photons ($s = 1$);

Gauge bosons ($s = 1$);

Gluons ($s = 1$);

The *Helium-isotopes* 4He ($s = 0$).

³Enrico Fermi, 1901–1954, Italian physicist, Nobel Prize 1938.

⁴Satyendra Nath Bose, 1894–1974, Indian physicist.

10.2.2 Interchanging Three Identical Particles

Before we get to the general case with N particles, let us examine the case of 3 identical particles as an intermediate stage. There are three transpositions \mathbf{P}_{12} , \mathbf{P}_{23} and \mathbf{P}_{13} . The symmetric and antisymmetric linear combinations of the $N! = 3! = 6$ states $\mathbf{e}_1^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_3^{(3)}$, $\mathbf{e}_1^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_2^{(3)}$, $\mathbf{e}_2^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_3^{(3)}$, $\mathbf{e}_2^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_1^{(3)}$, $\mathbf{e}_3^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_2^{(3)}$ and $\mathbf{e}_3^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_1^{(3)}$ are given by

$$\begin{aligned} (\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3)^{(+)} &\stackrel{\text{def}}{=} \frac{1}{\sqrt{6}}(\mathbf{e}_1^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_3^{(3)} + \mathbf{e}_1^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_2^{(3)} + \mathbf{e}_2^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_3^{(3)} + \\ &+ \mathbf{e}_2^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_1^{(3)} + \mathbf{e}_3^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_2^{(3)} + \mathbf{e}_3^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_1^{(3)}) \end{aligned} \quad (10.29)$$

and

$$\begin{aligned} (\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3)^{(-)} &\stackrel{\text{def}}{=} \frac{1}{\sqrt{6}}(\mathbf{e}_1^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_3^{(3)} - \mathbf{e}_1^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_2^{(3)} - \mathbf{e}_2^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_3^{(3)} + \\ &+ \mathbf{e}_2^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_1^{(3)} + \mathbf{e}_3^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_2^{(3)} - \mathbf{e}_3^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_1^{(3)}), \end{aligned} \quad (10.30)$$

where $\mathbf{e}_i^{(j)}$ is the i th eigenvector of the j th particle. Both the symmetric vector $(\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3)^{(+)}$ and the antisymmetric vector $(\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3)^{(-)}$ span a *one-dimensional* subspace (also called a *ray*) of the Hilbert space \mathcal{H} . The shape of the antisymmetric vector $(\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3)^{(-)}$ reminds us of the shape of a 3×3 determinant. Remember that for an ordinary determinant we get

$$\det \begin{vmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{vmatrix} =$$

$$= e_{11}e_{22}e_{33} - e_{11}e_{23}e_{32} - e_{12}e_{21}e_{33} + e_{12}e_{23}e_{31} + e_{13}e_{21}e_{32} - e_{13}e_{22}e_{31}.$$

If we replace the entries of the determinant by vectors and the normal multiplication by the Kronecker product (indicated by the index \otimes at the right hand side of the determinant), then we find

$$\det \begin{vmatrix} \mathbf{e}_1^{(1)} & \mathbf{e}_2^{(1)} & \mathbf{e}_3^{(1)} \\ \mathbf{e}_1^{(2)} & \mathbf{e}_2^{(2)} & \mathbf{e}_3^{(2)} \\ \mathbf{e}_1^{(3)} & \mathbf{e}_2^{(3)} & \mathbf{e}_3^{(3)} \end{vmatrix}_{\otimes} \stackrel{\text{def}}{=}$$

$$\begin{aligned} &= \mathbf{e}_1^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_3^{(3)} - \mathbf{e}_1^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_2^{(3)} - \mathbf{e}_2^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_3^{(3)} + \\ &+ \mathbf{e}_2^{(1)} \otimes \mathbf{e}_3^{(2)} \otimes \mathbf{e}_1^{(3)} + \mathbf{e}_3^{(1)} \otimes \mathbf{e}_1^{(2)} \otimes \mathbf{e}_2^{(3)} - \mathbf{e}_3^{(1)} \otimes \mathbf{e}_2^{(2)} \otimes \mathbf{e}_1^{(3)}, \end{aligned}$$

where $\mathbf{e}_i^{(j)}$ is the i th eigenvector of the j th particle. Summing up, we get

$$(\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3)^{(-)} = \frac{1}{\sqrt{6}} \cdot \det \begin{vmatrix} \mathbf{e}_1^{(1)} & \mathbf{e}_2^{(1)} & \mathbf{e}_3^{(1)} \\ \mathbf{e}_1^{(2)} & \mathbf{e}_2^{(2)} & \mathbf{e}_3^{(2)} \\ \mathbf{e}_1^{(3)} & \mathbf{e}_2^{(3)} & \mathbf{e}_3^{(3)} \end{vmatrix}_{\otimes}. \quad (10.31)$$

Each particle is assigned to one row. A transposition of two particles thus corresponds to interchanging two rows of the determinant. From basic linear algebra we know that such an interchange changes the sign of the determinant. Therefore the particle is indeed a fermion! Also, the determinant is equal to zero if two lines are equal. This fact brings us to a crucial theorem in quantum theory.

Pauli Principle Two identical fermions cannot occupy the same quantum state simultaneously.

There is an alternative version of Pauli's principle.

Theorem *The state of an atom is fully described by four quantum numbers n, ℓ, m_J and m_S , and it can be occupied by at most one electron.*

This important principle essentially determines the structure of the periodic table of the elements. Note, however, that the Pauli principle *only applies to fermions, not to bosons* (especially not for atomic nuclei)!

10.2.3 Interchanging N Identical Particles

For a system with N identical particles (think of an atom with N electrons on different shells), the state vector of each state $\xi_k^{(i)}$ (i.e., the k th eigenvector of the i th particle) is composed of the vectors

$$\xi(1, \dots, N) \stackrel{\text{def}}{=} \xi_{i_1}^{(1)} \otimes \xi_{i_2}^{(2)} \otimes \cdots \otimes \xi_{i_N}^{(N)}, \quad (10.32)$$

where i_1, i_2, \dots, i_N are permutations of the numbers $1, 2, \dots, N$. Remember once again that each of the state vectors $\xi^{(i)}$ of the single particle depends on both the location $\mathbf{x}^{(i)}$ and the spin $s^{(i)}$ of this particle. With the *transposition operator* \mathbf{P}_{ij} , which swaps two particles i and j , we get

$$\mathbf{P}_{ij} \xi(1, \dots, i, \dots, j, \dots, N) = \xi(1, \dots, j, \dots, i, \dots, N),$$

and obviously also

$$\mathbf{P}_{ij}^2 \xi(1, \dots, i, \dots, j, \dots, N) = \xi(1, \dots, i, \dots, j, \dots, N).$$

Due to $\mathbf{P}_{ij}^2 = \mathbf{I}$, the eigenvalues are bound to be $+1$. The transposition operator \mathbf{P}_{ij} therefore can only have the eigenvalues $+1$ and -1 .

If a state vector ξ and a permuted state vector ξ' are given, then the expectation values of any observable A must be the same:

$$\langle A \rangle = \xi^\dagger A \xi = \xi'^\dagger A \xi'.$$

We now write formally

$$\xi' = \mathbf{P}\xi,$$

where the permutation operator is given by

$$\mathbf{P} = \begin{pmatrix} 1 & 2 & \cdots & N \\ i_1 & i_2 & \cdots & i_N, \end{pmatrix}$$

i.e., the index i_1 occurs in the first place, the index i_2 in the second place, etc. Remember that all permutations can be assembled from transpositions \mathbf{T}_{ij} . A permutation can be either composed of an even, or of an odd number of transpositions. Accordingly, these permutations are called *even permutations* or *odd permutations*. For example, for three particles we find

$$\text{even : } \mathbf{P} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix};$$

$$\text{odd : } \mathbf{P} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

With these permutations, we can reshape the above requirement for observables as

$$(\mathbf{P}\xi)^\dagger A \mathbf{P}\xi = (\xi)^\dagger \mathbf{P}^\dagger A \mathbf{P}\xi \stackrel{!}{=} (\xi)^\dagger A \xi, \quad \text{for all } \mathbf{P}, \xi.$$

It follows that $\mathbf{P}^\dagger A \mathbf{P} = A$, or with $\mathbf{P}^\dagger = \mathbf{P}^{-1}$

$$A \mathbf{P} = \mathbf{P} A \text{ for all } \mathbf{P}.$$

In other words: In a system of identical particles, only those observables are allowed which commute with all permutations. Such observables are, e.g., the total momentum and the total angular momentum.

Next, we claim that the *symmetric state* $\xi(1, \dots, N)^{(+)}$ of a system can be calculated as

$$\boxed{\xi(1, \dots, N)^{(+)} = \frac{1}{\sqrt{N!}} \sum_{\beta} P_{\beta} (\xi_{i_1}^{(1)} \otimes \xi_{i_2}^{(2)} \otimes \cdots \otimes \xi_{i_N}^{(N)})}. \quad (10.33)$$

P_{β} is the permutation operator which swaps the $\xi^{(i)}$ ($i = 1, \dots, N$), and β numbers the $N!$ permutations of $\xi^{(i)}$. The normalization factor $1/\sqrt{N!}$ normalizes the sum to 1 and ensures

$$\xi^{\dagger}(1, \dots, N)^{(+)} \xi(1, \dots, N)^{(+)} = 1.$$

We shall now prove that there is only one symmetric state $\xi(1, \dots, N)^{(+)}$, i.e., the application of an arbitrary permutation operator P_{α} to $\xi(1, \dots, N)^{(+)}$ returns $\xi(1, \dots, N)^{(+)}$. This is indeed the case, since

$$\begin{aligned} P_{\alpha} \xi(1, \dots, N)^{(+)} &= \frac{1}{\sqrt{N!}} \sum_{\beta} \underbrace{P_{\alpha} P_{\beta}}_{P_{\gamma}} (\xi_{i_1}^{(1)} \otimes \xi_{i_2}^{(2)} \otimes \cdots \otimes \xi_{i_N}^{(N)}) = \\ &= \frac{1}{\sqrt{N!}} \sum_{\gamma} P_{\gamma} (\xi_{i_1}^{(1)} \otimes \xi_{i_2}^{(2)} \otimes \cdots \otimes \xi_{i_N}^{(N)}) = \xi(1, \dots, N)^{(+)}. \end{aligned}$$

Furthermore, we now claim that the *antisymmetric state* of a system $\xi(1, \dots, N)^{(-)}$ can be calculated as

$$\boxed{\xi(1, \dots, N)^{(-)} = \frac{1}{\sqrt{N!}} \sum_{\beta} (-1)^{\pi_{\beta}} P_{\beta} (\xi_{i_1}^{(1)} \otimes \xi_{i_2}^{(2)} \otimes \cdots \otimes \xi_{i_N}^{(N)})}. \quad (10.34)$$

π_{β} denotes the minimum number of transpositions from which P_{β} is constructed. In other words,

$$(-1)^{\pi_{\beta}} = \begin{cases} +1, & \text{if } P_{\beta} \text{ is even} \\ -1, & \text{if } P_{\beta} \text{ is odd.} \end{cases}$$

This formula strongly reminds us of the Leibniz formula (see Appendix F) for $n \times n$ determinants:

$$\det A = \sum_{\beta} \left((-1)^{\pi_{\beta}} \prod_{i=1}^n A_{i, p_i} \right).$$

Therefore, the antisymmetric state (10.34) can also be written with the help of the *Slater determinant*⁵ as

⁵John Clarke Slater, 1900–1976, American physicist.

$$\xi(1, \dots, N)^{(-)} = \frac{1}{\sqrt{N!}} \cdot \det \begin{vmatrix} \xi_1^{(1)} & \xi_2^{(1)} & \cdots & \xi_N^{(1)} \\ \xi_1^{(2)} & \xi_2^{(2)} & \cdots & \xi_N^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1^{(N)} & \xi_2^{(N)} & \cdots & \xi_N^{(N)} \end{vmatrix}_{\otimes}. \quad (10.35)$$

Remember that the symbol \otimes on the right side of the determinant indicates that the Kronecker product is to be used. The determinant's shape immediately reveals its antisymmetry under transpositions. After all, an exchange of two particles corresponds to exchanging two rows of the determinant! Furthermore, the Pauli principle follows immediately, since two equal lines in the determinant render the determinant zero.

The Pauli principle can also be derived by considering an antisymmetric N -particle state in which two single-particle states $\xi^{(i)}$ and $\xi^{(j)}$ are equal. Then

$$\xi_{i_1}^{(1)} \otimes \cdots \otimes \xi_{i_i}^{(i)} \otimes \cdots \otimes \xi_{i_j}^{(j)} \otimes \cdots \otimes \xi_{i_N}^{(N)} = -\xi_{i_1}^{(1)} \otimes \cdots \otimes \xi_{i_j}^{(j)} \otimes \cdots \otimes \xi_{i_i}^{(i)} \otimes \cdots \otimes \xi_{i_N}^{(N)}.$$

It follows immediately that

$$\xi_{i_1}^{(1)} \otimes \cdots \otimes \xi_{i_i}^{(i)} \otimes \cdots \otimes \xi_{i_j}^{(j)} \otimes \cdots \otimes \xi_{i_N}^{(N)} = \mathbf{0}, \quad \text{if } \xi^{(i)} = \xi^{(j)}.$$

Summing up, two identical fermions may never have the same single-particle quantum numbers simultaneously. When filling up available states with fermions, the Pauli principle must always be taken into account! This is how periodic table of the elements arises.

10.3 Problems

10.1 Entangled States: Arrange the four components c_{ij} in a 2×2 matrix and derive the condition (10.8) from it.

10.2 Commutator: In (10.4), we defined the matrices

$$\mathbf{A}_{1,2} = \mathbf{A} \otimes \mathbf{I}_B, \quad \text{and} \quad \mathbf{B}_{1,2} = \mathbf{I}_A \otimes \mathbf{B}.$$

What is the commutator $[\mathbf{A}_{1,2}, \mathbf{B}_{1,2}]$ of these two matrices?

10.3 Expectation Value: In (10.4), we defined the matrices

$$\mathbf{A}_{1,2} = \mathbf{A} \otimes \mathbf{I}_B, \quad \text{and} \quad \mathbf{B}_{1,2} = \mathbf{I}_A \otimes \mathbf{B}.$$

What are the expectation values $\langle \mathbf{A}_{1,2} \rangle$ and $\langle \mathbf{B}_{1,2} \rangle$ of these two observables?

10.4 Permutation Matrix: Which permutation matrix \mathbf{P} swaps two spin states? What is \mathbf{P}^{-1} and \mathbf{P}^\dagger ? What are the eigenvalues of \mathbf{P} ?

Chapter 11

Equivalence of Matrix and Wave Mechanics

Abstract Schrödinger's wave mechanics is briefly introduced, and its equivalence with the matrix formalism of Heisenberg (and others) is demonstrated.

11.1 The De Broglie Wavelength

Schrödinger's wave mechanics is based on the matter-wave theory of de Broglie. For his hypothesis of the wave nature of material particles, de Broglie was inspired by the classical analogy between the mechanical Maupertuis-Euler principle of least action and Fermat's principle on the shortest light path in optics. According to de Broglie's hypothesis, each free particle induces a *wave field* that is linked in such a way that a particle with momentum \mathbf{p} and energy E corresponds to a plane wave propagating in the direction of \mathbf{p} . The wavelength of the particle satisfies the same relationship as for photons, namely

$$p = \frac{h}{\lambda} = \hbar k,$$

that is, the condition

$$\mathbf{k} = \mathbf{p}/\hbar$$

for the wave vector, and for the frequency

$$\omega = E/\hbar, \tag{11.1}$$

where E is the total energy of the particle and \hbar is the reduced Planck quantum of action $\hbar = h/2\pi$.

In general, one can represent a wave as

$$\varphi = A \cos \left[2\pi \left(\frac{r}{\lambda} - \frac{t}{T} \right) \right], \tag{11.2}$$

where r is the distance from the origin, λ denotes the wavelength, and T is the oscillation period. For a wave in three-dimensional space, we need to write

$$\varphi = A \cos\left(\mathbf{k}^\top \mathbf{x} - \frac{2\pi t}{T}\right), \quad (11.3)$$

where the wave vector \mathbf{k} is defined such that it represents the direction of the wave and has the length $|\mathbf{k}| = 2\pi/\lambda$. From a mathematical point of view, a harmonic plane wave in three-dimensional space can be written as

$$A(\mathbf{x}, t) = \operatorname{Re}\left(A_0 e^{i(\mathbf{k}^\top \mathbf{x} - \omega t)}\right) = \operatorname{Re}(A_0) \cos(\mathbf{k}^\top \mathbf{x} - \omega t) - \operatorname{Im}(A_0) \sin(\mathbf{k}^\top \mathbf{x} - \omega t), \quad (11.4)$$

with the complex amplitude A_0 .

11.2 Operators in the Schrödinger Formalism

The *differentiation* $\frac{\partial}{\partial q}$ is an example of a *linear operator* because it has the property

$$\frac{\partial}{\partial q} [f(q) + g(q)] = \frac{\partial}{\partial q} f(q) + \frac{\partial}{\partial q} g(q). \quad (11.5)$$

The same property is given for the *multiplication* with q :

$$q [f(q) + g(q)] = q f(q) + q g(q). \quad (11.6)$$

Let R be a general operator acting on the function $f(q)$ with the resulting function $R f(q)$. If another operator T is applied to $R f(q)$, the result is yet another function $T R f(q)$. The composed operator $T R$ is called the *operator product*. By the way, adding operators follows the simple rule

$$(T + R)f(q) = T f(q) + R f(q).$$

For such operators, there exists an eigenvalue equation as well. However, we now find *eigenfunctions* corresponding to a given eigenvalue (instead of eigenvectors as in the case of matrix operators). For example, the operator $D = \frac{\partial^2}{\partial q^2}$ applied to the function $f(q) = \cos(\omega q)$ yields

$$Df(q) = \frac{\partial^2}{\partial q^2} \cos(\omega q) = -\omega^2 \cos(\omega q) = -\omega^2 f(q).$$

The function $f(q) = \cos(\omega q)$ is thus an eigenfunction to D and $-\omega^2$ is the corresponding eigenvalue.

Operators that are defined on a space of functions can thus be used similarly to matrices. As in the matrix case, however, the commutative law $RT = TR$ generally does not apply! For example, we find that

$$\left(\frac{\partial}{\partial q} q - q \frac{\partial}{\partial q} \right) f(q) = \underbrace{\frac{\partial q}{\partial q}}_1 f + q \frac{\partial f}{\partial q} - q \frac{\partial f}{\partial q} = f(q). \quad (11.7)$$

As an operator equation, this reads

$$\frac{\partial}{\partial q} q - q \frac{\partial}{\partial q} = 1, \quad (11.8)$$

that is, an equation that is similar to the commutation relation for position and momentum in matrix theory.

If we denote the operator “multiplication by q ” with Q , if P is the operation $\frac{h}{2\pi i} \frac{\partial}{\partial q}$, and if the operator “multiplication with 1” is denoted by I , then (11.8) reads

$$P Q - Q P = \frac{h}{2\pi i} I, \quad (11.9)$$

which corresponds exactly to the commutation relation (3.15)

$$P X - X P = \frac{h}{2\pi i} I$$

from matrix mechanics. Formally speaking, there is an “isomorphism” between the operators P , Q and the matrices P , X . For the harmonic oscillator, for example, we get the Hamilton operator (here again with x instead of q)

$$H = \frac{P^2}{2m} + \frac{m\omega^2 X^2}{2} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \frac{m\omega^2 x^2}{2}, \quad (11.10)$$

where m is the particle mass, ω is the angular eigenfrequency of the oscillator, X is the position operator, and $P = -i\hbar \frac{\partial}{\partial x}$ is the momentum operator.

11.3 Schrödinger’s Wave Mechanics

The starting points for the Schrödinger equation are Louis de Broglie’s idea of matter-waves and the Hamilton-Jacobi theory of classical mechanics. The so-called *wave function* ψ can be interpreted as the solution of a linear partial differential equation.

What does the differential equation look like? According to de Broglie, a free particle is always associated with a plane wave. Therefore, it seems plausible to assume the wave function ψ for a particle to be

$$\psi = Ae^{i(\mathbf{k}^T \mathbf{x} - \omega t)}, \quad (11.11)$$

according to (11.4). Equation (11.1) states that the total energy E of the particle is proportional to the frequency ω , namely

$$E = \hbar\omega.$$

The total energy E of the particle consists of kinetic energy E_{kin} and potential energy V :

$$E = E_{kin} + V.$$

A particle of mass m and momentum \mathbf{p} has the kinetic energy

$$E_{kin} = \frac{1}{2} \frac{\mathbf{p}^2}{m}, \quad (11.12)$$

where $\mathbf{p}^2 = \mathbf{p}^T \mathbf{p}$. In general, the potential energy of a particle is a function of the position \mathbf{x} of the particle; that is, $V = V(\mathbf{x})$. Therefore we get

$$E = E_{kin} + V(\mathbf{x}) = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x}). \quad (11.13)$$

Assuming that the particle is represented by the wave function (11.11), we can calculate the derivative of this wave function with respect to time t :

$$\frac{\partial}{\partial t} \psi(t, \mathbf{x}) = -i\omega A e^{i(\mathbf{k}^T \mathbf{x} - \omega t)} = -i\omega \psi(t, \mathbf{x}) = -i \left(\frac{E}{\hbar} \right) \psi(t, \mathbf{x}),$$

or

$$\underline{i\hbar \frac{\partial}{\partial t} \psi(t, \mathbf{x}) = E \psi(t, \mathbf{x})}. \quad (11.14)$$

The derivative of a plane wave (which in a first step depends only on x) with respect to the space variable x is

$$\frac{\partial}{\partial x} \psi(t, x) = ik A e^{i(kx - \omega t)} = \left(\frac{ip}{\hbar} \right) \psi(t, x),$$

therefore

$$\frac{\partial^2}{\partial x^2} \psi(t, x) = - \left(\frac{p}{\hbar} \right)^2 \psi(t, x).$$

Dividing by $2m$ and rearranging finally yields

$$\underline{-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi(t, x)} = \underline{\frac{p^2}{2m} \psi(t, x)} = \underline{E_{kin} \psi(t, x)}. \quad (11.15)$$

If we now insert $E = E_{kin} + V(x)$ into (11.14), we find

$$i\hbar \frac{\partial}{\partial t} \psi(t, x) = (E_{kin} + V(x)) \psi(t, x) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi(t, x) + V(x) \psi(t, x). \quad (11.16)$$

In general, however, the plane wave depends not only on the x -direction, but also on the y - and z -directions, that is, on the full position vector \mathbf{x} of the particle. If we differentiate the wave function (11.11) with respect to the vector components x_ν ($x_1 = x$, $x_2 = y$, $x_3 = z$), we obtain

$$\frac{\partial}{\partial x_\nu} \psi(t, \mathbf{x}) = ik_\nu A e^{i(\mathbf{k}^\top \mathbf{x} - \omega t)} = \left(\frac{ip_\nu}{\hbar} \right) \psi(t, \mathbf{x})$$

and

$$\frac{\partial^2}{\partial x_\nu^2} \psi(t, \mathbf{x}) = ik_\nu^2 A e^{i(\mathbf{k}^\top \mathbf{x} - \omega t)} = \left(\frac{ip_\nu}{\hbar} \right)^2 \psi(t, \mathbf{x}). \quad (11.17)$$

Dividing this equation by $2m$ and rearranging yields

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_\nu^2} \psi(t, \mathbf{x}) = \frac{p_\nu^2}{2m} \psi(t, \mathbf{x}). \quad (11.18)$$

We can add these three equations for $\nu = 1$, $\nu = 2$, and $\nu = 3$ and use the definition

$$\nabla^2 \stackrel{\text{def}}{=} \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2},$$

and we finally obtain

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(t, \mathbf{x}) = \frac{p^2}{2m} \psi(t, \mathbf{x}) = E_{kin} \psi(t, \mathbf{x}). \quad (11.19)$$

Putting $E = E_{kin} + V(t, \mathbf{x})$ into (11.14) similar to the one-dimensional case, one obtains

$$i\hbar \frac{\partial}{\partial t} \psi(t, \mathbf{x}) = (E_{kin} + V(t, \mathbf{x})) \psi(t, \mathbf{x}) = \left(-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x}) \right) \psi(t, \mathbf{x}). \quad (11.20)$$

If we now introduce the *Hamilton operator*

$$H \stackrel{\text{def}}{=} -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x}), \quad (11.21)$$

we obtain *Schrödinger's wave equation*

$i\hbar \frac{\partial}{\partial t} \psi(t, \mathbf{x}) = H\psi(t, \mathbf{x}) .$

(11.22)

Note that we “derived” this equation only for the plane wave. In other words, if a particle is assumed to have a wave function (11.11), it is bound to satisfy the above equation. With this plausibility check in mind, Schrödinger postulated that this equation is valid for *all* quantum mechanical wave functions.

In the Schrödinger equation, we find both the wave function and the Hamilton operator. In Heisenberg’s picture, however, the equations of motion only considered the operators in question. Still, we now show that the two formulations are mathematically equivalent.

11.4 Equivalence of Heisenberg and Schrödinger Pictures

The Schrödinger equation (11.22) is a linear partial differential equation. Solutions can be found by separating the variables, that is, by the ansatz

$$\psi(t, \mathbf{x}) = f(t) \cdot \varphi(\mathbf{x}). \quad (11.23)$$

Substituting (11.23) in (11.22), one obtains

$$i\hbar \varphi(\mathbf{x}) \frac{df(t)}{dt} = \left[-\frac{\hbar^2}{2m} \nabla^2 \varphi(\mathbf{x}) + V(\mathbf{x}) \varphi(\mathbf{x}) \right] f(t). \quad (11.24)$$

Dividing by $f(t) \cdot \varphi(\mathbf{x})$ yields

$$\frac{i\hbar df(t)/dt}{f(t)} = \frac{-(\hbar^2/2m)(\nabla^2 \varphi(\mathbf{x})) + V(\mathbf{x}) \varphi(\mathbf{x})}{\varphi(\mathbf{x})}. \quad (11.25)$$

Because the left side does not depend on \mathbf{x} and the right side does not depend on t , both sides must be equal to the same constant! We can therefore set

$i\hbar \frac{df(t)}{dt} = E f(t)$

(11.26)

and

$$\boxed{-\frac{\hbar^2}{2m} \nabla^2 \varphi(\mathbf{x}) + V(\mathbf{x}) \varphi(\mathbf{x}) = H \varphi(\mathbf{x}) = E \varphi(\mathbf{x})}. \quad (11.27)$$

Equation (11.26) describes the temporal development of $\psi(t, \mathbf{x})$. Equation (11.27) is an eigenvalue equation, which is generally referred to as a *time-independent* or stationary Schrödinger equation. It is a linear, second-order partial differential equation of the variable \mathbf{x} . In contrast, (11.26) is an ordinary linear differential equation of first order, which can be solved by the ansatz

$$f(t) = c \exp(-i E t / \hbar).$$

In the eigenvalue equation for H ,

$$H \varphi_\nu = h_\nu \varphi_\nu,$$

the h_ν are the *eigenvalues* of the operator H , and the functions φ_ν represent the associated *eigenfunctions*. In other words, if we apply the operator H to an eigenfunction, we obtain the same function multiplied by the corresponding eigenvalue. If we consider, for example, the one-dimensional free motion of a particle on the interval $[0, 1]$ of the x -axis, then the Schrödinger equation contains the Hamilton operator

$$H = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2},$$

and the eigenfunctions are

$$\varphi_\nu(x) = \sqrt{2} \sin(\nu \pi x). \quad (11.28)$$

Applying H to those eigenfunctions yields

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \sqrt{2} \sin(\nu \pi x) = \frac{\hbar^2 (\nu \pi)^2}{2m} \sqrt{2} \sin(\nu \pi x),$$

thus the eigenvalues are

$$h_\nu = \frac{\hbar^2 (\nu \pi)^2}{2m}, \nu = 1, 2, 3, \dots$$

For two different eigenvalues h_i and h_j , the corresponding eigenfunctions φ_i and φ_j have the property

$$\int_0^1 \varphi_i^*(x) \varphi_j(x) dx = 0 \quad \text{for } i \neq j; \quad (11.29)$$

that is, they are *orthogonal* to each other. Also, the eigenfunctions $\varphi_i(x)$ are *normalized*, because

$$\int_0^1 \varphi_i^*(x) \varphi_i(x) dx = \int_0^1 |\varphi_i(x)|^2 dx = 1. \quad (11.30)$$

The general property for a set of normalized eigenfunctions is therefore

$$\int_0^1 \varphi_i^*(x) \varphi_j(x) dx = \delta_{ij}. \quad (11.31)$$

Now remember from mathematics that every continuous function $f(x)$ that is also continuously differentiable can be represented by a series with these eigenfunctions:

$$f(x) = \sum_{\nu=1}^{\infty} a_{\nu} \sqrt{2} \sin(\nu \pi x), \quad (11.32)$$

where the coefficients a_{ν} of the series expansion can be calculated according to

$$a_{\nu} = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) f(x) dx. \quad (11.33)$$

We now have all the necessary tools to prove the equivalence of Schrödinger's and Heisenberg's formalism. The proof is carried out in close analogy with the proof of Pasqual Jordan in [14]. First, we claim that the components $(X)_{\nu\mu}$ and $(P)_{\nu\mu}$ of the matrices X and P of Heisenberg's theory can be calculated from the normalized eigenfunctions φ_{ν} as follows.

$$(X)_{\nu\mu} = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) \cdot x \cdot \varphi_{\mu}(x) dx, \quad (11.34)$$

and

$$(P)_{\nu\mu} = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) \cdot \frac{\hbar}{i} \frac{d\varphi_{\mu}(x)}{dx} \cdot dx. \quad (11.35)$$

In other words, if a quantum problem is completely solved in the Schrödinger picture, we can also construct the corresponding matrix solution according to Heisenberg's matrix mechanics. The proof is divided into three steps.

1. First we show that the matrices as defined in (11.34) and (11.35) in fact satisfy the commutation relation (3.15)

$$P X - X P = \frac{\hbar}{i} I. \quad (11.36)$$

For the elements $(\mathbf{P}X)_{\nu\mu}$ of the matrix product $\mathbf{P}X$, we get

$$(\mathbf{P}X)_{\nu\mu} = \sum_{\kappa} (\mathbf{P})_{\nu\kappa} (X)_{\kappa\mu} = \frac{\hbar}{i} \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) \frac{d}{dx} \sum_{\kappa} \varphi_{\kappa}(x) (X)_{\kappa\mu} dx. \quad (11.37)$$

Now the function

$$f(x) \stackrel{\text{def}}{=} x \cdot \varphi_{\mu}(x)$$

can also be written as a series according to (11.32):

$$f(x) = x \cdot \varphi_{\mu}(x) = \sum_{\nu=1}^{\infty} a_{\nu} \varphi_{\nu}(x). \quad (11.38)$$

For the a_{ν} , (11.33) yields

$$a_{\nu} = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) f(x) dx = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) x \cdot \varphi_{\mu}(x) dx.$$

Putting this in (11.38) provides

$$x \cdot \varphi_{\mu}(x) = \sum_{\nu=1}^{\infty} \varphi_{\nu}(x) \int_{-\infty}^{+\infty} \varphi_{\nu}^*(y) y \cdot \varphi_{\mu}(y) dy, \quad (11.39)$$

and with (11.34), we finally get

$$x \cdot \varphi_{\mu}(x) = \sum_{\nu=1}^{\infty} \varphi_{\nu}(x) (X)_{\nu\mu}. \quad (11.40)$$

Thus we find that

$$(\mathbf{P}X)_{\nu\mu} = \frac{\hbar}{i} \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) \frac{d}{dx} x \cdot \varphi_{\mu}(x) dx. \quad (11.41)$$

For the elements $(X\mathbf{P})_{\nu\mu}$ of the matrix product $X\mathbf{P}$, we similarly get with the help of (11.34)

$$(X\mathbf{P})_{\nu\mu} = \sum_{\kappa} (X)_{\nu\kappa} (\mathbf{P})_{\kappa\mu} = \frac{\hbar}{i} \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) x \sum_{\kappa} \varphi_{\kappa}(x) (\mathbf{P})_{\kappa\mu} dx. \quad (11.42)$$

This requires a series expansion of the function

$$g(x) \stackrel{\text{def}}{=} \frac{\hbar}{i} \frac{d}{dx} \varphi(x) = \frac{\hbar}{i} \sum_{\nu} b_{\nu} \varphi_{\nu}(x),$$

with

$$b_\nu = \int_{-\infty}^{+\infty} \varphi_\nu^*(x) \frac{d}{dx} \varphi(x) dx.$$

This leads to

$$g(x) = \frac{\hbar}{i} \frac{d}{dx} \varphi(x) = \frac{\hbar}{i} \sum_\nu \varphi_\nu(x) \int_{-\infty}^{+\infty} \varphi_\nu^*(y) \frac{d}{dy} \varphi(y) dy = \frac{\hbar}{i} \sum_\nu \varphi_\nu(x) (\mathbf{P})_{\nu\mu}. \quad (11.43)$$

Inserting (11.43) into (11.42) finally gives

$$(\mathbf{X}\mathbf{P})_{\nu\mu} = \frac{\hbar}{i} \int_{-\infty}^{+\infty} \varphi_\nu^*(x) x \cdot \frac{d}{dx} \varphi_\mu(x) dx. \quad (11.44)$$

Now we can combine the results (11.41) and (11.44), and we find

$$\begin{aligned} (\mathbf{P}\mathbf{X} - \mathbf{X}\mathbf{P})_{\nu\mu} &= \frac{\hbar}{i} \int_{-\infty}^{+\infty} \varphi_\nu^*(x) \left[\frac{d}{dx} x - x \cdot \frac{d}{dx} \right] \varphi_\mu(x) dx = \\ &= \frac{\hbar}{i} \int_{-\infty}^{+\infty} \varphi_\nu^*(x) \varphi_\mu(x) dx = \frac{\hbar}{i} \delta_{\nu\mu}, \end{aligned}$$

thus we get the commutation relation (11.36) indeed.

2. Now we show that the matrices \mathbf{X} and \mathbf{P} as constructed in (11.34) and (11.35) lead to the following representations of the matrices $\mathbf{U}(\mathbf{X})$ and \mathbf{P}^2 that occur in the Hamilton matrix \mathbf{H} .

$$\mathbf{U}(\mathbf{X})_{\nu\mu} = \int_{-\infty}^{+\infty} \varphi_\nu^*(x) \cdot V(x) \cdot \varphi_\mu(x) dx \quad (11.45)$$

and

$$(\mathbf{P}^2)_{\nu\mu} = \int_{-\infty}^{+\infty} \varphi_\nu^*(x) \cdot -\hbar^2 \frac{d^2 \varphi_\mu(x)}{dx^2} dx. \quad (11.46)$$

Note that $V(x)$ is the operator that corresponds to the matrix $\mathbf{V}(x)$. Equation (11.46) can be proved with the help of (11.35) and (11.43):

$$\begin{aligned} (\mathbf{P}^2)_{\nu\mu} &= \sum_\kappa (\mathbf{P})_{\nu\kappa} (\mathbf{P})_{\kappa\mu} = \int_{-\infty}^{+\infty} \varphi_\nu^*(x) \frac{\hbar}{i} \frac{d}{dx} \sum_\kappa \varphi_\kappa(x) (\mathbf{P})_{\kappa\mu} = \\ &= \int_{-\infty}^{+\infty} \varphi_\nu^*(x) \left(\frac{\hbar}{i} \right)^2 \frac{d^2 \varphi_\mu(x)}{dx^2} dx. \quad (11.47) \end{aligned}$$

Equation (11.45) can be proved by first showing that the assumption is correct for a power $V(X) = X^r$. Because the potential can be written as a series $V(X) = \sum_r c_r X^r$, the assumption then holds for an arbitrary $V(X)$. We prove the simplified claim inductively. For this, let us assume that the assertion is true for $V(X) = X^r$. Then according to (11.34) and (11.40), we indeed get

$$(X^{r+1})_{\nu\mu} = \sum_{\kappa} (X^r)_{\nu\kappa} (X)_{\kappa\mu} = \\ = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) x^r \sum_{\kappa} \varphi_{\kappa}(x) (X)_{\kappa\mu} \cdot dx = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) x^r \cdot x \varphi_{\mu}(x) dx. \quad (11.48)$$

3. It remains to be shown that the matrices X and P as constructed in (11.34) and (11.35) meet the commutation relation (11.36) and have the property

$$H(P, X) = \frac{1}{2m} P^2 + V(X) = \text{diagonal matrix.} \quad (11.49)$$

With (11.45), (11.46), and (11.27) for $\varphi_{\mu}(x)$, we find

$$(H)_{\nu\mu} = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) \left[\frac{-\hbar}{2m} \frac{d^2}{dx^2} \varphi_{\mu}(x) + V(x) \varphi_{\mu}(x) \right] dx = \\ = \int_{-\infty}^{+\infty} \varphi_{\nu}^*(x) E_{\mu} \varphi_{\mu}(x) dx = E_{\mu} \delta_{\nu\mu}. \quad (11.50)$$

Indeed, the eigenvalues of the Hamilton matrix $H(P, X)$ equal the eigenvalues E_{μ} of the Schrödinger equation.

11.5 Example: The Harmonic Oscillator

In accordance with (11.10) and (11.27), the stationary Schrödinger equation for the harmonic oscillator is

$$\left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \frac{m\omega^2 x^2}{2} \right) \varphi_{\nu}(x) = E_{\nu} \varphi_{\nu}(x). \quad (11.51)$$

With

$$q \stackrel{\text{def}}{=} x \sqrt{m\omega/\hbar}, \quad (11.52)$$

we get

$$\frac{d^2\varphi_\nu(q)}{dq^2} + \left(\frac{2E_\nu}{\hbar\omega} - q^2 \right) \varphi_\nu(q) = 0. \quad (11.53)$$

This nonlinear ordinary differential equation of second order is solved by the *Hermite polynomials*

$$H_\nu \stackrel{\text{def}}{=} (-1)^\nu e^{q^2} \frac{d^\nu}{dq^\nu} e^{-q^2}, \quad \nu = 0, 1, 2, \dots \quad (11.54)$$

We therefore obtain the normalized eigenfunctions of (11.53) as

$$\varphi_\nu(q) = (2^\nu \nu! \sqrt{\pi})^{-1/2} e^{-q^2/2} H_\nu(q), \quad \nu = 0, 1, 2, \dots \quad (11.55)$$

The eigenvalues are given by

$$\frac{2E_\nu}{\hbar\omega} = 2\nu + 1. \quad (11.56)$$

Compare this result to the energy values

$$E_\nu = \hbar\omega \left(\nu + \frac{1}{2} \right) \quad (11.57)$$

that we found in Chap. 5 with the help of quantum matrix mechanics.

11.6 Problems

11.1 Schrödinger's Cat: A cat is locked in a steel chamber, along with a machine that the cat cannot manipulate by any means. The machine includes a Geiger counter tube that contains a tiny amount of radioactive substance. The amount is so small that in the course of an hour perhaps one of the atoms decays, but also (with equal probability) perhaps none decays. If an atom decays, the counter tube discharges and releases a hammer that smashes a small bottle with cyanide. If we leave the entire system to itself for an hour, we can predict that the cat is still alive if no atom has decayed in the meantime. However, one single atomic decay would have poisoned it. How do you interpret this experiment from a quantum mechanical perspective?

11.2 Commutators: What are the commutators $[p, x]$, $[p, x^n]$, and $[p^n, x]$ for $p = \frac{\hbar}{i} \frac{\partial}{\partial x}$?

11.3 Differentiation of Operators: Derive the differentiation rules for functions of operators.

Chapter 12

Relativistic Quantum Mechanics

Abstract The basic concepts of the special theory of relativity are provided, followed by Dirac's application of these concepts to quantum mechanics.

12.1 Special Relativity

12.1.1 Four-Dimensional Spacetime

Dirac writes in his book [8]: "Let us now try to make the theory invariant under Lorentz transformations, so that it conforms to the principle of special relativity. This is necessary in order for the theory to apply to particles at high speed." In this section, we present the most important results of the special theory of relativity, as shown, for example, in my book [15]. Note that the components of the vectors that occur here are numbers, not matrices.

The theory of special relativity is based on two postulates:

- The laws of physics are the same for all inertial frames.
- The speed of light c is constant in all inertial frames.

An *inertial frame* is a frame of reference in which Newton's laws apply. In special relativity, one speaks of an *event* if something happens at a time t at some spatial location

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

With the help of the speed of light c , the time t is transformed into a space coordinate as ct . Space and time therefore form the four-dimensional *spacetime*. We denote an event in spacetime by a four-dimensional vector

$$\vec{\mathbf{x}} = \begin{pmatrix} ct \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

One consequence of the second postulate is the *invariance of spacetime intervals*. If a flash of light is emitted at the origin at $t = 0$, the spherical wavefront of the light at some later time Δt is described by

$$c^2 \Delta t^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 = \Delta \mathbf{x}^\top \Delta \mathbf{x},$$

or

$$c^2 \Delta t^2 - \Delta \mathbf{x}^\top \Delta \mathbf{x} = 0. \quad (12.1)$$

With the *Minkowski matrix*

$$\mathbf{M} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and the vector

$$\Delta \vec{\mathbf{x}} = \begin{pmatrix} c \Delta t \\ \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{pmatrix},$$

this can be rewritten as a quadratic form

$$\Delta \vec{\mathbf{x}}^\top \mathbf{M} \Delta \vec{\mathbf{x}} = 0.$$

Due to the second postulate, this equation must also hold for an observer in another frame with the coordinates $\vec{\mathbf{x}}'$:

$$\Delta \vec{\mathbf{x}}'^\top \mathbf{M} \Delta \vec{\mathbf{x}}' = 0.$$

12.1.2 Lorentz Transformation

If an inertial reference frame $\vec{\mathbf{x}}'$ moves with respect to another inertial reference system $\vec{\mathbf{x}}$ with a constant speed \mathbf{v} , then the *Lorentz transformation* connects the two frames (see [8]):

$$\mathbf{L}(\mathbf{v}) \stackrel{\text{def}}{=} \left(\begin{array}{c|c} \gamma & -\frac{\gamma}{c} \mathbf{v}^\top \\ \hline -\frac{\gamma}{c} \mathbf{v} & \mathbf{I} + (\gamma - 1) \frac{\mathbf{v} \mathbf{v}^\top}{v^2} \end{array} \right), \quad (12.2)$$

where the factor γ is given by

$$\gamma = (1 - v^2/c^2)^{-1/2}.$$

The transformation is then given by

$$\vec{x}' = \mathbf{L}(\mathbf{v})\vec{x}.$$

Componentwise, this reads

$$c t' = \gamma c t - \frac{\gamma}{c} \mathbf{v}^\top \mathbf{x},$$

and

$$\mathbf{x}' = \mathbf{x} + (\gamma - 1) \frac{\mathbf{v}^\top \mathbf{x}}{v^2} \mathbf{v} - \gamma \mathbf{v} t. \quad (12.3)$$

Note that these relations assume the coordinate axes of both systems to be parallel.

12.1.3 Velocity and Its Lorentz Transformation

If we define the four-dimensional velocity vector

$$\vec{u} \stackrel{\text{def}}{=} \gamma_u \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \quad (12.4)$$

with

$$\gamma_u \stackrel{\text{def}}{=} \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}},$$

the Lorentz transformed velocity vector \vec{u}' is [15]

$$\vec{u}' = \mathbf{L} \vec{u}, \quad (12.5)$$

with

$$\vec{u}' = \gamma_{u'} \begin{pmatrix} c \\ \mathbf{u}' \end{pmatrix}.$$

Inasmuch as such a velocity vector \vec{u} looks the same in every inertial frame, it is much better suited for the formulation of physical laws. The quadratic form for the velocity is given by

$$\vec{u}^\top \mathbf{M} \vec{u} = \gamma_u^2 c^2 - \gamma_u^2 \mathbf{u}^\top \mathbf{u} = \frac{c^4}{c^2 - u^2} - \frac{c^2 u^2}{c^2 - u^2} = c^2, \quad (12.6)$$

invariant with respect to Lorentz transformations, because $\vec{u}'^\top \mathbf{M} \vec{u}' = c^2$ holds as well.

12.1.4 Momentum and Its Lorentz Transformation

Multiplying (12.5) with the rest mass m_0 , we obtain

$$\begin{pmatrix} m_0 \gamma_{u'} c \\ m_0 \gamma_{u'} \mathbf{u}' \end{pmatrix} = \mathbf{L} \begin{pmatrix} m_0 \gamma_u c \\ m_0 \gamma_u \mathbf{u} \end{pmatrix}. \quad (12.7)$$

We define the *momentum* as usual:

$$\mathbf{p} \stackrel{\text{def}}{=} m_0 \gamma_u \mathbf{u} = m_u \mathbf{u} = m_u \frac{d\mathbf{x}}{dt}, \quad (12.8)$$

where

$$m_u \stackrel{\text{def}}{=} m_0 \gamma_u = \frac{m_0}{\sqrt{1 - \frac{u^2}{c^2}}}.$$

The four-dimensional momentum vector

$$\vec{\mathbf{p}} \stackrel{\text{def}}{=} \begin{pmatrix} m_u c \\ \mathbf{p} \end{pmatrix} = m_0 \vec{\mathbf{u}} = m_0 \gamma_u \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix} \quad (12.9)$$

transforms according to (12.7) as

$$\vec{\mathbf{p}}' = \mathbf{L} \vec{\mathbf{p}}. \quad (12.10)$$

Also, the quadratic form associated with the momentum vector

$$\vec{\mathbf{p}}^\top \mathbf{M} \vec{\mathbf{p}} = m_0^2 \vec{\mathbf{u}}^\top \mathbf{M} \vec{\mathbf{u}} = m_0^2 c^2 \quad (12.11)$$

is invariant with respect to Lorentz transformations, because we also find

$$\vec{\mathbf{p}}'^\top \mathbf{M} \vec{\mathbf{p}}' = m_0^2 c^2.$$

12.1.5 Equation of Motion and Force

The relativistic equation of motion for a particle has to be Lorentz invariant. Additionally, Newton's equation of motion has to hold in the inertial frame of the considered particle:

$$m_0 \frac{d\mathbf{u}}{dt} = \mathbf{f} \in \mathbb{R}^3. \quad (12.12)$$

Let the respective inertial system be \mathcal{X} . Furthermore, suppose that \mathcal{X}' is an inertial system that moves relative to \mathcal{X} with the constant speed $\mathbf{u}(t_0)$. Then the particle rests

momentarily at the time $t = t_0$ in \mathcal{X}' . Now note that the equation of motion (12.12) refers to a point of time and its neighborhood. For the neighborhood $t = t_0 \pm dt$, the speed in \mathcal{X}' is arbitrarily small. For speeds $v \ll c$, (12.12) holds. Hence in \mathcal{X}' ,

$$m_0 \frac{d\mathbf{u}'}{dt'} = \mathbf{f}' \in \mathbb{R}^3 \quad (12.13)$$

also holds exactly. Here, m_0 is the rest mass and \mathbf{f}' the three-dimensional force in \mathcal{X}' . From (12.13), we can derive the relativistic equations of motion in an arbitrary reference frame.

Expanding the vector \mathbf{f}' in (12.13) to a four-vector and calling the result $\vec{\mathbf{f}}'$, we get

$$m_0 \frac{d}{dt'} \begin{pmatrix} c \\ \mathbf{u}' \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{f}' \end{pmatrix} \stackrel{\text{def}}{=} \vec{\mathbf{f}}'. \quad (12.14)$$

Note that $\vec{\mathbf{f}}'$ is specified in the rest system \mathcal{X}' . In the inertial system \mathcal{X} (in which the particle moves with the velocity \mathbf{u}), $\vec{\mathbf{f}}$ is obtained by a Lorentz transformation $\mathbf{L}(-\mathbf{u})$:

$$\vec{\mathbf{f}} = \mathbf{L}(-\mathbf{u}) \begin{pmatrix} 0 \\ \mathbf{f}' \end{pmatrix} = \begin{pmatrix} \frac{\gamma_u}{c} \mathbf{u}^\top \mathbf{f}' \\ \mathbf{A}(\mathbf{u}) \mathbf{f}' \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} f_0 \\ \mathbf{f} \end{pmatrix}, \quad (12.15)$$

where

$$\mathbf{A}(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{I} + (\gamma_u - 1) \frac{\mathbf{u}\mathbf{u}^\top}{u^2}$$

is the known part from the Lorentz matrix. Finally, the equation

$$m_0 \gamma \frac{d}{dt} \begin{pmatrix} \gamma c \\ \gamma \mathbf{u} \end{pmatrix} = \begin{pmatrix} f_0 \\ \mathbf{f} \end{pmatrix};$$

that is,

$$m_0 \vec{\mathbf{a}} = \vec{\mathbf{f}}, \quad (12.16)$$

has all the properties that we need. Namely, the four-vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{f}}$ are Lorentz-invariant and the equation changes into Newton's equation of motion in the rest frame of the particle:

$$m_0 \begin{pmatrix} 0 \\ \frac{d\mathbf{u}'}{dt'} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{f}' \end{pmatrix}.$$

For the last three components of the equation of motion (12.16), we get

$$\frac{d(m_u \mathbf{u})}{dt} = \frac{1}{\gamma_u} \mathbf{f}, \quad (12.17)$$

with the velocity-dependent mass

$$\underline{m_u \stackrel{\text{def}}{=} \gamma_u m_0}. \quad (12.18)$$

In the theory of relativity, the time derivative of the momentum $m_u \mathbf{u}$ is also interpreted as a force. According to (12.15) and (12.16), the components f_i of the relativistic equation of motion are thus

$$\underline{\underline{f_0 = \gamma_u \frac{d}{dt} (m_u c) = \frac{\gamma_u}{c} \mathbf{u}^\top \mathbf{f}'}} \quad (12.19)$$

and

$$\underline{\underline{\mathbf{f} = \gamma_u \frac{d}{dt} (m_u \mathbf{u}) = \mathbf{A}(\mathbf{u}) \mathbf{f}'}} \quad (12.20)$$

12.1.6 Energy and Rest Mass

Equation (12.19) multiplied with c/γ_u provides

$$\underline{\underline{\frac{d}{dt} (m_u c^2) = \mathbf{u}^\top \mathbf{f}}}, \quad (12.21)$$

where $\mathbf{u}^\top \mathbf{f}$ is the instantaneous power, that is, the work per time unit accomplished by the force \mathbf{f} . Therefore, the left-hand side of (12.21) must be the temporal change of energy, which requires that $m_u c^2 = \gamma_u m_0 c^2$ is an energy. For the *relativistic energy*, we obtain the formula

$$\underline{\underline{E = m_u c^2}}. \quad (12.22)$$

For $\mathbf{u} = \mathbf{0}$ (i.e., the particle is at rest), we find $\gamma_u = 1$ and therefore

$$\underline{\underline{E_0 = m_0 c^2}}. \quad (12.23)$$

This is Einstein's famous formula for the “rest energy”.

The four-dimensional momentum vector $\vec{\mathbf{p}}$ can now be recognized as a combination of energy and momentum:

$$\underline{\underline{\vec{\mathbf{p}} = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}}}. \quad (12.24)$$

The quadratic form (12.11),

$$\underline{\underline{\vec{\mathbf{p}}^\top \mathbf{M} \vec{\mathbf{p}} = m_0^2 c^2}},$$

now reads

$$\underline{\underline{\vec{\mathbf{p}}^\top \mathbf{M} \vec{\mathbf{p}} = (E/c, \mathbf{p}^\top) \begin{pmatrix} E/c \\ -\mathbf{p} \end{pmatrix} = E^2/c^2 - p^2 = m_0^2 c^2}},$$

that is, Einstein's *relativistic energy-momentum relationship*

$$\boxed{E^2 = p^2 c^2 + m_0^2 c^4.} \quad (12.25)$$

12.2 The Dirac Equation

12.2.1 The Wave Equation for a Free Particle

In classical mechanics, a moving particle has the energy

$$E = \frac{p^2}{2m}, \quad (12.26)$$

where

$$p^2 = \mathbf{p}^\top \mathbf{p}, \quad \mathbf{p} \in \mathbb{R}^3.$$

On the other hand, we saw in Sect. 11.3 that (12.26) transforms into the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi(t, \mathbf{x}) = -\frac{\hbar^2}{2m} \nabla^2 \psi(t, \mathbf{x}) \quad (12.27)$$

for the wave function $\psi(t, \mathbf{x})$. The temporal and spatial derivatives occur in different orders, therefore the Schrödinger equation cannot be invariant under Lorentz transformations. Rather, its structure changes during transition from one inertial system to another. The Schrödinger equation is a nonrelativistic approximation, for low velocities, of the fully relativistic equation of motion.

Still, let us note the two formal correspondences (substitutions)

$$E \rightarrow i\hbar \frac{\partial}{\partial t} \quad \text{and} \quad \mathbf{p} \rightarrow -i\hbar \nabla \quad (12.28)$$

that transformed (12.26) into (12.27). If we now consider a relativistic quantum particle and assume the same correspondence in the energy-momentum relationship (12.25), we obtain the so-called *Klein–Gordon equation*

$$\boxed{\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \psi(t, \mathbf{x}) = \frac{m_0^2 c^2}{\hbar^2} \psi(t, \mathbf{x})}. \quad (12.29)$$

The Hamiltonian is obtained from the relationship (12.25) as

$$E = H = \pm \sqrt{p^2 c^2 + m_0^2 c^4}, \quad (12.30)$$

where m_0 is the rest mass of the moving particle. However, the minus sign causes a problem. After all, what is a negative energy? At the end of this chapter we learn more about Dirac's interpretation of this negative energy. Still, the solutions of the Klein–Gordon equation can even contain negative probability densities!

Let us first turn to yet another problem, namely how to calculate the square root of an operator. Dirac made the ingenious ansatz

$$\sqrt{p^2c^2 + m_0^2c^4} = c(\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta m_0 c). \quad (12.31)$$

Squaring this ansatz gives

$$\begin{aligned} p^2c^2 + m_0^2c^4 &= c^2(\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta m_0 c)^2 = \\ &= c^2(\alpha_1^2 p_1^2 + \alpha_2^2 p_2^2 + \alpha_3^2 p_3^2 + \beta^2 m_0^2 c^2 + \\ &\quad + \alpha_1 \alpha_2 p_1 p_2 + \alpha_1 \alpha_3 p_1 p_3 + \alpha_2 \alpha_3 p_2 p_3 + \\ &\quad + \alpha_2 \alpha_1 p_2 p_1 + \alpha_3 \alpha_1 p_3 p_1 + \alpha_3 \alpha_2 p_3 p_2 + \\ &\quad + \alpha_1 p_1 \beta m_0 c + \alpha_2 p_2 \beta m_0 c + \alpha_3 p_3 \beta m_0 c + \\ &\quad + \beta m_0 c \alpha_1 p_1 + \beta m_0 c \alpha_2 p_2 + \beta m_0 c \alpha_3 p_3). \end{aligned} \quad (12.32)$$

Under which conditions for the α 's and β is (12.32) valid? The conditions are

$$\alpha_1^2 = \alpha_2^2 = \alpha_3^2 = \beta^2 = 1, \quad (12.33)$$

$$\alpha_1 \alpha_2 + \alpha_2 \alpha_1 = \alpha_1 \alpha_3 + \alpha_3 \alpha_1 = \alpha_2 \alpha_3 + \alpha_3 \alpha_2 = 0, \quad (12.34)$$

$$\alpha_1 \beta + \beta \alpha_1 = \alpha_2 \beta + \beta \alpha_2 = \alpha_3 \beta + \beta \alpha_3 = 0. \quad (12.35)$$

In other words: The α 's all anti-commute with one another and with β , and their square is unity. These properties of the α 's and β suggest that they are not just (real or complex) numbers. Now recall that the Pauli 2×2 matrices satisfy very similar conditions. Therefore, let us assume that the α 's and β are $n \times n$ matrices,¹ where n is initially unknown. We obtain for the energy

$$E = \pm c(\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta m_0 c). \quad (12.36)$$

Inasmuch as the squares of all matrices are equal to the unit matrix, all eigenvalues must be equal to ± 1 . Also, remember that the sum of the diagonal elements of a matrix is equal to the sum of its eigenvalues. Now, however,

¹Written now in boldface letters.

$$\alpha_i = \alpha_i \beta^2 = \alpha \beta \beta = -\beta \alpha \beta,$$

and therefore

$$\text{trace } \alpha_i = \text{trace}(-\beta \alpha_i \beta) = -\text{trace}(\alpha_i \beta^2) = -\text{trace } \alpha_i. \quad (12.37)$$

Note that we exploited the fact that the trace operation is *cyclic*; that is,

$$\text{trace}(\alpha \beta \gamma) = \text{trace}(\beta \gamma \alpha).$$

Also, we used the fact that

$$\text{trace}(a \alpha) = a \text{trace}(\alpha), a \in \mathbb{C}.$$

Equation (12.37) can only be true if $\text{trace } \alpha_i = 0$, which is possible only if the number of +1 eigenvalues is equal to the number of -1 eigenvalues. Hence, n must be an even number. Now how can we make use of the three Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

for finding a solution to our problem? Perhaps finding a fourth matrix that anti-commutes with these three? It turns out, however, that this is impossible. Therefore let us move on to the next even number, namely $n = 4$. In other words, we are now looking for 4×4 matrices that satisfy our conditions (12.33)–(12.35). One possibility for the α 's and β are the four matrices called *Dirac matrices*:

$$\alpha_1 = \sigma_1 \otimes \sigma_1 = \begin{pmatrix} \mathbf{0} & \sigma_1 \\ \sigma_1 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad (12.38)$$

$$\alpha_2 = \sigma_1 \otimes \sigma_2 = \begin{pmatrix} \mathbf{0} & \sigma_2 \\ \sigma_2 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \quad (12.39)$$

$$\alpha_3 = \sigma_1 \otimes \sigma_3 = \begin{pmatrix} \mathbf{0} & \sigma_3 \\ \sigma_3 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad (12.40)$$

$$\beta = \sigma_3 \otimes I_2 = \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (12.41)$$

If we interpret (12.36) as the Hamiltonian and use the known relation $E = i\hbar \frac{\partial}{\partial t} \psi(\mathbf{x}, t)$, we obtain the *Dirac equation*

$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{x}, t) = c (\alpha_1 \mathbf{p}_1 + \alpha_2 \mathbf{p}_2 + \alpha_3 \mathbf{p}_3 + \beta m_0 c) \psi(\mathbf{x}, t).$

(12.42)

Because the matrices α_i and β are 4×4 matrices, the *Dirac wave function* must be a four-column vector, also called a *Dirac spinor*:

$$\psi(\mathbf{x}, t) = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}.$$

Note that the \mathbf{p}_i are also represented as 4×4 matrices

$$\mathbf{p}_i = \begin{pmatrix} p_i & 0 & 0 & 0 \\ 0 & p_i & 0 & 0 \\ 0 & 0 & p_i & 0 \\ 0 & 0 & 0 & p_i \end{pmatrix}.$$

12.2.2 Invariant Form of the Dirac Equation

Multiplying the Dirac equation (12.42) from the left by the matrix β , one obtains with $\beta\beta = I_4$

$$i\hbar \beta \frac{\partial}{\partial t} \psi(\mathbf{x}, t) = c (\beta \alpha_1 \mathbf{p}_1 + \beta \alpha_2 \mathbf{p}_2 + \beta \alpha_3 \mathbf{p}_3 + m_0 c I_4) \psi(\mathbf{x}, t). \quad (12.43)$$

If we define the alternative Dirac matrices

$$\gamma_0 \stackrel{\text{def}}{=} \beta \quad \text{and} \quad \gamma_j \stackrel{\text{def}}{=} \beta \alpha_j \quad (j = 1, 2, 3), \quad (12.44)$$

the Dirac equation looks like

$$i\hbar \frac{\partial}{\partial t} \gamma_0 \psi(\mathbf{x}, t) = c (\gamma_1 \mathbf{p}_1 + \gamma_2 \mathbf{p}_2 + \gamma_3 \mathbf{p}_3 + m_0 c I_4) \psi(\mathbf{x}, t). \quad (12.45)$$

If we set $p_i \stackrel{\text{def}}{=} -i\hbar \frac{\partial}{\partial x_i}$ and $x_0 = c t$, we get the equation

$$\left(i\hbar \sum_{i=0}^3 \gamma_i \frac{\partial}{\partial x_i} - m_0 \mathbf{I}_4 \right) \psi(\mathbf{x}, t) = \mathbf{0}. \quad (12.46)$$

With the Feynman² dash-notation

$$\not{d} \stackrel{\text{def}}{=} \sum_{i=0}^3 \gamma_i \frac{\partial}{\partial x_i},$$

where \not{d} is a 4×4 matrix, we finally get the compact Feynman notation of Dirac's equation:

$$\boxed{(i\not{d} - \frac{m_0}{\hbar} \mathbf{I}_4) \psi(\mathbf{x}, t) = \mathbf{0}}. \quad (12.47)$$

A detailed proof of the invariance of the Dirac equation under Lorentz transformations can be found, for example, in Klaus Schulten's "Notes on Quantum Mechanics" (freely available for download at the website of the University of Illinois).

12.2.3 Solution of the Dirac Equation

The Dirac equation is a system of four linear differential equations. As an ansatz, we try the four wave functions

$$\psi = \theta \exp(i(\mathbf{p}^\top \mathbf{x}/\hbar - \omega t)) \quad (12.48)$$

as eigenfunctions of energy and momentum with the eigenvalues $E = \hbar\omega$, p_1 , p_2 , and p_3 . If we insert the ansatz (12.48) in the Dirac equation (12.42), we obtain the algebraic equation

$$E\theta = c(\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta m_0 c)\theta. \quad (12.49)$$

Due to the block structure of the α -matrices and of β , it is useful to split the Dirac spinor ψ into two two-dimensional spinors χ and η :

$$\psi(\mathbf{x}, t) = \begin{pmatrix} \chi \\ \eta \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}. \quad (12.50)$$

²Richard Phillips Feynman, 1918–1988, American physicist, Nobel Prize 1965.

In the solution (12.48), we similarly split the vector

$$\theta = \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix}. \quad (12.51)$$

Also, we introduce the notation

$$\sigma \cdot \bar{p} \stackrel{\text{def}}{=} \sigma_1 \bar{p}_1 + \sigma_2 \bar{p}_2 + \sigma_3 \bar{p}_3, \quad (12.52)$$

where

$$\bar{p}_i \stackrel{\text{def}}{=} \begin{pmatrix} p_i & 0 \\ 0 & p_i \end{pmatrix}.$$

The algebraic Dirac equation (12.49) now reads

$$E \begin{pmatrix} \xi \\ \zeta \end{pmatrix} = c \begin{pmatrix} \mathbf{0} & \sigma \cdot \bar{p} \\ \sigma \cdot \bar{p} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix} + m_0 c^2 \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix}. \quad (12.53)$$

Rearranging yields

$$(E - m_0 c^2) \xi = c(\sigma \cdot \bar{p}) \zeta, \quad (12.54)$$

$$(E + m_0 c^2) \zeta = c(\sigma \cdot \bar{p}) \xi. \quad (12.55)$$

If we multiply (12.54) with $(E + m_0 c^2)$ and insert (12.55), we obtain

$$(E^2 - m_0^2 c^4) \xi = c^2 (\sigma \cdot \bar{p})^2 \xi. \quad (12.56)$$

For the product $\sigma \cdot \bar{p}$, we can show that

$$(\sigma \cdot \bar{p})^2 = p^2, \quad (12.57)$$

which brings us back to the known condition

$$E^2 = m_0^2 c^4 + c^2 p^2. \quad (12.58)$$

With the definitions

$$p_+ \stackrel{\text{def}}{=} p_1 + i p_2 \quad \text{and} \quad p_- \stackrel{\text{def}}{=} p_1 - i p_2, \quad (12.59)$$

the two (12.54) and (12.55) are then

$$(E - m_0 c^2) \theta_1 = c(p_3 \theta_3 + p_- \theta_4),$$

$$(E - m_0 c^2) \theta_2 = c(p_+ \theta_3 + p_3 \theta_4),$$

$$(E + m_0 c^2) \theta_3 = c(p_3 \theta_1 + p_- \theta_2),$$

$$(E + m_0 c^2) \theta_4 = c(p_+ \theta_1 + p_3 \theta_2).$$

If θ_3 and θ_4 are given, these four equations determine θ_1 and θ_2 (and vice versa). For a given p with $E = +\sqrt{c^2 p^2 + m_0^2 c^4}$, there are therefore two independent solutions of the four equations, namely

$$\begin{pmatrix} 1 \\ 0 \\ \frac{cp_3}{E+m_0 c^2} \\ \frac{cp_+}{E+m_0 c^2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \\ \frac{cp_-}{E+m_0 c^2} \\ \frac{-cp_3}{E+m_0 c^2} \end{pmatrix}. \quad (12.60)$$

These solutions represent the two *spin states* of an electron with the given momentum p , as is physically required. This becomes particularly clear if we consider the nonrelativistic limit $v \ll c$ of the solution. In this case, we get the two vectors (see Chap. 8)

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

12.2.4 Dirac's Interpretation of the Negative Energy

Dirac developed the idea that the states in an atom with negative energies are already occupied with electrons, like the filled electron shells in the Pauli exclusion scheme. If the states with negative energies were unoccupied, all electrons would fall “down” into these states, which would lead to a release of huge amounts of energy. Because this does not happen in reality, these states must already be occupied. Countless numbers of electrons fill the so-called “Dirac sea” of negative energies and make sure that all such states are occupied. The vacuum state is characterized by the fact that all the negative energy states are filled. Based on this idea of occupied negative energy states, Dirac developed his theory of *holes*. According to this theory, an electron that absorbs a photon can make a transition from a negative energy state to a state with positive energy, leaving behind a state with a positive charge that can move

like a particle. Casually speaking, this state is a piece of *anti-matter*. In other words, Dirac predicted the positron! In 1932, Anderson³ finally discovered the positron with a positive electric charge in cosmic rays.

12.3 Problems

12.1 Gamma Matrices: What is the form of the gamma matrices and what relationships exist between them?

12.2 Dirac Equation: Solve the Dirac equation

$$(i\cancel{\partial} - \frac{m_0}{\hbar} \mathbf{I}_4) \psi(\mathbf{x}, t) = \mathbf{0} \quad (12.61)$$

with the ansatz

$$\psi(\mathbf{x}, t) = \mathbf{c} \exp(i(\mathbf{k}^\top \mathbf{x} - \omega t)). \quad (12.62)$$

³Carl David Anderson, 1905–1991, American physicist, Nobel Prize 1936.

Appendix A

Solutions to Problems

- 2.1.** Given the two Hermite matrices $\mathbf{A} = \mathbf{A}^\dagger$ and $\mathbf{B} = \mathbf{B}^\dagger$, we find $(\mathbf{AB})^\dagger = \mathbf{B}^\dagger \mathbf{A}^\dagger = \mathbf{BA}$. This is equal to \mathbf{AB} only if the matrices commute; that is, $\mathbf{AB} = \mathbf{BA}$.
- 2.2.** Let $\mathbf{A}\mathbf{e}_1 = \lambda_1 \mathbf{e}_1$ and $\mathbf{A}\mathbf{e}_2 = \lambda_2 \mathbf{e}_2$ for $\lambda_1 \neq \lambda_2$. Proof by contradiction. We assume that there are $c_1, c_2 \neq 0$ such that

$$c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 = \mathbf{0}. \quad (\text{A.1})$$

Multiplying this equation by \mathbf{A} and taking into account the above eigenvalue equations, we receive

$$c_1 \lambda_1 \mathbf{e}_1 + c_2 \lambda_2 \mathbf{e}_2 = \mathbf{0}. \quad (\text{A.2})$$

If we multiply (A.1) by λ_1 and subtract the result from (A.2), we obtain

$$c_2(\lambda_2 - \lambda_1) \mathbf{e}_2 = \mathbf{0}.$$

Because $\lambda_1 \neq \lambda_2$ and $\mathbf{e}_2 \neq \mathbf{0}$, it must be $c_2 = 0$, which contradicts the assumption.

- 2.3.** An Hermitian matrix satisfies the equation

$$\mathbf{A}\mathbf{e} = \lambda\mathbf{e}.$$

Because $\mathbf{A}^\dagger = \mathbf{A}$, we have

$$(\mathbf{A}\mathbf{e})^\dagger \mathbf{e} = \mathbf{e}^\dagger \underbrace{\mathbf{A}^\dagger \mathbf{e}}_{\lambda^* \mathbf{e}} = \mathbf{e}^\dagger \underbrace{\mathbf{A}\mathbf{e}}_{\lambda \mathbf{e}},$$

or $\lambda^* \mathbf{e}^\dagger \mathbf{e} = \lambda \mathbf{e}^\dagger \mathbf{e}$. Because $\mathbf{e}^\dagger \mathbf{e} > 0$, we have $\lambda^* = \lambda$; that is, the eigenvalue λ is real.

2.4. For unitary matrices, we have $\mathbf{U}^\dagger \mathbf{U} = \mathbf{I}$. With $\mathbf{U}\mathbf{e} = \lambda\mathbf{e}$, we find from

$$(\mathbf{U}\mathbf{e})^\dagger (\mathbf{U}\mathbf{e}) = \mathbf{e}^\dagger \mathbf{U}^\dagger \mathbf{U}\mathbf{e} = \mathbf{e}^\dagger \mathbf{e}$$

that

$$\lambda^* \lambda \mathbf{e}^\dagger \mathbf{e} = \mathbf{e}^\dagger \mathbf{e},$$

thus $|\lambda| = 1$.

2.5. Take \mathbf{a}_1 and construct

$$\mathbf{e}_1 = \frac{\mathbf{a}_1}{\sqrt{\mathbf{a}_1^\dagger \mathbf{a}_1}}.$$

Then we get indeed

$$\mathbf{e}_1^\dagger \mathbf{e}_1 = \frac{\mathbf{a}_1^\dagger \mathbf{a}_1}{\sqrt{\mathbf{a}_1^\dagger \mathbf{a}_1} \sqrt{\mathbf{a}_1^\dagger \mathbf{a}_1}} = 1.$$

Now we subtract the vector $(\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1$ from the vector \mathbf{a}_2 and normalize the result to get

$$\mathbf{e}_2 = \frac{\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1}{\sqrt{(\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)^\dagger (\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)}}.$$

We find again that

$$\mathbf{e}_2^\dagger \mathbf{e}_2 =$$

$$= \frac{(\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)^\dagger (\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)}{(\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)^\dagger (\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)} = 1.$$

Also, \mathbf{e}_2 is orthogonal to \mathbf{e}_1 because of

$$\begin{aligned} \mathbf{e}_1^\dagger \mathbf{e}_2 &= \frac{\mathbf{a}_1^\dagger}{\sqrt{\mathbf{a}_1^\dagger \mathbf{a}_1}} \frac{\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1}{\sqrt{(\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)^\dagger (\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1)}} = \\ &= \frac{\mathbf{a}_1^\dagger \mathbf{a}_2 - \mathbf{a}_1^\dagger (\mathbf{e}_1^\dagger \mathbf{a}_2)\mathbf{e}_1}{\sqrt{\dots}} = \frac{\mathbf{a}_1^\dagger \mathbf{a}_2 - \frac{\mathbf{a}_1^\dagger (\mathbf{a}_1^\dagger \mathbf{a}_2)\mathbf{a}_1}{\mathbf{a}_1^\dagger \mathbf{a}_1}}{\sqrt{\dots}} = 0. \end{aligned}$$

Similarly, we obtain the general formula

$$\mathbf{e}_j = \frac{\mathbf{a}_j - \sum_{i=1}^{j-1} (\mathbf{e}_i^\dagger \mathbf{a}_j)\mathbf{e}_i}{\sqrt{(\mathbf{a}_j - \sum_{l=1}^{j-1} (\mathbf{e}_l^\dagger \mathbf{a}_j)\mathbf{e}_l)^\dagger (\mathbf{a}_j - \sum_{l=1}^{j-1} (\mathbf{e}_l^\dagger \mathbf{a}_j)\mathbf{e}_l)}}.$$

2.6. (a)

$$\mathbf{T} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$\mathbf{T} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

provides the four conditions

$$t_{11} + t_{12} = 1, \quad t_{21} + t_{22} = 0, \quad -t_{11} + t_{12} = 0, \quad -t_{21} + t_{22} = 1.$$

The transformation matrix is therefore given by

$$\mathbf{T} = \underbrace{\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}}_{\underline{\underline{\mathbf{T}}}}.$$

Because of $\mathbf{T}^\dagger \mathbf{T} = \mathbf{I}$, \mathbf{T} is a unitary matrix.

(b) According to Problem 2.5, we get

$$\underline{\underline{\mathbf{e}_1}} = \frac{\underline{\underline{\mathbf{a}_1}}}{\sqrt{\mathbf{a}_1^\dagger \mathbf{a}_1}} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

and

$$\begin{aligned} \underline{\underline{\mathbf{e}_2}} &= \frac{\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2) \mathbf{e}_1}{\sqrt{(\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2) \mathbf{e}_1)^\dagger (\mathbf{a}_2 - (\mathbf{e}_1^\dagger \mathbf{a}_2) \mathbf{e}_1)}} = \\ &= \frac{\begin{pmatrix} -1 \\ 1 \end{pmatrix} - 0 \cdot \mathbf{e}_1}{\sqrt{2}} = \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}. \end{aligned}$$

3.1. We have

$$\begin{aligned} &[\mathbf{A}, [\mathbf{B}, \mathbf{C}]] + [\mathbf{B}, [\mathbf{C}, \mathbf{A}]] + [\mathbf{C}, [\mathbf{A}, \mathbf{B}]] = \\ &= [\mathbf{A}, \mathbf{BC} - \mathbf{CB}] + [\mathbf{B}, \mathbf{CA} - \mathbf{AC}] + [\mathbf{C}, \mathbf{AB} - \mathbf{BA}] \\ &= \mathbf{ABC} - \mathbf{ACB} - \mathbf{BCA} + \mathbf{CBA} + \\ &\quad + \mathbf{BCA} - \mathbf{BAC} - \mathbf{CAB} + \mathbf{ACB} + \\ &\quad + \mathbf{CAB} - \mathbf{CBA} - \mathbf{ABC} + \mathbf{BAC} \\ &= \underline{\underline{0}}. \end{aligned}$$

3.2. If we multiply the relation

$$[X, \mathbf{P}^n] = X\mathbf{P}^n - \mathbf{P}^nX = n i\hbar \mathbf{P}^{n-1}$$

with \mathbf{P} from the left, we find

$$\mathbf{P}X\mathbf{P}^n - \mathbf{P}^{n+1}X = n i\hbar \mathbf{P}^n. \quad (\text{A.3})$$

From

$$[X, \mathbf{P}] = i\hbar \mathbf{I}$$

we get

$$\mathbf{P}X = X\mathbf{P} - i\hbar \mathbf{I}.$$

A substitution in (A.3) finally yields the assertion for $n + 1$:

$$X\mathbf{P}^{n+1} - \mathbf{P}^{n+1}X = (n + 1)i\hbar \mathbf{P}^n.$$

3.3. Because $(AB)^\dagger = B^\dagger A^\dagger$ and $(A + B)^\dagger = A^\dagger + B^\dagger$, we find

$$(AB + BA)^\dagger = B^\dagger A^\dagger + A^\dagger B^\dagger = AB + BA.$$

3.4. With

$$\mathbf{H} = \mathbf{E} = \begin{pmatrix} E_1 & 0 & \cdots & \cdots \\ 0 & E_2 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & E_N \end{pmatrix},$$

we can write

$$\begin{aligned} X(t) &= \exp(-it\mathbf{E}/\hbar)X(0)\exp(it\mathbf{E}/\hbar) = \\ &= \begin{pmatrix} e^{(-itE_1/\hbar)} & 0 & \cdots & 0 \\ 0 & e^{(-itE_2/\hbar)} & 0 & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & e^{(-itE_N/\hbar)} \end{pmatrix} X(0) \exp(it\mathbf{E}/\hbar) = \\ &= \begin{pmatrix} X_{11}(0) & X_{12}(0)e^{(-it(E_1-E_2)/\hbar)} & \cdots & X_{1N}(0)e^{(-it(E_1-E_N)/\hbar)} \\ \vdots & X_{22}(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1}(0)e^{(-it(E_N-E_1)/\hbar)} & \cdots & \cdots & X_{NN}(0) \end{pmatrix}. \end{aligned}$$

This implies

$$\underline{X_{\mu\nu}(t) = e^{(it(E_\nu - E_\mu)/\hbar)} X_{\mu\nu}(0)}.$$

3.5. The matrix $\exp(i \mathbf{H})$ is a power series in \mathbf{H} , therefore we can easily show that

$$(\exp(i \mathbf{H}))^\dagger = \exp(-i \mathbf{H}^\dagger) = \exp(-i \mathbf{H}).$$

This leads to

$$(\exp(i \mathbf{H}))^\dagger (\exp(i \mathbf{H})) = \exp(-i \mathbf{H}) (\exp(i \mathbf{H})) = \exp(\mathbf{0}) = \mathbf{I}.$$

In other words, $\exp(i \mathbf{H})$ is indeed a unitary matrix.

3.6. This question is equivalent to the question of whether $-\hbar \mathbf{C} = i[\mathbf{AB} - \mathbf{BA}]$ is Hermitian. We find that

$$\begin{aligned} (i[\mathbf{A}, \mathbf{B}])^\dagger &= -i([\mathbf{AB} - \mathbf{BA}])^\dagger \\ &= -i((\mathbf{AB})^\dagger - (\mathbf{BA})^\dagger) \\ &= -i\mathbf{B}^\dagger \mathbf{A}^\dagger + i\mathbf{A}^\dagger \mathbf{B}^\dagger \\ &= -i\mathbf{BA} + i\mathbf{AB} \\ &= i[\mathbf{A}, \mathbf{B}]. \end{aligned} \tag{A.4}$$

Therefore, $-\hbar \mathbf{C}$ and thus $\hbar \mathbf{C} = \frac{1}{i}[\mathbf{A}, \mathbf{B}]$ are indeed Hermitian. Note, however, that $[\mathbf{A}, \mathbf{B}]$ is *anti-hermitian*!

3.7.

$$(a) [\mathbf{A}, \mathbf{BC}] = \mathbf{ABC} - \mathbf{BCA} + \underbrace{\mathbf{BAC} - \mathbf{BAC}}_0 = \mathbf{B}[\mathbf{A}, \mathbf{C}] + [\mathbf{A}, \mathbf{B}]\mathbf{C}.$$

$$(b) [\mathbf{AB}, \mathbf{C}] = \mathbf{ABC} - \mathbf{CAB} + \underbrace{\mathbf{ACB} - \mathbf{ACB}}_0 = \mathbf{A}[[\mathbf{B}, \mathbf{C}] + [\mathbf{A}, \mathbf{C}]\mathbf{B}].$$

3.8. A nilpotent matrix is a square matrix N such that $N^k = \mathbf{0}$ for some positive integer k . In the present case, we find

$$N^2 = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad N^3 = \mathbf{0},$$

which leads to

$$\exp(tN) = \mathbf{I} + tN + \frac{t^2}{2}N^2 = \begin{pmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ 0 & 0 & 1 \end{pmatrix}.$$

- 4.1.** The momentum of the electron is

$$p = m_e \cdot v = (9.11 \cdot 10^{-31} \text{ kg})(10^3 \text{ ms}^{-1}) = 9.11 \cdot 10^{-28} \text{ m kg s}^{-1}.$$

The percentage of the momentum accuracy is

$$\frac{\Delta p}{p} \cdot 100 = 0.1,$$

therefore

$$\Delta p = \frac{p \cdot 0.1}{100} = 9.11 \cdot 10^{-31} \text{ m kg s}^{-1}.$$

Heisenberg's uncertainty principle then yields

$$\Delta x \geq \frac{\hbar}{2\Delta p} = \frac{1.055 \cdot 10^{-34} \text{ Js}}{2 \cdot 9.11 \cdot 10^{-31} \text{ kg m s}^{-1}} = 0.0579 \cdot 10^{-3} \text{ m} = \underline{\underline{0.0579 \text{ mm}}}.$$

- 4.2.** A matrix is a projection matrix if (1) \mathbf{P} is Hermitian and (2) $\mathbf{P}^2 = \mathbf{P}$. Let us check these conditions for the product matrix.

(1) Because \mathbf{P}_1 and \mathbf{P}_2 are projection matrices, they are Hermitian; that is, $\mathbf{P}_1 = \mathbf{P}_1^\dagger$ and $\mathbf{P}_2 = \mathbf{P}_2^\dagger$. Therefore

$$(\mathbf{P}_1 \mathbf{P}_2)^\dagger = \mathbf{P}_2^\dagger \mathbf{P}_1^\dagger = \mathbf{P}_2 \cdot \mathbf{P}_1.$$

A necessary condition for the product matrix being Hermitian is therefore that $\mathbf{P}_1 \cdot \mathbf{P}_2 = \mathbf{P}_2 \cdot \mathbf{P}_1$; that is,

$$[\mathbf{P}_1, \mathbf{P}_2] = \mathbf{0}.$$

In other words, \mathbf{P}_1 and \mathbf{P}_2 must commute.

(2) If \mathbf{P}_1 and \mathbf{P}_2 commute, we have

$$(\mathbf{P}_1 \mathbf{P}_2)^2 = \mathbf{P}_1 \mathbf{P}_2 \mathbf{P}_1 \mathbf{P}_2 = \mathbf{P}_1 (\mathbf{P}_2 \mathbf{P}_1) \mathbf{P}_2 = \mathbf{P}_1 (\mathbf{P}_1 \mathbf{P}_2) \mathbf{P}_2 = \mathbf{P}_1^2 \mathbf{P}_2^2 = \mathbf{P}_1 \mathbf{P}_2.$$

We see that the second condition is automatically fulfilled if the two projection matrices commute.

- 4.3.** (1) The density matrix is

$$\mathbf{D} = \mathbf{e}_{3,1} \mathbf{e}_{3,1}^\dagger = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and it is indeed $\mathbf{D}^2 = \mathbf{D}$.

(2) In this case, the density matrix is

$$\mathbf{D} = \mathbf{e}_{2,1}\mathbf{e}_{2,1}^\dagger = \frac{1}{2} \begin{pmatrix} 1 \\ i \end{pmatrix} (1 - i) = \begin{pmatrix} 0.5 & -0.5i \\ 0.5i & 0.5 \end{pmatrix},$$

and it is again $\mathbf{D}^2 = \mathbf{D}$.

In both cases $\text{trace}(\mathbf{D}) = 1$, therefore both systems are in a pure state.

4.4. Because

$$\mathbf{M}\mathbf{e}_1 = (\mathbf{e}_1\mathbf{e}_1^\top - \mathbf{e}_2\mathbf{e}_2^\top)\mathbf{e}_1 = \mathbf{e}_1,$$

and

$$\mathbf{M}\mathbf{e}_2 = (\mathbf{e}_1\mathbf{e}_1^\top - \mathbf{e}_2\mathbf{e}_2^\top)\mathbf{e}_2 = -\mathbf{e}_2,$$

the eigenvectors are \mathbf{e}_1 and \mathbf{e}_2 , and the eigenvalues are $+1$ and -1 .

5.1. With

$$\mathbf{X} = \sqrt{\frac{\hbar}{2m\omega_0}} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ 1 & 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

and

$$\mathbf{P} = i\sqrt{\frac{\hbar m\omega_0}{2}} \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & \dots \\ 1 & 0 & -\sqrt{2} & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & -\sqrt{3} & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & -\sqrt{4} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

we get

$$\mathbf{X}\mathbf{P} = i\frac{\hbar}{2} \begin{pmatrix} 1 & 0 & -\sqrt{2} & 0 & 0 & \dots \\ 0 & 1 & 0 & -\sqrt{6} & 0 & \dots \\ \sqrt{2} & 0 & 1 & 0 & -\sqrt{12} & \dots \\ 0 & \sqrt{6} & 0 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

and

$$\mathbf{P}\mathbf{X} = i\frac{\hbar}{2} \begin{pmatrix} -1 & 0 & -\sqrt{2} & 0 & 0 & \dots \\ 0 & -1 & 0 & -\sqrt{6} & 0 & \dots \\ \sqrt{2} & 0 & -1 & 0 & -\sqrt{12} & \dots \\ 0 & \sqrt{6} & 0 & -1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix};$$

that is, indeed

$$\mathbf{X}\mathbf{P} - \mathbf{P}\mathbf{X} = i\hbar\mathbf{I}.$$

5.2. We have

$$\tilde{X} = \sqrt{\frac{m\omega_0}{2\hbar}} X$$

and

$$\tilde{P} = \sqrt{\frac{1}{2m\omega_0\hbar}} P.$$

\hbar has the dimension of an action ($M \cdot L^2 \cdot T^{-1}$). The factor $\sqrt{\frac{m\omega_0}{2\hbar}}$ therefore has the dimension L^{-1} . Because X has the dimension of a length L , the first term of A is dimensionless. We find the same result for the second term, therefore A is dimensionless.

5.3. A is not Hermitian, inasmuch as

$$\begin{aligned} A^\dagger &= \frac{1}{\sqrt{2\hbar}} \left(\sqrt{m\omega_0} X^\dagger - \frac{i}{\sqrt{m\omega_0}} P^\dagger \right) \\ &= \frac{1}{\sqrt{2\hbar}} \left(\sqrt{m\omega_0} X - \frac{i}{\sqrt{m\omega_0}} P \right) \\ &= \tilde{X} - i \tilde{P} \\ &\neq A. \end{aligned}$$

Therefore, A cannot be an observable.

5.4. Because

$$N^\dagger = (A^\dagger A)^\dagger = A^\dagger A = N,$$

N is Hermitian.

5.5. We prove the formula by induction.

1. $n = 1$: Is $[A, A^\dagger A] = A$? According to Problem 3.7,

$$[A, BC] = B[A, C] + [A, B]C.$$

Together with $[A, A^\dagger] = I$ from (6.10), we get

$$[A, A^\dagger A] = A^\dagger \underbrace{[A, A]}_0 + \underbrace{[A, A^\dagger]}_I A = A.$$

2. $n - 1 \rightarrow n$: Does $[A^n, N] = nA^n$ follow from

$$[A^{n-1}, N] = (n-1)A^{n-1}?$$

Again according to Problem 3.7,

$$[AB, C] = A[B, C] + [A, C]B.$$

We can apply the above induction assumption to find

$$\begin{aligned} [A^n, N] &= A[A^{n-1}, N] + [A, N]A^{n-1} = A(n-1)A^{n-1} + (1 \cdot A^1)A^{n-1} = \\ &= (n-1)A^n + A^n = nA^n. \end{aligned}$$

5.6. With (5.32) and (5.33), we get

$$\begin{aligned} N = A^\dagger A &= \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \\ 0 & \sqrt{2} & 0 & 0 & \\ 0 & 0 & \sqrt{3} & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{2} & 0 & \\ 0 & 0 & 0 & \sqrt{3} & \ddots \\ \vdots & \ddots & \ddots & \ddots & \end{pmatrix} = \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \\ 0 & 0 & 2 & 0 & \\ 0 & 0 & 0 & 3 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \end{pmatrix}. \end{aligned}$$

The eigenvalues are the numbers at the main diagonal, namely 0, 1, 2, 3,

5.7. The eigenvalue equations (5.29) and (5.30) are

$$A\mathbf{e}_n = \sqrt{n}\mathbf{e}_{n-1}$$

and

$$A^\dagger \mathbf{e}_n = \sqrt{n+1}\mathbf{e}_{n+1}.$$

With these equations, (5.34) supplies

$$X\mathbf{e}_n = \sqrt{\frac{\hbar}{2m\omega_0}}(A + A^\dagger)\mathbf{e}_n = \sqrt{\frac{\hbar}{2m\omega_0}}(\sqrt{n}\mathbf{e}_{n-1} + \sqrt{n+1}\mathbf{e}_{n+1}). \quad (\text{A.5})$$

Multiplying by the transposed eigenvector \mathbf{e}_m from the left gives the matrix element in the mth row and nth column of X (with $\mathbf{e}_m \mathbf{e}_n = \delta_{mn}$):

$$X_{mn} = \mathbf{e}_m^\top X \mathbf{e}_n = \sqrt{\frac{\hbar}{2m\omega_0}}(\sqrt{n}\delta_{m,n-1} + \sqrt{n+1}\delta_{m,n+1}).$$

Multiplying (A.5) with X yields

$$X^2 \mathbf{e}_n = \sqrt{\frac{\hbar}{2m\omega_0}} (\sqrt{n} X \mathbf{e}_{n-1} + \sqrt{n+1} X \mathbf{e}_{n+1}), \quad (\text{A.6})$$

and with (5.29) and (5.30) we get

$$\begin{aligned} X^2 \mathbf{e}_n &= \frac{\hbar}{2m\omega_0} (\sqrt{n}(A + A^\dagger) \mathbf{e}_{n-1} + \sqrt{n+1}(A + A^\dagger) \mathbf{e}_{n+1}) = \\ &= \frac{\hbar}{2m\omega_0} (\sqrt{n(n-1)} \mathbf{e}_{n-2} + (2n+1) \mathbf{e}_n + \sqrt{(n+1)(n+2)} \mathbf{e}_{n+2}). \end{aligned}$$

Multiplying this equation also by the transposed eigenvector \mathbf{e}_m from the left gives the matrix element in the m th row and n th column of X^2 :

$$X_{mn}^2 = \frac{\hbar}{2m\omega_0} (\sqrt{n(n-1)} \delta_{m,n-2} + (2n+1) \delta_{m,n} + \sqrt{(n+1)(n+2)} \delta_{m,n+2}).$$

In the same way, we obtain

$$\begin{aligned} X_{mn}^3 &= \left(\frac{\hbar}{2m\omega_0} \right)^{3/2} \left(\sqrt{n(n-1)(n-2)} \delta_{m,n-3} + 3\sqrt{(n+1)^3} \delta_{m,n+1} + \right. \\ &\quad \left. + 3\sqrt{n^3} \delta_{m,n-1} + \sqrt{(n+1)(n+2)(n+3)} \delta_{m,n+3} \right). \end{aligned}$$

The matrix X^3 therefore has the form

$$X^3 = \left(\frac{\hbar}{2m\omega_0} \right)^{3/2} \begin{pmatrix} 0 & 3 & 0 & \sqrt{2 \cdot 3} & 0 & \cdots \\ 3 & 0 & 6\sqrt{2} & 0 & \sqrt{2 \cdot 3 \cdot 4} & \cdots \\ 0 & 6\sqrt{2} & 0 & 9\sqrt{3} & 0 & \cdots \\ \sqrt{2 \cdot 3} & 0 & 9\sqrt{3} & 0 & & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

6.1.

$$\begin{aligned} \mathfrak{X} \cdot \mathfrak{P} - \mathfrak{P} \cdot \mathfrak{X} &= X_1 P_1 + X_2 P_2 + X_3 P_3 - P_1 X_1 - P_2 X_2 - P_3 X_3 = \\ &= i\hbar I + i\hbar I + i\hbar I = \underline{\underline{3i\hbar I}}. \end{aligned}$$

6.2. We have

$$(\mathfrak{A} \times \mathfrak{B}) \cdot \mathfrak{C} =$$

$$\begin{aligned}
&= \left(\begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{pmatrix} \times \begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{B}_3 \end{pmatrix} \right) \cdot \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \end{pmatrix} = \\
&= \begin{pmatrix} \mathbf{A}_2 \mathbf{B}_3 - \mathbf{A}_3 \mathbf{B}_2 \\ \mathbf{A}_3 \mathbf{B}_1 - \mathbf{A}_1 \mathbf{B}_3 \\ \mathbf{A}_1 \mathbf{B}_2 - \mathbf{A}_2 \mathbf{B}_1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \end{pmatrix} = \\
&= \mathbf{A}_2 \mathbf{B}_3 \mathbf{C}_1 - \mathbf{A}_3 \mathbf{B}_2 \mathbf{C}_1 + \mathbf{A}_3 \mathbf{B}_1 \mathbf{C}_2 - \mathbf{A}_1 \mathbf{B}_3 \mathbf{C}_2 + \mathbf{A}_1 \mathbf{B}_2 \mathbf{C}_3 - \mathbf{A}_2 \mathbf{B}_1 \mathbf{C}_3 = \\
&= \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{B}_2 \mathbf{C}_3 - \mathbf{B}_3 \mathbf{C}_2 \\ \mathbf{B}_3 \mathbf{C}_1 - \mathbf{B}_1 \mathbf{C}_3 \\ \mathbf{B}_1 \mathbf{C}_2 - \mathbf{B}_2 \mathbf{C}_1 \end{pmatrix} = \underline{\underline{\mathbf{A}}} \cdot (\underline{\underline{\mathbf{B}}} \times \underline{\underline{\mathbf{C}}}).
\end{aligned}$$

6.3. $\mathbf{L}_{\pm}^{\dagger} = (\mathbf{L}_1 \pm i\mathbf{L}_2)^{\dagger} = \mathbf{L}_1^{\dagger} \pm (i\mathbf{L}_2)^{\dagger} = \mathbf{L}_1 \mp i\mathbf{L}_2 = \mathbf{L}_{\mp} \neq \mathbf{L}_{\pm}$, so neither \mathbf{L}_+ nor \mathbf{L}_- is Hermitian.

6.4. We have

$$|\mathbf{L}_+ \mathbf{e}(j, m)|^2 = |\alpha \mathbf{e}(j, m+1)|^2 = |\alpha|^2 \underbrace{\mathbf{e}(j, m+1)^{\dagger} \mathbf{e}(j, m+1)}_1 = |\alpha|^2.$$

Also, it applies

$$(\mathbf{L}_+ \mathbf{e}(j, m))^{\dagger} (\mathbf{L}_+ \mathbf{e}(j, m)) = \mathbf{e}(j, m)^{\dagger} (\mathbf{L}_- \mathbf{L}_+) \mathbf{e}(j, m).$$

With $\mathbf{L}_- \mathbf{L}_+ = \mathbf{L}^2 - \mathbf{L}_3^2 - \hbar \mathbf{L}_3$ from (6.21), we obtain

$$\begin{aligned}
(\mathbf{L}_+ \mathbf{e}(j, m))^{\dagger} (\mathbf{L}_+ \mathbf{e}(j, m)) &= \mathbf{e}(j, m)^{\dagger} (\mathbf{L}^2 - \mathbf{L}_3^2 - \hbar \mathbf{L}_3) \mathbf{e}(j, m) \\
&= \hbar^2 (j(j+1) - m^2 - m) \underbrace{\mathbf{e}(j, m)^{\dagger} \mathbf{e}(j, m)}_1.
\end{aligned}$$

This implies

$$|\alpha|^2 = \hbar^2 (j(j+1) - m^2 - m)$$

and therefore

$$\alpha = \hbar \sqrt{j(j+1) - m^2 - m} = \hbar \sqrt{(j-m)(j+m+1)}.$$

The factor for \mathbf{L}_- can be calculated in a similar way.

7.1. First, we know that

$$\dot{\mathbf{P}}_i = -\frac{\partial \mathbf{H}}{\partial X_i}.$$

If the Hamiltonian is

$$\mathbf{H} = \frac{1}{2m} \mathfrak{P}^2 - Ze^2 \mathbf{R}^{-1},$$

then we find

$$\begin{aligned} \frac{\partial \mathbf{H}}{\partial X_i} &= -Ze^2 \frac{\partial \mathbf{R}^{-1}}{\partial X_i} = -Ze^2 \frac{\partial (X_1^2 + X_2^2 + X_3^2)^{-\frac{1}{2}}}{\partial X_i} = \\ &= Ze^2 X_i (X_1^2 + X_2^2 + X_3^2)^{-\frac{3}{2}} = Ze^2 X_i \mathbf{R}^{-3}, \end{aligned}$$

in other words,

$$\dot{\mathfrak{P}} = -Ze^2 \mathfrak{R} \mathbf{R}^{-3}. \quad (\text{A.7})$$

- 7.2.** For the first component matrix $X_1 \mathbf{R}^{-1}$ of $\mathfrak{R} \mathbf{R}^{-1}$, we obtain with the help of the Heisenberg formula $\frac{d}{dt} A = \frac{i}{\hbar} [E, A]$

$$\frac{d}{dt} (X_1 \mathbf{R}^{-1}) = \frac{i}{\hbar} \{ E(X_1 \mathbf{R}^{-1}) - (X_1 \mathbf{R}^{-1}) E \} = \quad (\text{A.8})$$

$$= \frac{i}{2m\hbar} \{ \mathfrak{P}^2 (X_1 \mathbf{R}^{-1}) - (X_1 \mathbf{R}^{-1}) \mathfrak{P}^2 \}. \quad (\text{A.9})$$

Note that we used the fact that

$$\mathbf{R} X_i = (X_1^2 X_i^2 + X_2^2 X_i^2 + X_3^2 X_i^2)^{1/2} = X_i \mathbf{R},$$

which implies $X_i \mathbf{R}^{-1} = \mathbf{R}^{-1} X_i$. Therefore, the component $-Ze^2 \mathbf{R}^{-1}$ of the Hamiltonian

$$E = \frac{1}{2m} \mathfrak{P}^2 - Ze^2 \mathbf{R}^{-1}$$

vanishes in (A.8). Now with

$$\mathfrak{P}^2 = \mathbf{P}_1^2 + \mathbf{P}_2^2 + \mathbf{P}_3^2$$

Equation (A.9) can be written as

$$\frac{d}{dt} (X_1 \mathbf{R}^{-1}) = \frac{i}{2m\hbar} \left\{ \sum_{j=1}^3 (\mathbf{P}_j^2 (X_1 \mathbf{R}^{-1}) - (X_1 \mathbf{R}^{-1}) \mathbf{P}_j^2) \right\}. \quad (\text{A.10})$$

We can add the entity $\mathbf{0} = -\mathbf{P}_j X_i \mathbf{R}^{-1} \mathbf{P}_j + \mathbf{P}_j X_i \mathbf{R}^{-1} \mathbf{P}_j$ into (A.10), which leads to

$$\frac{d}{dt} (X_1 \mathbf{R}^{-1}) =$$

$$= \frac{i}{2m\hbar} \left\{ \sum_{j=1}^3 \left\{ (\mathbf{P}_j(\mathbf{P}_j X_1 \mathbf{R}^{-1} - X_1 \mathbf{R}^{-1} \mathbf{P}_j) + (\mathbf{P}_j X_1 \mathbf{R}^{-1} - X_1 \mathbf{R}^{-1} \mathbf{P}_j) \mathbf{P}_j) \right\} \right\}. \quad (\text{A.11})$$

Multiplication with $\mathbf{R}^3 \mathbf{R}^{-3} = \mathbf{I}$ from the right supplies

$$\begin{aligned} & \frac{d}{dt} (X_1 \mathbf{R}^{-1}) = \\ &= \frac{i}{2m\hbar} \sum_{j=1}^3 \left\{ \mathbf{P}_j (\mathbf{P}_j X_1 \mathbf{R}^2 - X_1 \mathbf{R}^{-1} \mathbf{P}_j \mathbf{R}^3) + \right. \\ & \quad \left. + (\mathbf{P}_j X_1 \mathbf{R}^{-1} - X_1 \mathbf{R}^{-1} \mathbf{P}_j) \mathbf{P}_j \mathbf{R}^3 \right\} \mathbf{R}^{-3}. \end{aligned} \quad (\text{A.12})$$

If we multiply (7.5), namely

$$\mathbf{P}_j \mathbf{R} - \mathbf{R} \mathbf{P}_j = \frac{\hbar}{2\pi i} X_j \mathbf{R}^{-1},$$

with \mathbf{R}^{-1} from both the left and the right, and if we also take into account that $\mathbf{R}^{-1} X_j = X_j \mathbf{R}^{-1}$, we get

$$\mathbf{R}^{-1} \mathbf{P}_j - \mathbf{P}_j \mathbf{R}^{-1} = \frac{\hbar}{i} X_j \mathbf{R}^{-3}.$$

Rearranging yields

$$\mathbf{R}^{-1} \mathbf{P}_j = \mathbf{P}_j \mathbf{R}^{-1} - i\hbar X_j \mathbf{R}^{-3}. \quad (\text{A.13})$$

We can now put this result into the round parentheses in (A.12) in order to get

$$\begin{aligned} & \frac{d}{dt} (X_1 \mathbf{R}^{-1}) = \\ &= \frac{i}{2m\hbar} \sum_{j=1}^3 \left\{ \mathbf{P}_j (\mathbf{P}_j X_1 \mathbf{R}^2 - X_1 \mathbf{P}_j \mathbf{R}^2 + i\hbar X_1 X_j) + \right. \\ & \quad \left. + (\mathbf{P}_j X_1 \mathbf{R}^{-1} - X_1 \mathbf{P}_j \mathbf{R}^{-1} + i\hbar X_1 X_j \mathbf{R}^{-3}) \mathbf{P}_j \right\}. \end{aligned} \quad (\text{A.14})$$

Next, we know that $\mathbf{P}_j X_1 - X_1 \mathbf{P}_j = \frac{\hbar}{i} \mathbf{I}$ for $j = 1$ and $= \mathbf{0}$ for $j \neq 1$. Using this result as well as inserting $\mathbf{I} = \mathbf{R}^3 \mathbf{R}^{-3}$ in the second round parentheses, we get

$$\frac{d}{dt} (X_1 \mathbf{R}^{-1}) =$$

$$\begin{aligned}
&= \frac{1}{2m} \left\{ (\mathbf{P}_1(X_2^2 + X_3^2)\mathbf{R}^{-3} - \mathbf{P}_2X_1X_2\mathbf{R}^{-3} - \mathbf{P}_3X_1X_3\mathbf{R}^{-3}) + \right. \\
&\quad \left. + ((X_2^2 + X_3^2)\mathbf{R}^{-3}\mathbf{P}_1 - X_1X_2\mathbf{R}^{-3}\mathbf{P}_2 - X_1X_3\mathbf{R}^{-3}\mathbf{P}_3) \right\}. \quad (\text{A.15})
\end{aligned}$$

It turns out that

$$\begin{aligned}
&(\mathbf{P}_1(X_2^2 + X_3^2)\mathbf{R}^{-3} - \mathbf{P}_2X_1X_2\mathbf{R}^{-3} - \mathbf{P}_3X_1X_3\mathbf{R}^{-3}) = \\
&= X_3\mathbf{P}_1X_3\mathbf{R}^{-3} - X_1\mathbf{P}_3X_3\mathbf{R}^{-3} - X_1\mathbf{P}_2X_2\mathbf{R}^{-3} + X_2\mathbf{P}_1X_2\mathbf{R}^{-3} = \\
&= (X_3\mathbf{P}_1 - X_1\mathbf{P}_3)X_3\mathbf{R}^{-3} - (X_1\mathbf{P}_2 - X_2\mathbf{P}_1)X_2\mathbf{R}^{-3} = \\
&= \mathbf{L}_2X_3\mathbf{R}^{-3} - \mathbf{L}_3X_2\mathbf{R}^{-3},
\end{aligned}$$

that is, the first component of $\mathfrak{L} \times \mathfrak{R}\mathbf{R}^{-3}$. So we finally get the desired result

$$\frac{d}{dt}(\mathfrak{R}\mathbf{R}^{-1}) = \frac{1}{2m} \{ \mathfrak{L} \times (\mathfrak{R}\mathbf{R}^{-3}) - (\mathfrak{R}\mathbf{R}^{-3}) \times \mathfrak{L} \}. \quad (\text{A.16})$$

7.3. According to (7.11), the Lenz matrix-vector is defined as

$$\mathfrak{A} \stackrel{\text{def}}{=} \frac{1}{Ze^2m} \frac{1}{2} (\mathfrak{L} \times \mathfrak{P} - \mathfrak{P} \times \mathfrak{L}) + \mathfrak{R}\mathbf{R}^{-1},$$

where the angular momentum \mathfrak{L} (in the round parentheses) is constant. For the time derivative of \mathfrak{A} , we therefore obtain

$$\frac{d}{dt}\mathfrak{A} = \frac{1}{Ze^2m} \frac{1}{2} (\mathfrak{L} \times \dot{\mathfrak{P}} - \dot{\mathfrak{P}} \times \mathfrak{L}) + \frac{d}{dt}\mathfrak{R}\mathbf{R}^{-1}. \quad (\text{A.17})$$

With (A.7) and (A.16), we get

$$\begin{aligned}
\underline{\underline{\frac{d}{dt}\mathfrak{A}}} &= -\frac{1}{2m} \{ \mathfrak{L} \times (\mathfrak{R}\mathbf{R}^{-3}) - (\mathfrak{R}\mathbf{R}^{-3}) \times \mathfrak{L} \} + \\
&\quad + \frac{1}{2m} \{ \mathfrak{L} \times (\mathfrak{R}\mathbf{R}^{-3}) - (\mathfrak{R}\mathbf{R}^{-3}) \times \mathfrak{L} \} = \underline{\underline{0}}. \quad (\text{A.18})
\end{aligned}$$

8.1. The Pauli spin matrix σ_1 has the form

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

With α a real number, the power series expansion of the exponential function is

$$\exp(i\alpha\sigma_1) = \sum_{\nu=0}^{\infty} \frac{i\alpha^\nu}{\nu!} \sigma_1^\nu.$$

We find that

$$\sigma_1^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2,$$

and therefore

$$\sigma_1^{2\nu} = I_2 \quad \text{and} \quad \sigma_1^{2\nu+1} = \sigma_1. \quad (\text{A.19})$$

We can now separate the sum into one part with even integers and one part with odd integers:

$$\exp(i\alpha\sigma_1) = \sum_{\nu=0}^{\infty} \frac{(i\alpha)^{2\nu}}{(2\nu)!} \sigma_1^{2\nu} + \sum_{\nu=0}^{\infty} \frac{(i\alpha)^{2\nu+1}}{(2\nu+1)!} \sigma_1^{2\nu+1}.$$

With (A.19), we get

$$\begin{aligned} \exp(i\alpha\sigma_1) &= I_2 \sum_{\nu=0}^{\infty} (-1)^\nu \frac{(i\alpha)^{2\nu}}{(2\nu)!} + i\sigma_1 \sum_{\nu=0}^{\infty} (-1)^\nu \frac{(i\alpha)^{2\nu+1}}{(2\nu+1)!} = \\ &= I_2 \cos \alpha + i\sigma_1 \sin \alpha \\ &= \begin{pmatrix} \cos \alpha & 0 \\ 0 & \cos \alpha \end{pmatrix} + \begin{pmatrix} 0 & i \sin \alpha \\ i \sin \alpha & 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos \alpha & i \sin \alpha \\ i \sin \alpha & \cos \alpha \end{pmatrix}. \end{aligned}$$

8.2. Each of the three Pauli matrices has the two eigenvalues $+1$ and -1 . The corresponding normalized eigenvectors are

$$\begin{aligned} \mathbf{e}_{1+} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{e}_{1-} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \\ \mathbf{e}_{2+} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \mathbf{e}_{2-} = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}, \\ \mathbf{e}_{3+} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_{3-} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

8.3. If we write the eigenvectors as columns of the transformation matrix \mathbf{T} , we obtain, for example, for the Pauli matrix σ_1

$$\mathbf{T} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

This yields

$$\mathbf{T}^\dagger \boldsymbol{\sigma}_1 \mathbf{T} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

that is, a diagonal matrix with the eigenvalues +1 and -1 on the diagonal.

- 8.4.** We start with the largest possible values $j = \frac{1}{2} + \frac{1}{2} = 1$ and $m = \frac{1}{2} + \frac{1}{2} = 1$. m can then have the values -1, 0, and +1. For the largest total angular momentum quantum number $j = m = \frac{1}{2} + \frac{1}{2} = 1$, there is exactly one state in the coupled and uncoupled basis:

$$\mathbf{e}_{j=1, m=1} = \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}}. \quad (\text{A.20})$$

Now let us remember that for spin- $\frac{1}{2}$ systems the Pauli matrix $\boldsymbol{\sigma}_3$ has the two eigenvectors

$$\mathbf{e}_{\frac{1}{2}, \frac{1}{2}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The four possible vector combinations are

$$\begin{aligned} \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, & \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}} &= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \\ \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, & \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \end{aligned}$$

We can apply the lowering operator

$$\mathbf{J}_- = \mathbf{J}_1 + i\mathbf{J}_2 = \mathbf{S}_- \otimes \mathbf{I}_2 + \mathbf{I}_2 \otimes \mathbf{S}_-$$

to (A.20) and get

$$\mathbf{J}_- \mathbf{e}_{j=1, m=1} = (\mathbf{S}_- \mathbf{e}_{\frac{1}{2}, \frac{1}{2}}) \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} + \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} \otimes (\mathbf{S}_- \mathbf{e}_{\frac{1}{2}, \frac{1}{2}}). \quad (\text{A.21})$$

With (6.49), namely

$$\mathbf{J}_- \mathbf{e}_{jm} = [j(j+1) - m(m-1)]^{1/2} \hbar \mathbf{e}_{j, m-1}$$

and

$$S_- \mathbf{e}_{sm} = [s(s+1) - m(m-1)]^{1/2} \hbar \mathbf{e}_{s,m-1},$$

we obtain from (A.21)

$$(2)^{\frac{1}{2}} \hbar \mathbf{e}_{j=1,m=0} = \hbar \mathbf{e}_{\frac{1}{2},-\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2},\frac{1}{2}} + \hbar \mathbf{e}_{\frac{1}{2},\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2},-\frac{1}{2}},$$

which leads to

$$\begin{aligned} \mathbf{e}_{j=1,m=0} &= \frac{1}{\sqrt{2}} \mathbf{e}_{\frac{1}{2},-\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2},\frac{1}{2}} + \frac{1}{\sqrt{2}} \mathbf{e}_{\frac{1}{2},\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2},-\frac{1}{2}} = \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}. \end{aligned}$$

According to (8.37), the Clebsch–Gordan coefficients are defined as

$$C(a, b; j, m) = (\mathbf{e}'_a \otimes \mathbf{e}''_b)^\dagger \mathbf{e}(j, m).$$

For our problem, this reads

$$C(a, b; j, m) = (\mathbf{e}_a \otimes \mathbf{e}_b)^\dagger \mathbf{e}(j, m).$$

In particular, we find the Clebsch–Gordan coefficients

$$\begin{aligned} \underline{\underline{C\left(\left(\frac{1}{2}, -\frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}\right); j=1, m=0\right)}} &= (\mathbf{e}_{\frac{1}{2},-\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2},\frac{1}{2}})^\dagger \mathbf{e}(1, 0) = \frac{1}{\sqrt{2}}, \\ \underline{\underline{C\left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, -\frac{1}{2}\right); j=1, m=0\right)}} &= (\mathbf{e}_{\frac{1}{2},\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2},-\frac{1}{2}})^\dagger \mathbf{e}(1, 0) = \frac{1}{\sqrt{2}}. \end{aligned}$$

Another Clebsch–Gordan coefficient can be obtained directly from (A.20):

$$\underline{\underline{C\left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}\right); j=1, m=1\right)}} = 1.$$

Applying the lowering operator \mathbf{J}_- once again yields

$$\underline{\underline{C\left(\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, -\frac{1}{2}\right); j=1, m=-1\right)}} = 1.$$

Until now we have had three basis vectors, namely

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

By choosing an orthonormal vector corresponding to $\mathbf{e}_{j=1,m=0}$, we get the missing fourth basis vector

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}.$$

This means that the additional Clebsch–Gordan coefficients are $\pm \frac{1}{\sqrt{2}}$.

- 8.5.** Because \mathbf{J}^2 and \mathbf{J}_3 commute, they have identical eigenvectors \mathbf{e}_{jm} . The eigenvalue equations of \mathbf{J} are then

$$\mathbf{J}^2 \mathbf{e}_{jm} = j(j+1)\hbar^2 \mathbf{e}_{jm}, \quad (\text{A.22})$$

$$\mathbf{J}_3 \mathbf{e}_{jm} = m\hbar \mathbf{e}_{jm}, \quad -j \leq m \leq j. \quad (\text{A.23})$$

The largest total angular momentum quantum number is $j = m = \ell + \frac{1}{2}$. For these values, there is exactly one state in the coupled and uncoupled basis:

$$\mathbf{e}_{j=\ell+\frac{1}{2},m=\ell+\frac{1}{2}} = \mathbf{e}_{\ell,\ell} \otimes \mathbf{e}_{\frac{1}{2},\frac{1}{2}}.$$

As in the previous problem, we apply the lowering operator

$$\mathbf{J}_- = \mathbf{J}_1 + i\mathbf{J}_2 = \mathbf{L}_- \otimes \mathbf{I}_S + \mathbf{I}_L \otimes \mathbf{S}_-$$

and receive

$$\mathbf{J}_- \mathbf{e}_{j=\ell+\frac{1}{2},m=\ell+\frac{1}{2}} = \mathbf{L}_- \mathbf{e}_{\ell,\ell} \otimes \mathbf{e}_{\frac{1}{2},\frac{1}{2}} + \mathbf{e}_{\ell,\ell} \otimes \mathbf{S}_- \mathbf{e}_{\frac{1}{2},\frac{1}{2}}. \quad (\text{A.24})$$

With (6.49), namely

$$\mathbf{L}_- \mathbf{e}_{\ell m} = [\ell(\ell+1) - m(m-1)]^{1/2} \hbar \mathbf{e}_{\ell,m-1},$$

$$\mathbf{J}_- \mathbf{e}_{jm} = [j(j+1) - m(m-1)]^{1/2} \hbar \mathbf{e}_{j,m-1},$$

and

$$\mathbf{S}_- \mathbf{e}_{sm} = [s(s+1) - m(m-1)]^{1/2} \hbar \mathbf{e}_{s,m-1},$$

we obtain from (A.24)

$$(2\ell + 1)^{\frac{1}{2}} \hbar \mathbf{e}_{j=\ell+\frac{1}{2}, m=\ell-\frac{1}{2}} = (2\ell)^{\frac{1}{2}} \hbar \mathbf{e}_{\ell, \ell-1} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} + (1)^{\frac{1}{2}} \hbar \mathbf{e}_{\ell, \ell} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}}.$$

This can be written as

$$\mathbf{e}_{j=\ell+\frac{1}{2}, m=\ell-\frac{1}{2}} = \left(\frac{2\ell}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell-1} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} + \left(\frac{1}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}}.$$

The vectors $\mathbf{e}_{\ell, \ell-1} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}}$ and $\mathbf{e}_{\ell, \ell} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}}$ are orthogonal to each other. The orthogonal linear combination is therefore the state of the total angular momentum $j = \ell - 1/2$ with the same $m = \ell - 1/2$:

$$\mathbf{e}_{j=\ell-\frac{1}{2}, m=\ell-\frac{1}{2}} = \left(\frac{\ell}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell-1} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} - \left(\frac{2\ell}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}}.$$

Applying the lowering operator \mathbf{J}_- as above and writing down the orthogonal linear combination, we find

$$\mathbf{e}_{j=\ell+\frac{1}{2}, m=\ell-\frac{3}{2}} = \left(\frac{2\ell - 1}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell-2} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} + \left(\frac{2}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}},$$

$$\mathbf{e}_{j=\ell-\frac{1}{2}, m=\ell-\frac{3}{2}} = \left(\frac{2}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell-2} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} - \left(\frac{2\ell - 1}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, \ell-1} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}}.$$

With this method, we get the general result

$$\mathbf{e}_{j=\ell+\frac{1}{2}, m} = \left(\frac{\ell + m + \frac{1}{2}}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, m-\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} + \left(\frac{\ell - m + \frac{1}{2}}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, m+\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}},$$

$$\mathbf{e}_{j=\ell-\frac{1}{2}, m} = \left(\frac{\ell - m + \frac{1}{2}}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, m-\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, \frac{1}{2}} - \left(\frac{\ell + m + \frac{1}{2}}{2\ell + 1} \right)^{\frac{1}{2}} \mathbf{e}_{\ell, m+\frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, -\frac{1}{2}}.$$

There are $2(2\ell + 1)$ states in total. The Clebsch–Gordan coefficients are

$$\underline{\underline{C \left(\left(\ell, m \mp \frac{1}{2} \right), \left(\frac{1}{2}, \pm \frac{1}{2} \right); j = \ell + \frac{1}{2}, m \right) =}}$$

$$\underline{\underline{(\mathbf{e}_{\ell, m \mp \frac{1}{2}} \otimes \mathbf{e}_{\frac{1}{2}, \pm \frac{1}{2}})^\dagger \mathbf{e}_{j=\ell+\frac{1}{2}, m} = \left(\frac{\ell \pm m + \frac{1}{2}}{2\ell + 1} \right)^{\frac{1}{2}},}}$$

and

$$\begin{aligned} & \underline{\underline{C\left(\left(\ell, m \mp \frac{1}{2}\right), \left(\frac{1}{2}, \pm \frac{1}{2}\right); j = \ell - \frac{1}{2}, m\right)}} = \\ & = (\boldsymbol{e}_{\ell, m \mp \frac{1}{2}} \otimes \boldsymbol{e}_{\frac{1}{2}, \pm \frac{1}{2}})^\dagger \boldsymbol{e}_{j = \ell - \frac{1}{2}, m} = \pm \underline{\underline{\left(\frac{\ell \mp m + \frac{1}{2}}{2\ell + 1}\right)^{\frac{1}{2}}}}. \end{aligned}$$

- 9.1.** In the normal Zeeman effect, the energy level E_n splits into $2\ell + 1$ levels under the influence of a magnetic field. After all, the highest ℓ -value for a fixed n is $\ell = n - 1$, and these include $2\ell + 1 = 2n - 1$ values of m . For $\ell = 2$, we therefore find $2\ell + 1 = 5$ values of m , namely $m = +2, +1, 0, -1$ and -2 . In other words, the level $\ell = 2$ is split into $\underline{\underline{5}}$ levels.

- 10.1.** From (10.7), we find in matrix form

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \mathbf{C} = \begin{pmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{pmatrix} = \begin{pmatrix} a_1 \mathbf{b}^\top \\ a_2 \mathbf{b}^\top \end{pmatrix} = \mathbf{a} \mathbf{b}^\top.$$

In the second matrix, the second line is a multiple of the first line, therefore the determinant of the matrix is zero. This implies that the determinant of \mathbf{C} must also be equal to zero:

$$c_{11} c_{22} - c_{12} c_{21} = 0.$$

By the way, an algebraic computer program for the calculation of the a_i and b_j can be found in [23].

- 10.2.**

$$\mathbf{A}_{1,2} \mathbf{B}_{1,2} = (\mathbf{A} \otimes \mathbf{I}_B)(\mathbf{I}_A \otimes \mathbf{B}) = \mathbf{A} \otimes \mathbf{B},$$

$$\mathbf{B}_{1,2} \mathbf{A}_{1,2} = (\mathbf{I}_A \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{I}_B) = \mathbf{A} \otimes \mathbf{B}.$$

It follows that

$$[\mathbf{A}_{1,2}, \mathbf{B}_{1,2}] = \mathbf{A} \otimes \mathbf{B} - \mathbf{A} \otimes \mathbf{B} = \mathbf{0}.$$

- 10.3.**

$$\langle \mathbf{A}_{1,2} \rangle = (\xi_1 \otimes \xi_2)^\dagger (\mathbf{A} \otimes \mathbf{I}_B) (\xi_1 \otimes \xi_2) = (\xi_1^\dagger \mathbf{A} \xi_1) \underbrace{(\xi_2^\dagger \xi_2)}_1 = \langle \mathbf{A} \rangle,$$

$$\langle \mathbf{B}_{1,2} \rangle = (\xi_1 \otimes \xi_2)^\dagger (\mathbf{I}_A \otimes \mathbf{B}) (\xi_1 \otimes \xi_2) = \underbrace{(\xi_1^\dagger \xi_1)}_1 (\xi_2^\dagger \mathbf{B} \xi_2) = \langle \mathbf{B} \rangle.$$

- 10.4.** According to (11.8), we have

$$\mathbf{P}(\xi_1 \otimes \xi_2) = \xi_2 \otimes \xi_1$$

with the permutation matrix

$$\mathbf{P} = \mathbf{U}_{2 \times 2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Because \mathbf{P} is symmetric and real, we find $\mathbf{P}^\dagger = \mathbf{P}$. Due to $\mathbf{P}^2 = \mathbf{I}_4$, we get $\mathbf{P}^{-1} = \mathbf{P}$. The matrix \mathbf{P} therefore only has eigenvalues equal to ± 1 .

- 11.1.** Schrödinger would assign a wave function to each state of the cat, that is, “alive after an hour” and “dead after an hour.” The probabilities for both states would be equal. The total state of the system would then be a superposition of the two wave functions. In other words, the ψ -function of the system would express that the living and the dead cat are mixed in equal parts.

According to the *Copenhagen interpretation* of quantum mechanics, the wave function of the system collapses at the time of measurement. Once you open the chamber and observe the system (i.e., a measurement), the atomic nucleus jumps from the earlier superposition state into one of the eigenstates according to the measurement device. It is therefore only at the time of the measurement (by an outside observer) that it is decided whether the cat is dead or alive. Prior to the measurement, we can only make a probability statement about the status of the cat.

From the perspective of *ensemble theory*, the experiment would be described by an ensemble of identical systems, say 10,000 boxes with one cat in each box. After a certain time interval, approximately 5000 cats would be dead and approximately 5000 cats would be alive. This result is due to the empirical law of large numbers. According to this law, the more often you repeat the experiment, the better the results approach the theoretical probability of 50%. To me, this is the interpretation of a physicist, whereas other interpretations (such as Everett’s parallel universe, with one universe for the living cat and one universe for the dead cat) rather belong to the field of philosophy. In any case, we had better speak of Schrödinger’s cats, that is, in *plural*.

- 11.2.**

$$(a) [p, x]\psi(x) = (px - xp)\psi(x) = \frac{\hbar}{i} \frac{\partial}{\partial x} x\psi(x) - x \frac{\hbar}{i} \frac{\partial}{\partial x} \psi(x) = \\ = \frac{\hbar}{i}\psi(x) + x \frac{\hbar}{i} \frac{\partial}{\partial x} \psi(x) - x \frac{\hbar}{i} \frac{\partial}{\partial x} \psi(x) = \frac{\hbar}{i}\psi(x).$$

In other words, $\underline{\underline{[p, x]}} = \frac{\hbar}{i}$.

$$(b) [p, x^n]\psi(x) = \frac{\hbar}{i} \frac{\partial}{\partial x} x^n \psi(x) - x^n \frac{\hbar}{i} \frac{\partial}{\partial x} \psi(x) = \\ = nx^{n-1} \frac{\hbar}{i} \psi(x) + x^n \frac{\hbar}{i} \frac{\partial}{\partial x} \psi(x) - x^n \frac{\hbar}{i} \frac{\partial}{\partial x} \psi(x) = nx^{n-1} \frac{\hbar}{i} \psi(x);$$

that is, $[p, x^n] = n \frac{\hbar}{i} x^{n-1}$.

(c) With the commutator rule from Problem 3.7,

$$[AB, C] = A[B, C] + [A, C]B,$$

we prove the statement inductively.

$n = 2$:

$$[p^2, x] = [p \cdot p, x] = p \underbrace{[p, x]}_{\frac{\hbar}{i} I} + [p, x]p = 2 \frac{\hbar}{i} p.$$

$n - 1 \rightarrow n$:

$$\underline{[p^n, x]} = [p \cdot p^{n-1}, x] = p \underbrace{[p^{n-1}, x]}_{(n-1) \frac{\hbar}{i} p^{n-2}} + \underbrace{[p, x] p^{n-1}}_{\frac{\hbar}{i}} = \underline{\underline{n \frac{\hbar}{i} p^{n-1}}}.$$

11.3. (a) The differentiation of an operator A with respect to time t is defined as

$$\frac{dA}{dt} \stackrel{\text{def}}{=} \lim_{\epsilon \rightarrow 0} \frac{A(t + \epsilon) - A(t)}{\epsilon}.$$

For two time-dependent operators $A(t)$ and $B(t)$, we find

$$\begin{aligned} \underline{\frac{d}{dt}(A(t)B(t))} &= \lim_{\epsilon \rightarrow 0} \frac{A(t + \epsilon)B(t + \epsilon) - A(t)B(t)}{\epsilon} = \\ &= \lim_{\epsilon \rightarrow 0} \left[\frac{[A(t + \epsilon) - A(t)]B(t)}{\epsilon} + \frac{A(t + \epsilon)[B(t + \epsilon) - B(t)]}{\epsilon} \right] = \\ &= \underline{\underline{\frac{dA(t)}{dt}B(t) + A(t)\frac{dB(t)}{dt}}}. \end{aligned} \quad (\text{A.25})$$

Next, we show inductively that $\frac{dA(t)^n}{dt} = \sum_{\nu=1}^n A(t)^{\nu-1} \frac{dA}{dt} A(t)^{n-\nu}$

$n = 2$:

$$\frac{dA(t)^2}{dt} = \frac{dA(t)}{dt} A(t) + A(t) \frac{dA(t)}{dt} = \sum_{\nu=1}^2 A(t)^{\nu-1} \frac{dA}{dt} A(t)^{2-\nu}.$$

$n \rightarrow n + 1$:

$$\underline{\underline{\frac{dA(t)^{n+1}}{dt}}} \stackrel{(\text{A.25})}{=} \frac{dA(t)^n}{dt} A(t) + A(t)^n \frac{dA(t)}{dt} =$$

$$\begin{aligned}
&= \sum_{\nu=1}^n A(t)^{\nu-1} \frac{dA}{dt} A(t)^{n-\nu} + A(t)^n \frac{dA(t)}{dt} = \\
&= \sum_{\nu=1}^{n+1} A(t)^{\nu-1} \frac{dA}{dt} A(t)^{n+1-\nu}.
\end{aligned}$$

The differentiation of a function $f(A)$ of an operator A with respect to the operator A is defined as

$$\frac{df(A)}{dA} \stackrel{\text{def}}{=} \lim_{\epsilon \rightarrow 0} \frac{f(A + \epsilon \mathbf{1}) - f(A)}{\epsilon},$$

with the neutral element **1**. We obtain a similar product rule

$$\begin{aligned}
&\frac{d}{dA}(f(A)g(A)) = \lim_{\epsilon \rightarrow 0} \frac{f(A + \epsilon \mathbf{1})g(A + \epsilon \mathbf{1}) - f(A)g(A)}{\epsilon} = \\
&= \lim_{\epsilon \rightarrow 0} \left[\frac{[f(A + \epsilon \mathbf{1}) - f(A)]g(A)}{\epsilon} + \frac{f(A + \epsilon \mathbf{1})[g(A + \epsilon \mathbf{1}) - g(A)]}{\epsilon} \right] = \\
&= \frac{df(A)}{dA}g(A) + f(A)\frac{dg(A)}{dA}.
\end{aligned}$$

12.1. The gamma matrices are defined as

$$\gamma_i \stackrel{\text{def}}{=} \beta \alpha_i.$$

With (12.38) to (12.41), where σ_j are the Pauli matrices, we get

$$\alpha_1 = \sigma_1 \otimes \sigma_1 = \begin{pmatrix} \mathbf{0} & \sigma_1 \\ \sigma_1 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

$$\alpha_2 = \sigma_1 \otimes \sigma_2 = \begin{pmatrix} \mathbf{0} & \sigma_2 \\ \sigma_2 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix},$$

$$\alpha_3 = \sigma_1 \otimes \sigma_3 = \begin{pmatrix} \mathbf{0} & \sigma_3 \\ \sigma_3 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix},$$

$$\beta = \sigma_3 \otimes I_2 = \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

The gamma matrices are therefore

$$\gamma_0 = \beta = \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (\text{A.26})$$

$$\gamma_1 = \beta \alpha_1 = \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} \begin{pmatrix} \mathbf{0} & \sigma_1 \\ \sigma_1 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad (\text{A.27})$$

$$\gamma_2 = \beta \alpha_2 = \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} \begin{pmatrix} \mathbf{0} & \sigma_2 \\ \sigma_2 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}, \quad (\text{A.28})$$

$$\gamma_3 = \beta \alpha_3 = \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} \begin{pmatrix} \mathbf{0} & \sigma_3 \\ \sigma_3 & \mathbf{0} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (\text{A.29})$$

Note that for $j = 1, 2$, and 3 , we can also write the gamma matrices as

$$\gamma_j = \begin{pmatrix} \mathbf{0} & \sigma_j \\ -\sigma_j & \mathbf{0} \end{pmatrix}. \quad (\text{A.30})$$

In this form, we immediately find that

$$\gamma_j^2 = \begin{pmatrix} \mathbf{0} & \sigma_j \\ -\sigma_j & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \sigma_j \\ -\sigma_j & \mathbf{0} \end{pmatrix} = \begin{pmatrix} -I_2 & \mathbf{0} \\ \mathbf{0} & -I_2 \end{pmatrix} = -I_4, \quad (\text{A.31})$$

because $\sigma_j^2 = I_2$ for all j . With the *Levi-Civita symbol*

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i, j, k) \text{ is } (1, 2, 3), (2, 3, 1) \text{ or } (3, 1, 2), \\ -1 & \text{if } (i, j, k) \text{ is } (3, 2, 1), (1, 3, 2) \text{ or } (2, 1, 3), \\ 0 & \text{if } i = j \text{ or } j = k \text{ or } k = i, \end{cases}$$

we find for the product of two Pauli matrices ($\mu \neq \nu$)

$$\boldsymbol{\sigma}_\mu \boldsymbol{\sigma}_\nu = i \sum_\kappa \varepsilon_{\mu\nu\kappa} \boldsymbol{\sigma}_\kappa. \quad (\text{A.32})$$

With (A.31), we finally get ($\mu \neq \nu$)

$$\begin{aligned} \gamma_\mu \gamma_\nu &= \begin{pmatrix} \mathbf{0} & \boldsymbol{\sigma}_\mu \\ -\boldsymbol{\sigma}_\mu & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \boldsymbol{\sigma}_\nu \\ -\boldsymbol{\sigma}_\nu & \mathbf{0} \end{pmatrix} = \\ &= \begin{pmatrix} -\boldsymbol{\sigma}_\mu \boldsymbol{\sigma}_\nu & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\sigma}_\mu \boldsymbol{\sigma}_\nu \end{pmatrix} = \begin{pmatrix} -i \sum_\kappa \varepsilon_{\mu\nu\kappa} \boldsymbol{\sigma}_\kappa & \mathbf{0} \\ \mathbf{0} & -i \sum_\kappa \varepsilon_{\mu\nu\kappa} \boldsymbol{\sigma}_\kappa \end{pmatrix}. \end{aligned} \quad (\text{A.33})$$

For example, we get

$$\gamma_1 \gamma_2 = \begin{pmatrix} -i \boldsymbol{\sigma}_3 & \mathbf{0} \\ \mathbf{0} & -i \boldsymbol{\sigma}_3 \end{pmatrix} = \begin{pmatrix} -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & i \end{pmatrix}.$$

From (A.33), we immediately see that

$$\gamma_\mu \gamma_\nu = -\gamma_\nu \gamma_\mu, \quad (\text{A.34})$$

which implies that the matrices *anti-commute*:

$$\{\gamma_\mu, \gamma_\nu\} \stackrel{\text{def}}{=} \gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = \mathbf{0}. \quad (\text{A.35})$$

The commutation law also follows from (A.33) as

$$[\gamma_\mu, \gamma_\nu] \stackrel{\text{def}}{=} \gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu = 2 \begin{pmatrix} -i \sum_\kappa \varepsilon_{\mu\nu\kappa} \boldsymbol{\sigma}_\kappa & \mathbf{0} \\ \mathbf{0} & -i \sum_\kappa \varepsilon_{\mu\nu\kappa} \boldsymbol{\sigma}_\kappa \end{pmatrix}. \quad (\text{A.36})$$

12.2. Using (2.5) to (2.4), we first obtain

$$i \gamma_0 \frac{\partial}{\partial x_0} = \begin{pmatrix} i \frac{\partial}{\partial x_0} & 0 & 0 & 0 \\ 0 & i \frac{\partial}{\partial x_0} & 0 & 0 \\ 0 & 0 & -i \frac{\partial}{\partial x_0} & 0 \\ 0 & 0 & 0 & -i \frac{\partial}{\partial x_0} \end{pmatrix}, \quad (\text{A.37})$$

$$i \gamma_1 \frac{\partial}{\partial x_1} = \begin{pmatrix} 0 & 0 & 0 & i \frac{\partial}{\partial x_1} \\ 0 & 0 & i \frac{\partial}{\partial x_1} & 0 \\ 0 & -i \frac{\partial}{\partial x_1} & 0 & 0 \\ -i \frac{\partial}{\partial x_1} & 0 & 0 & 0 \end{pmatrix}, \quad (\text{A.38})$$

$$i \gamma_1 \frac{\partial}{\partial x_2} = \begin{pmatrix} 0 & 0 & 0 & \frac{\partial}{\partial x_2} \\ 0 & 0 & -\frac{\partial}{\partial x_2} & 0 \\ 0 & -\frac{\partial}{\partial x_2} & 0 & 0 \\ \frac{\partial}{\partial x_2} & 0 & 0 & 0 \end{pmatrix}, \quad (\text{A.39})$$

$$i \gamma_1 \frac{\partial}{\partial x_3} = \begin{pmatrix} 0 & 0 & i \frac{\partial}{\partial x_3} & 0 \\ 0 & 0 & 0 & -i \frac{\partial}{\partial x_3} \\ -i \frac{\partial}{\partial x_3} & 0 & 0 & 0 \\ 0 & i \frac{\partial}{\partial x_3} & 0 & 0 \end{pmatrix}. \quad (\text{A.40})$$

Adding these four matrices yields

$$\begin{aligned} i\phi &= i \sum_{j=0}^3 \gamma_j \frac{\partial}{\partial x_j} = \\ &= \begin{pmatrix} i \frac{\partial}{\partial x_0} & 0 & i \frac{\partial}{\partial x_3} & i \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2} \\ 0 & i \frac{\partial}{\partial x_0} & i \frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2} & -i \frac{\partial}{\partial x_3} \\ -i \frac{\partial}{\partial x_3} & -i \frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2} & -i \frac{\partial}{\partial x_0} & 0 \\ \frac{\partial}{\partial x_2} - i \frac{\partial}{\partial x_1} & i \frac{\partial}{\partial x_3} & 0 & -i \frac{\partial}{\partial x_0} \end{pmatrix}. \quad (\text{A.41}) \end{aligned}$$

This takes us to four linear differential equations:

$$(i\phi - \frac{m_0}{\hbar} \mathbf{I}_4) \psi =$$

$$\begin{pmatrix} i \frac{\partial}{\partial x_0} - \frac{m_0}{\hbar} & 0 & i \frac{\partial}{\partial x_3} & i \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2} \\ 0 & i \frac{\partial}{\partial x_0} - \frac{m_0}{\hbar} & i \frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2} & -i \frac{\partial}{\partial x_3} \\ -i \frac{\partial}{\partial x_3} & -i \frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2} & -i \frac{\partial}{\partial x_0} - \frac{m_0}{\hbar} & 0 \\ \frac{\partial}{\partial x_2} - i \frac{\partial}{\partial x_1} & i \frac{\partial}{\partial x_3} & 0 & -i \frac{\partial}{\partial x_0} - \frac{m_0}{\hbar} \end{pmatrix} \begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} =$$

$$\begin{pmatrix} i \frac{\partial \psi_0}{\partial x_0} + i \frac{\partial \psi_3}{\partial x_1} + \frac{\partial \psi_3}{\partial x_2} + i \frac{\partial \psi_2}{\partial x_3} - \frac{m_0}{\hbar} \psi_0 \\ i \frac{\partial \psi_1}{\partial x_0} + i \frac{\partial \psi_2}{\partial x_1} - \frac{\partial \psi_2}{\partial x_2} - i \frac{\partial \psi_3}{\partial x_3} - \frac{m_0}{\hbar} \psi_1 \\ -i \frac{\partial \psi_2}{\partial x_0} - i \frac{\partial \psi_1}{\partial x_1} - \frac{\partial \psi_1}{\partial x_2} - i \frac{\partial \psi_0}{\partial x_3} - \frac{m_0}{\hbar} \psi_2 \\ -i \frac{\partial \psi_3}{\partial x_0} - i \frac{\partial \psi_0}{\partial x_1} + \frac{\partial \psi_0}{\partial x_2} + i \frac{\partial \psi_1}{\partial x_3} - \frac{m_0}{\hbar} \psi_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (\text{A.42})$$

With the ansatz of a plane wave

$$\psi(\mathbf{x}, t) = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} \exp(i(\mathbf{k}^\top \mathbf{x} - \omega t)), \quad (\text{A.43})$$

we obtain for the first derivative

$$\frac{\partial}{\partial x_0} \psi = \mathbf{c} (-i \omega/c) \exp(i(\mathbf{k}^\top \mathbf{x} - \omega t)). \quad (\text{A.44})$$

With $\mathbf{k}^\top \mathbf{x} = k_1 x_1 + k_2 x_2 + k_3 x_3$, we further get

$$\frac{\partial}{\partial x_j} \psi = \mathbf{c} (i k_j) \exp(i(\mathbf{k}^\top \mathbf{x} - \omega t)). \quad (\text{A.45})$$

If we insert these derivatives into (A.42), we finally get the linear algebraic equation

$$\begin{pmatrix} \frac{\omega}{c} - \frac{m_0}{\hbar} & 0 & -k_3 & ik_2 - k_1 \\ 0 & \frac{\omega}{c} - \frac{m_0}{\hbar} & -k_1 - ik_2 & k_3 \\ k_3 & k_1 - ik_2 & -\frac{\omega}{c} - \frac{m_0}{\hbar} & 0 \\ k_1 + ik_2 & -k_3 & 0 & -\frac{\omega}{c} - \frac{m_0}{\hbar} \end{pmatrix} \mathbf{c} = \mathbf{0}. \quad (\text{A.46})$$

Note that there is an algebraic computer program (written in MAXIMA) in [23].

Appendix B

The Kronecker Product

B.1 Definitions

The Kronecker product of two matrices $\mathbf{A} \in \mathbb{C}^{n \times m}$ and $\mathbf{B} \in \mathbb{C}^{p \times q}$ leads to a matrix $\mathbf{C} \in \mathbb{C}^{np \times mq}$ and is denoted by

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{C}.$$

The submatrices $\mathbf{C}_{ij} \in \mathbb{C}^{p \times q}$ for $i = 1$ to n and for $j = 1$ to m are defined as

$$\mathbf{C}_{ij} \stackrel{\text{def}}{=} a_{ij} \mathbf{B},$$

the matrix \mathbf{C} therefore has the form

$$\mathbf{C} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2m}\mathbf{B} \\ \dots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{pmatrix}.$$

B.2 Properties of the Kronecker Product

Without proof, the following important properties also hold (see, e.g., [7]):

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}), \quad (\text{B.1})$$

$$(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top, \quad (\text{B.2})$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}. \quad (\text{B.3})$$

Appendix C

Fourier Decomposition of Periodic Functions

Let $f(t)$ be an arbitrary periodic function. Its Fourier decomposition is given by the sum of simple sine waves

$$f(t) = \frac{y_0}{2} + \sum_{n=1}^{\infty} y_n \cdot \sin(n\omega t + \varphi_n),$$

where the y_n are the *amplitudes* and the φ_n are the *phases*. With the help of the trigonometric addition theorem, the function can be decomposed into

$$f(t) = \frac{y_0}{2} + \sum_{n=1}^{\infty} a_n \cdot \cos(n\omega t) + \sum_{n=1}^{\infty} b_n \cdot \sin(n\omega t). \quad (\text{C.1})$$

The goal of *Fourier analysis* is to compute the coefficients a_n and b_n (with the help of integrals over a full period).

We can also write the Fourier series with the help of complex numbers as

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega t}.$$

The connection with the above notation with sine and cosine expressions is obtained as

$$\begin{aligned} \sum_{n=-\infty}^{\infty} c_n e^{in\omega t} &= c_0 + \sum_{n=1}^{\infty} c_n e^{in\omega t} + \sum_{n=-1}^{-\infty} c_n e^{in\omega t} \\ &= c_0 + \sum_{n=1}^{\infty} c_n (\cos n\omega t + i \sin n\omega t) + \sum_{n=-1}^{-\infty} c_n (\cos n\omega t + i \sin n\omega t) \end{aligned}$$

$$\begin{aligned}
&= c_0 + \sum_{n=1}^{\infty} c_n (\cos n\omega t + i \sin n\omega t) + \sum_{n=1}^{\infty} c_n (\cos n\omega t - i \sin n\omega t) \\
&= c_0 + \sum_{n=1}^{\infty} [(c_n + c_{-n}) \cos n\omega t + i(c_n - c_{-n}) \sin n\omega t]. \tag{C.2}
\end{aligned}$$

A comparison of (C.1) with (C.2) yields

$$c_o = \frac{y_0}{2},$$

$$a_n = c_n + c_{-n}$$

and

$$b_n = i(c_n - c_{-n}),$$

or

$$c_n = \frac{a_n - ib_n}{2} \quad \text{and} \quad c_{-n} = \frac{a_n + ib_n}{2} = c_n^*.$$

The main idea of Fourier analysis is to compare the signal with sinusoids of various frequencies ω_n . Each such sinusoid may be thought of as a prototype oscillation. As a result, we obtain for each considered frequency parameter ω_n a magnitude coefficient $y_n \geq 0$, along with a phase coefficient φ_n . In the case that the coefficient y_n is large, there is a high similarity between the signal and the sinusoid of frequency ω_n , and the signal contains a periodic oscillation at that frequency. In the case that y_n is small, the signal does not contain a remarkable periodic component at that frequency. The original signal and the Fourier transform contain the same amount of information. This information, however, is represented in different ways. Whereas the signal displays the information across time, the Fourier transform displays the information across frequency.

Appendix D

Laplace–Runge–Lenz Vector

In a potential of the form $V(r) = -\alpha/r$ (e.g., the Coulomb potential or the gravitational potential), the Laplace–Runge–Lenz vector is a conserved entity. In classical mechanics, the vector is primarily used to describe the shape and orientation of orbits of astronomical bodies, such as the path of a planet around its star. For two bodies that move according to Newtonian physics, the Laplace–Runge–Lenz vector is a constant of motion; that is, it is the same on every point of the path. With

$$\boldsymbol{\ell} = \mathbf{r} \times \mathbf{p} \quad (\text{D.1})$$

the (time-constant) angular momentum of the electron around the nucleus, and

$$\mathbf{p} = m \mathbf{v}$$

its linear momentum, the respective (constant) Laplace–Runge–Lenz vector is defined as

$$\boldsymbol{a} \stackrel{\text{def}}{=} \frac{1}{Ze^2m}(\boldsymbol{\ell} \times \mathbf{p}) + \mathbf{r}/r. \quad (\text{D.2})$$

Let us now prove that the Laplace–Runge–Lenz vector is indeed constant over time. With $k \stackrel{\text{def}}{=} Z^2e$, we get

$$\boldsymbol{a} = \frac{1}{k m}(\boldsymbol{\ell} \times \mathbf{p}) + \mathbf{r}/r. \quad (\text{D.3})$$

In addition, the time derivative of the momentum for the potential $V = -k/r$ is

$$\dot{\mathbf{p}} = -k \mathbf{r}/r^3, \quad (\text{D.4})$$

that is, Newton's law of motion. Now

$$\begin{aligned}
\dot{\mathbf{a}} &= \frac{d}{dt} \left(\frac{1}{k m} (\ell \times \mathbf{p}) + \mathbf{r}/r \right) = \\
&= \frac{1}{k m} \underbrace{(\dot{\ell} \times \mathbf{p} + \ell \times \dot{\mathbf{p}})}_0 + \frac{\dot{\mathbf{r}}}{r} + \mathbf{r} \left(\left(\frac{\partial}{\partial r} \frac{1}{r} \right) \cdot \dot{\mathbf{r}} \right) = \\
&= \frac{1}{k m} (\mathbf{r} \times \mathbf{p}) \times \dot{\mathbf{p}} + \frac{1}{m r} \mathbf{p} + \mathbf{r} \left(\mathbf{r} \frac{-1}{r^3} \cdot \dot{\mathbf{r}} \right).
\end{aligned}$$

With (D.4) and the vector product rule $\alpha \times \beta \times \gamma = (\alpha \cdot \gamma)\beta - (\alpha \cdot \beta)\gamma$, we find indeed

$$\begin{aligned}
\dot{\mathbf{a}} &= -\frac{1}{mr^3} \underbrace{(\mathbf{r} \times \mathbf{p}) \times \mathbf{r}}_{(\mathbf{r} \cdot \mathbf{r})\mathbf{p} - (\mathbf{r} \cdot \mathbf{p})\mathbf{r}} + \frac{1}{m r} \mathbf{p} + \mathbf{r} \left(\mathbf{r} \frac{-1}{r^3} \cdot \underbrace{\dot{\mathbf{r}}}_{\frac{1}{m}\mathbf{p}} \right) = \\
&= -\frac{1}{m r} \mathbf{p} + \frac{1}{mr^3} (\mathbf{r} \cdot \mathbf{p}) \mathbf{r} + \frac{1}{m r} \mathbf{p} - \mathbf{r} (\mathbf{r} \cdot \mathbf{p}) \frac{1}{mr^3} = 0.
\end{aligned}$$

D.1 Further Properties of the Vector \mathbf{a}

Lemma *The vector \mathbf{a} is perpendicular to the angular momentum vector ℓ .*

Proof The vector $\ell \times \mathbf{p}$ is perpendicular to ℓ . Also, the vector $\ell = \mathbf{r} \times \mathbf{p}$ is perpendicular to \mathbf{r} . This implies

$$\ell \cdot \mathbf{a} = \frac{1}{Ze^2 m} \underbrace{\ell \cdot (\ell \times \mathbf{p})}_0 + \underbrace{\ell \cdot \mathbf{r}}_0 / r = 0. \quad \mathbf{q.e.d.}$$

Lemma *The scalar product of \mathbf{a} and \mathbf{r} is*

$$\mathbf{a} \cdot \mathbf{r} = -\frac{1}{Ze^2 m} \ell^2 + r.$$

Proof With $(\alpha \times \beta) \cdot \gamma = \alpha \cdot (\beta \times \gamma)$, we find

$$\mathbf{a} \cdot \mathbf{r} = \frac{1}{Ze^2m} \underbrace{\ell(\mathbf{p} \times \mathbf{r}) \cdot \mathbf{r}}_{\ell} + \frac{(\mathbf{r} \cdot \mathbf{r})}{r} = -\frac{1}{Ze^2m} \ell^2 + r. \quad \mathbf{q.e.d.}$$

Lemma

$$|\mathbf{a}| = \sqrt{\frac{2H}{Z^2e^4m} \ell^2 + 1}. \quad (\text{D.5})$$

Proof With the Lagrange identity

$$(\boldsymbol{\alpha} \times \boldsymbol{\beta}) \cdot (\boldsymbol{\gamma} \times \boldsymbol{\delta}) = (\boldsymbol{\alpha} \cdot \boldsymbol{\gamma})(\boldsymbol{\beta} \cdot \boldsymbol{\delta}) - (\boldsymbol{\alpha} \cdot \boldsymbol{\delta})(\boldsymbol{\beta} \cdot \boldsymbol{\gamma})$$

and the Hamilton function $H = \frac{1}{2m} \mathbf{p}^2 - \frac{Ze^2}{r}$, we find

$$\begin{aligned} |\mathbf{a}|^2 &= \mathbf{a} \cdot \mathbf{a} = \left| \frac{1}{Ze^2m} (\ell \times \mathbf{p}) + \mathbf{r}/r \right|^2 = \\ &= \frac{1}{Z^2e^4m^2} \underbrace{(\ell \times \mathbf{p}) \cdot (\ell \times \mathbf{p})}_{(\ell^2)(p^2) - \underbrace{(\ell p)(p \ell)}_0} + \frac{2}{Ze^2m} \underbrace{(\ell \times \mathbf{p}) \cdot \mathbf{r}}_{-\ell^2}/r + 1 = \\ &= \frac{2}{Z^2e^4m} \ell^2 \underbrace{\left(\frac{1}{2m} \mathbf{p}^2 - \frac{Ze^2}{r} \right)}_H + 1. \quad \mathbf{q.e.d.} \end{aligned}$$

Appendix E

Permutation

A permutation without repetition is an arrangement of n objects. For the first object, we have n placement possibilities, for the second object only $n - 1$ possibilities, for the third object only $n - 2$, and so on. For the last item, there is only one empty seat left. The number of possible permutations of n objects is therefore $n! = n \cdot (n - 1) \cdots 1$. For example, there are $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$ possible ways to arrange four differently colored balls.

Let us now introduce a more precise representation of an n digit permutation π as Cauchy's two-line notation with two rows and n columns. In the top row, we place the numbers 1 to n (in any order). Under each number j , we put the function value $\pi(j)$ in the second line. For example,

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{pmatrix}.$$

Obviously, the second row also contains each number 1 to n exactly once. As an example, let us consider the permutation

$$\pi : \{1, 2, 3, 4\} \rightarrow \{2, 4, 3, 1\}$$

with $\pi(1) = 2$, $\pi(2) = 4$, $\pi(3) = 3$, and $\pi(4) = 1$. In Cauchy's two-line notation, this permutation looks like

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}.$$

One can represent a permutation of n objects as an $n \times n$ matrix. If a permutation $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ of n objects is given, then the permutation matrix \mathbf{P}_π is defined as

$$\mathbf{P}_\pi \stackrel{\text{def}}{=} (\vec{p}_1 \dots \vec{p}_n),$$

where \vec{p}_i is a *canonical basis vector*,¹ thus a permutation matrix is a square binary matrix that has exactly one entry of 1 in each row and each column and 0s elsewhere. As an example, consider the permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 5 & 3 \end{pmatrix}.$$

The corresponding permutation matrix now has the form

$$\mathbf{P}_\pi = (\vec{p}_1 \ \vec{p}_2 \ \vec{p}_3 \ \vec{p}_4 \ \vec{p}_5) =$$

$$= (\vec{e}_4 \ \vec{e}_2 \ \vec{e}_1 \ \vec{e}_5 \ \vec{e}_3) = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

For a vector $\vec{v} = (v_1, v_2, v_3, v_4, v_5)^\top$, this yields

$$\mathbf{P}_\pi \vec{v} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix} = \begin{pmatrix} v_3 \\ v_2 \\ v_5 \\ v_1 \\ v_4 \end{pmatrix}.$$

¹The canonical basis vector \vec{e}_k is the k th column of the $n \times n$ -identity matrix \mathbf{I}_n .

Appendix F

Determinants

F.1 Axiomatic Definition of the Determinant of a Square Matrix

A mapping from the space of square matrices onto the underlying field maps each matrix $\mathbf{A} = (\mathbf{v}_1 | \dots | \mathbf{v}_n)$ on its determinant $\det \mathbf{A}$, if it satisfies the following three properties.

- It is *multilinear*, that is,

$$\begin{aligned}\det(\mathbf{v}_1 | \dots | \mathbf{v}_{i-1} | b\mathbf{v}_i + c\mathbf{w} | \mathbf{v}_{i+1} | \dots | \mathbf{v}_n) &= \\ &= b \det(\mathbf{v}_1 | \dots | \mathbf{v}_{i-1} | \mathbf{v}_i | \mathbf{v}_{i+1} | \dots | \mathbf{v}_n) + \\ &\quad + c \det(\mathbf{v}_1 | \dots | \mathbf{v}_{i-1} | \mathbf{w} | \mathbf{v}_{i+1} | \dots | \mathbf{v}_n)\end{aligned}$$

for all $\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w} \in V$, and for all $b, c \in K$.

- If we interchange two columns or two rows, the sign is changed

$$\det(\mathbf{v}_1 | \dots | \mathbf{v}_i | \dots | \mathbf{v}_j | \dots | \mathbf{v}_n) = - \det(\mathbf{v}_1 | \dots | \mathbf{v}_j | \dots | \mathbf{v}_i | \dots | \mathbf{v}_n).$$

- If two columns or two rows are equal, the determinant is zero.
- It is normalized, that is,

$$\det \mathbf{I}_n = 1$$

for the identity matrix \mathbf{I}_n .

F.2 The Leibniz Formula

For an $n \times n$ matrix A with elements $a_{i,j}$, the determinant can be defined by the Leibniz formula

$$\det A = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} \right).$$

The sum stretches over all permutations² σ of the symmetric group S_n of degree n . $\operatorname{sgn}(\sigma)$ denotes the sign of the permutation σ , that is, $+1$ if σ is an even permutation, and -1 if it is odd. We can recognize whether a permutation is even or odd by counting the number of transpositions that are required to generate the permutation. An even number of permutations means that the permutation is even; an odd number of permutations means that the permutation is odd.

²See Appendix E.

Appendix G

Dirac's Bra-Ket Notation

In [9], Dirac introduced a special notation for quantum mechanical states. Today, this notation is used in many books on quantum mechanics and should therefore be briefly described here as well.

The state space in quantum mechanics is a complex finite or infinite vector space. Dirac denotes an element f of the vector space by $|f\rangle$, which he then calls a *ket* vector. An example for a one-dimensional ket is Schrödinger's wave function $|\psi\rangle$, whose representation in position space is the well-known complex-valued wave function $\psi(x)$. An example for a four-dimensional ket is the vector

$$|\psi\rangle \stackrel{\text{def}}{=} \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$$

in the Dirac equation. Note that the ket $|\psi\rangle$ stands for the *entire* wave function ψ !

Dirac then defines a *dual* to each ket called the *bra*. We get a bra from the respective ket by taking its conjugate complex (if the ket is a vector, we also need to transpose):

$$\langle f | \stackrel{\text{def}}{=} (f^*)^\top = f^\dagger.$$

The *scalar product* of two vectors can then be written with the bra and the ket as

$$\langle f | g \rangle \stackrel{\text{def}}{=} f^\dagger g.$$

We then have

$$\langle f | g \rangle = \langle g | f \rangle^*.$$

For an operator C , the definition can be extended to

$$\langle f | C | g \rangle \stackrel{\text{def}}{=} f^\dagger C g.$$

This definition allows us to write the operation of C in two equivalent versions, depending on to which state we prefer to apply the operator:

$$\langle f | C | g \rangle = \langle C^\dagger f | g \rangle = \langle f | Cg \rangle.$$

The length of a vector f is given by

$$|f| = \sqrt{\langle f | f \rangle}.$$

For the case of wave functions in position space, the scalar product is defined as

$$\langle f | g \rangle \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} f^*(x)g(x)dx.$$

Appendix H

Proofs of Pauli's Formulas

H.1 ‘Commutator Gymnastics’

First we note formulas that we need to perform the proofs. We have ($k = 1, 2$, and 3) for $\mathbf{F} = \mathbf{F}(X_k, \mathbf{P}_k)$

$$(3.27) \quad [X_k, \mathbf{F}] = i\hbar \frac{\partial \mathbf{F}}{\partial \mathbf{P}_k},$$

$$(3.28) \quad [\mathbf{P}_k, \mathbf{F}] = -i\hbar \frac{\partial \mathbf{F}}{\partial X_k}.$$

For $\mathbf{F} = \mathbf{R} = \sqrt{X_1^2 + X_2^2 + X_3^2}$ we get

$$\underline{[X_k, \mathbf{R}]} = i\hbar \frac{\partial \mathbf{R}}{\partial \mathbf{P}_k} = \underline{\underline{\mathbf{0}}}, \quad (\text{H.1})$$

and

$$\underline{[\mathbf{P}_k, \mathbf{R}]} = -i\hbar \frac{\partial \mathbf{R}}{\partial X_k} = \underline{\underline{i\hbar X_k \mathbf{R}^{-1}}}. \quad (\text{H.2})$$

Also

$$\underline{[\mathbf{P}_k, \mathbf{R}^{-1}]} = -i\hbar \frac{\partial(\mathbf{R}^{-1})}{\partial X_k} = \underline{\underline{-i\hbar X_k (\mathbf{R}^3)^{-1}}}. \quad (\text{H.3})$$

Furthermore

$$[\mathbf{L}_1, \mathbf{R}] = (X_2 \mathbf{P}_3 - X_3 \mathbf{P}_2) \mathbf{R} - \mathbf{R} (X_2 \mathbf{P}_3 - X_3 \mathbf{P}_2) = X_2 \underbrace{[\mathbf{P}_3, \mathbf{R}]}_{-i\hbar X_3 \mathbf{R}^{-1}} + X_3 \underbrace{[\mathbf{R}, \mathbf{P}_2]}_{i\hbar X_2 \mathbf{R}^{-1}} = \mathbf{0},$$

generally

$$\underline{[\mathbf{L}_k, \mathbf{R}]} = \mathbf{0}. \quad (\text{H.4})$$

This implies that \mathbf{R}^{-1} commutes with \mathbf{L}_k . For example, from $\mathbf{L}_k \mathbf{R} = \mathbf{R} \mathbf{L}_k$ it follows that

$$\mathbf{R}^{-1} \mathbf{L}_k \mathbf{R} \mathbf{R}^{-1} = \mathbf{R}^{-1} \mathbf{R} \mathbf{L}_k \mathbf{R}^{-1}$$

so

$$\mathbf{R}^{-1} \mathbf{L}_k = \mathbf{L}_k \mathbf{R}^{-1}$$

or

$$\underline{\underline{[\mathbf{L}_k, \mathbf{R}^{-1}] = \mathbf{0}}} \quad (\text{H.5})$$

It is

$$[\mathbf{X}_1, \mathbf{L}_1] = i\hbar \frac{\partial \mathbf{L}_1}{\partial \mathbf{P}_k} = i\hbar \frac{\partial (X_2 \mathbf{P}_3 - X_3 \mathbf{P}_2)}{\partial \mathbf{P}_1} = \mathbf{0}$$

and generally

$$\underline{\underline{[\mathbf{X}_k, \mathbf{L}_k] = \mathbf{0}}} \quad (\text{H.6})$$

Next we have

$$[\mathbf{X}_1, \mathbf{L}_2] = i\hbar \frac{\partial \mathbf{L}_2}{\partial \mathbf{P}_1} = i\hbar \frac{\partial (X_3 \mathbf{P}_1 - X_1 \mathbf{P}_3)}{\partial \mathbf{P}_1} = i\hbar \mathbf{X}_3,$$

and general³

$$\underline{\underline{[\mathbf{X}_i, \mathbf{L}_j] = i\hbar \varepsilon_{ijk} \mathbf{X}_k}} \quad (\text{H.7})$$

Similarly for the commutator with \mathbf{P} we get

$$\underline{\underline{[\mathbf{P}_k, \mathbf{L}_k] = \mathbf{0}}} \quad (\text{H.8})$$

and

$$\underline{\underline{[\mathbf{P}_i, \mathbf{L}_j] = i\hbar \varepsilon_{ijk} \mathbf{P}_k}} \quad (\text{H.9})$$

Notice these commutator properties [29]

$$\underline{\underline{[\mathbf{A}, \mathbf{BC}] = \mathbf{B}[\mathbf{A}, \mathbf{C}] + [\mathbf{A}, \mathbf{B}]\mathbf{C}}}.$$

³The Levi-Civita symbol ε_{ijk} is defined as follows.

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i, j, k) \text{ is } (1, 2, 3), (2, 3, 1), \text{ or } (3, 1, 2), \\ -1 & \text{if } (i, j, k) \text{ is } (3, 2, 1), (1, 3, 2), \text{ or } (2, 1, 3), ; \\ 0 & \text{if } i = j \text{ or } j = k \text{ or } k = i \end{cases}$$

that is, ε_{ijk} is 1 if (i, j, k) is an even permutation of $(1, 2, 3)$, -1 if it is an odd permutation, and 0 if any index is repeated.

Proof: $B[A, C] + [A, B]C = B(AC - CA) + (AB - BA)C = ABC - BCA = [A, BC]$. q.e.d.

Similarly,

$$\underline{[AB, C] = A[B, C] + [A, C]B}. \quad (\text{H.10})$$

Suppose now that $[A, B] = \mathbf{0}$. If n is a positive integer, then so is

$$\underline{[A^n, B] = \mathbf{0}}.$$

Proof By induction: The base case, $n = 1$, is given. Then using the inductive hypothesis, we get

$$[A^n, B] = [A^{n-1}, B]A + A^{n-1}[A, B] = \mathbf{0}. \quad \text{q.e.d.}$$

H.2 Proof of Pauli's Formula (7.14)

$$[A_i, L_i] = \mathbf{0} \quad (7.14)$$

First we prove

$$-\mathfrak{P} \times \mathfrak{L} = \mathfrak{L} \times \mathfrak{P} - 2i\hbar\mathfrak{P}. \quad (\text{H.11})$$

We have with (H.9) for the first component, and similarly for the other two components,

$$(\mathfrak{L} \times \mathfrak{P})_1 - 2i\hbar P_1 = \underbrace{L_2 P_3}_{-P_3 L_2 + i\hbar P_1} - \underbrace{L_3 P_2}_{-P_2 L_3 - i\hbar P_1} - 2i\hbar P_1 = -(\mathfrak{P} \times \mathfrak{L})_1.$$

With (H.11) we can write (7.11) now

$$\mathfrak{A} = \frac{1}{Ze^2m} (\mathfrak{L} \times \mathfrak{P} - i\hbar\mathfrak{P}) + \mathfrak{R}\mathbf{R}^{-1}. \quad (\text{H.12})$$

With

$$A_1 = \frac{1}{mZe^2} (L_2 P_3 - L_3 P_2 - i\hbar P_1) + X_1 \mathbf{R}^{-1}$$

we have

$$[A_1, L_1] = \left[\frac{1}{mZe^2} (L_2 P_3 - L_3 P_2 - i\hbar P_1), L_1 \right] + [X_1 \mathbf{R}^{-1}, L_1],$$

and with (H.8)

$$[A_1, \mathbf{L}_1] = \frac{1}{mZe^2} \left(\underbrace{\mathbf{L}_2 \mathbf{P}_3 \mathbf{L}_1 - \mathbf{L}_1 \mathbf{L}_2 \mathbf{P}_3}_{[\mathbf{L}_2 \mathbf{P}_3, \mathbf{L}_1]} \underbrace{- \mathbf{L}_3 \mathbf{P}_2 \mathbf{L}_1 + \mathbf{L}_1 \mathbf{L}_3 \mathbf{P}_2}_{-[\mathbf{L}_3 \mathbf{P}_2, \mathbf{L}_1]} \right) + [X_1 \mathbf{R}^{-1}, \mathbf{L}_1],$$

and with (H.13)

$$\begin{aligned} [A_1, \mathbf{L}_1] &= \frac{1}{mZe^2} \left(\underbrace{\frac{[\mathbf{L}_2 \mathbf{P}_3, \mathbf{L}_1]}{L_2 [\mathbf{P}_3, \mathbf{L}_1] + [\mathbf{L}_2, \mathbf{L}_1] P_3} - \frac{[\mathbf{L}_3 \mathbf{P}_2, \mathbf{L}_1]}{L_3 [\mathbf{P}_2, \mathbf{L}_1] + [\mathbf{L}_3, \mathbf{L}_1] P_2}}_{\frac{i\hbar P_2}{-i\hbar L_3}} \right) + [\dots] = \\ &= \mathbf{0} + [X_1 \mathbf{R}^{-1}, \mathbf{L}_1], \end{aligned}$$

and with (H.13), (H.5), and (H.6) we finally get

$$[A_1, \mathbf{L}_1] = X_1 \underbrace{[\mathbf{R}^{-1}, \mathbf{L}_1]}_0 + \underbrace{[X_1, \mathbf{L}_1]}_0 \mathbf{R}^{-1} = \mathbf{0}.$$

Similarly we get for $k = 2$, and 3

$$[A_k, \mathbf{L}_k] = \mathbf{0}.$$

H.3 Proof of Pauli's Formula (7.15)

$$(7.15) \quad [A_i, \mathbf{L}_j] = i\hbar \varepsilon_{ijk} A_k.$$

with

$$A_1 = \frac{1}{mZe^2} (\mathbf{L}_2 \mathbf{P}_3 - \mathbf{L}_3 \mathbf{P}_2 - i\hbar \mathbf{P}_1) + X_1 \mathbf{R}^{-1}$$

we have

$$[A_1, \mathbf{L}_2] = \left[\frac{1}{mZe^2} (\mathbf{L}_2 \mathbf{P}_3 - \mathbf{L}_3 \mathbf{P}_2 - i\hbar \mathbf{P}_1), \mathbf{L}_2 \right] + [X_1 \mathbf{R}^{-1}, \mathbf{L}_2],$$

and with (H.8)

$$[A_1, \mathbf{L}_2] = \frac{1}{mZe^2} \left(\underbrace{\mathbf{L}_2 \mathbf{P}_3 \mathbf{L}_2 - \mathbf{L}_2 \mathbf{L}_2 \mathbf{P}_3}_{[\mathbf{L}_2 \mathbf{P}_3, \mathbf{L}_2]} \underbrace{- \mathbf{L}_3 \mathbf{P}_2 \mathbf{L}_2 + \mathbf{L}_2 \mathbf{L}_3 \mathbf{P}_2}_{-[\mathbf{L}_3 \mathbf{P}_2, \mathbf{L}_2]} \underbrace{- i\hbar \mathbf{P}_1 \mathbf{L}_2 + i\hbar \mathbf{L}_2 \mathbf{P}_1}_{-i\hbar [\mathbf{P}_1, \mathbf{L}_2]} \right) +$$

$$+[X_1 \mathbf{R}^{-1}, \mathbf{L}_2],$$

and with (H.13)

$$\begin{aligned} [\mathbf{A}_1, \mathbf{L}_2] &= \frac{1}{mZe^2} \left(\begin{array}{ccc} \underbrace{[\mathbf{L}_2 \mathbf{P}_3, \mathbf{L}_2]}_{\mathbf{L}_2 \underbrace{[\mathbf{P}_3, \mathbf{L}_2]}_{-i\hbar P_1} + \underbrace{[\mathbf{L}_2, \mathbf{L}_2]}_0 \mathbf{P}_3} & - & \underbrace{[\mathbf{L}_3 \mathbf{P}_2, \mathbf{L}_2]}_{\mathbf{L}_3 \underbrace{[\mathbf{P}_2, \mathbf{L}_2]}_0 + \underbrace{[\mathbf{L}_3, \mathbf{L}_2]}_{-i\hbar L_1} \mathbf{P}_2} & -(i\hbar)^2 \mathbf{P}_3 \\ \end{array} \right) + [\dots] = \\ &= \frac{i\hbar}{mZe^2} (-\mathbf{L}_2 \mathbf{P}_1 + \mathbf{L}_1 \mathbf{P}_2 - i\hbar \mathbf{P}_3) + [X_1 \mathbf{R}^{-1}, \mathbf{L}_2]. \end{aligned}$$

With (H.13), (H.5), and (H.6) we finally get

$$\underline{[\mathbf{A}_1, \mathbf{L}_2]} = \frac{i\hbar}{mZe^2} \left(\underbrace{-\mathbf{L}_2 \mathbf{P}_1 + \mathbf{L}_1 \mathbf{P}_2 - i\hbar \mathbf{P}_3}_{(\mathfrak{L} \times \mathfrak{P})_3} \right) + X_1 \underbrace{[\mathbf{R}^{-1}, \mathbf{L}_2]}_0 + \underbrace{[X_1, \mathbf{L}_2]}_{i\hbar X_3} \mathbf{R}^{-1} = \underline{i\hbar \mathbf{A}_3}.$$

Similarly we get

$$\underline{[\mathbf{A}_i, \mathbf{L}_j]} = i\hbar \varepsilon_{ijk} \mathbf{A}_k.$$

H.4 Proof of Pauli's Formula (7.16)

$$(7.16) \quad \mathfrak{A} \cdot \mathfrak{L} = \mathfrak{L} \cdot \mathfrak{A} = 0.$$

We have

$$\mathfrak{A} \cdot \mathfrak{L} = \left\{ \frac{1}{Z e^2 m} ((\mathfrak{L} \times \mathfrak{P}) - i\hbar \mathfrak{P}) + (\mathfrak{R} \mathbf{R}^{-1}) \right\} \cdot \mathfrak{L} \quad (\text{H.13})$$

The angular momentum vector matrix \mathfrak{L} is orthogonal to each of the three terms in (H.13). To see this, we first look for

$$\begin{aligned} (\mathfrak{L} \times \mathfrak{P}) \cdot \mathfrak{L} &= \begin{pmatrix} \mathbf{L}_2 \mathbf{P}_3 - \mathbf{L}_3 \mathbf{P}_2 \\ \mathbf{L}_3 \mathbf{P}_1 - \mathbf{L}_1 \mathbf{P}_3 \\ \mathbf{L}_1 \mathbf{P}_2 - \mathbf{L}_2 \mathbf{P}_1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \mathbf{L}_3 \end{pmatrix} = \\ &= \underbrace{\mathbf{L}_2 \mathbf{P}_3}_{P_3 L_2 - i\hbar P_1} \mathbf{L}_1 - \underbrace{\mathbf{L}_1 \mathbf{P}_3}_{P_3 L_1 - i\hbar P_2} \mathbf{L}_2 + \underbrace{\mathbf{L}_3 \mathbf{P}_1 \mathbf{L}_2 - \mathbf{L}_2 \mathbf{P}_1 \mathbf{L}_3}_{-i\hbar P_1 L_1 + i\hbar P_3 L_3 - i\hbar P_2 L_2} + \\ &+ \underbrace{\mathbf{L}_1 \mathbf{P}_2 \mathbf{L}_3 - \mathbf{L}_3 \mathbf{P}_2 \mathbf{L}_1}_{-i\hbar P_2 L_2 + i\hbar P_1 L_1 + i\hbar P_3 L_3} = -i\hbar \mathbf{P}_1 \mathbf{L}_1 - i\hbar \mathbf{P}_2 \mathbf{L}_2 - i\hbar \mathbf{P}_3 \mathbf{L}_3 = -i\hbar \mathfrak{P} \cdot \mathfrak{L}. \end{aligned} \quad (\text{H.14})$$

This is the same as the second term in (H.13). But for this term we get

$$\mathfrak{P} \cdot \mathfrak{L} = \mathfrak{P} \cdot (\mathfrak{R} \times \mathfrak{P}) = \begin{pmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{P}_3 \end{pmatrix} \cdot \begin{pmatrix} X_2 \mathbf{P}_3 - X_3 \mathbf{P}_2 \\ X_3 \mathbf{P}_1 - X_1 \mathbf{P}_3 \\ X_1 \mathbf{P}_2 - X_2 \mathbf{P}_1 \end{pmatrix},$$

and because X_i commutes with \mathbf{P}_k for $i \neq k$, we finally get

$$\mathfrak{P} \cdot \mathfrak{L} = \mathbf{0}. \quad (\text{H.15})$$

The same is true for the third term in (H.13), inasmuch as X_i commutes with X_k ,

$$\mathfrak{R} \cdot \mathfrak{L} = \mathfrak{R} \cdot (\mathfrak{R} \times \mathfrak{P}) = \mathbf{0}. \quad (\text{H.16})$$

H.5 Proof of Pauli's Formula (7.17)

$$(7.17) \quad \mathfrak{A} \times \mathfrak{A} = -i\hbar \frac{2}{mZ^2e^4} \mathfrak{L}E.$$

With

$$\mathfrak{A} = \frac{1}{Ze^2m} \frac{1}{2} (\mathfrak{L} \times \mathfrak{P} - i\hbar \mathfrak{P}) + \mathfrak{R} \mathbf{R}^{-1},$$

$$\mathfrak{B} \stackrel{\text{def}}{=} \frac{1}{2} (\mathfrak{L} \times \mathfrak{P} - i\hbar \mathfrak{P}),$$

and

$$\mathfrak{A} = \frac{1}{Ze^2m} \mathfrak{B} + \mathfrak{R} \mathbf{R}^{-1},$$

we have

$$\mathfrak{A} \times \mathfrak{A} = \frac{1}{Z^2e^4m^2} (\mathfrak{B} \times \mathfrak{B}) + \frac{1}{Ze^2m} \{(\mathfrak{R} \mathbf{R}^{-1}) \times (\mathfrak{L} \times \mathfrak{P}) + (\mathfrak{L} \times \mathfrak{P}) \times (\mathfrak{R} \mathbf{R}^{-1})\}. \quad (\text{H.17})$$

First, we obtain for $(\mathfrak{B} \times \mathfrak{B})$

$$\mathfrak{B} \times \mathfrak{B} = ((\mathfrak{L} \times \mathfrak{P}) \times (\mathfrak{L} \times \mathfrak{P})) - i\hbar((\mathfrak{L} \times \mathfrak{P}) \times \mathfrak{P}) - i\hbar(\mathfrak{P} \times (\mathfrak{L} \times \mathfrak{P})). \quad (\text{H.18})$$

and for the first component

$$\begin{aligned} ((\mathfrak{L} \times \mathfrak{P}) \times (\mathfrak{L} \times \mathfrak{P}))_1 &= (\mathfrak{L} \times \mathfrak{P})_2 (\mathfrak{L} \times \mathfrak{P})_3 - (\mathfrak{L} \times \mathfrak{P})_3 (\mathfrak{L} \times \mathfrak{P})_2 = \\ &= [(\mathfrak{L} \times \mathfrak{P})_2, (\mathfrak{L} \times \mathfrak{P})_3] = [\mathbf{L}_3 \mathbf{P}_1 - \mathbf{L}_1 \mathbf{P}_3, \mathbf{L}_1 \mathbf{P}_2 - \mathbf{L}_2 \mathbf{P}_1] = \\ &= (\mathbf{L}_3 \mathbf{P}_1 - \mathbf{L}_1 \mathbf{P}_3)(\mathbf{L}_1 \mathbf{P}_2 - \mathbf{L}_2 \mathbf{P}_1) - (\mathbf{L}_1 \mathbf{P}_2 - \mathbf{L}_2 \mathbf{P}_1)(\mathbf{L}_3 \mathbf{P}_1 - \mathbf{L}_1 \mathbf{P}_3). \end{aligned}$$

With (7.13), (H.8), and (H.9) we obtain

$$((\mathfrak{L} \times \mathfrak{P}) \times (\mathfrak{L} \times \mathfrak{P}))_1 = -i\hbar \mathbf{L}_1 \mathbf{P}_1^2 - i\hbar \mathbf{L}_1 \mathbf{P}_2^2 - i\hbar \mathbf{L}_1 \mathbf{P}_3^2 = -i\hbar \mathbf{L}_1 \mathfrak{P}^2, \quad (\text{H.19})$$

and generally

$$((\mathfrak{L} \times \mathfrak{P}) \times (\mathfrak{L} \times \mathfrak{P})) = -i\hbar \mathfrak{L} \mathfrak{P}^2. \quad (\text{H.20})$$

Next, we obtain for the first component of the second and third terms in (H.18)

$$\begin{aligned} & ((\mathfrak{L} \times \mathfrak{P}) \times \mathfrak{P})_1 + (\mathfrak{P} \times (\mathfrak{L} \times \mathfrak{P}))_1 = \\ &= \{(\mathfrak{L} \times \mathfrak{P})_2 \mathbf{P}_3 - (\mathfrak{L} \times \mathfrak{P})_3 \mathbf{P}_2\} + \{\mathbf{P}_2 (\mathfrak{L} \times \mathfrak{P})_3 - \mathbf{P}_3 (\mathfrak{L} \times \mathfrak{P})_2\} = \\ &= [\underbrace{(\mathfrak{L} \times \mathfrak{P})_2}_{\mathbf{L}_3 \mathbf{P}_1 - \mathbf{L}_1 \mathbf{P}_3}, \mathbf{P}_3] - [\underbrace{(\mathfrak{L} \times \mathfrak{P})_3}_{\mathbf{L}_1 \mathbf{P}_2 - \mathbf{L}_2 \mathbf{P}_1}, \mathbf{P}_2] = \mathbf{0}. \end{aligned}$$

Finally, we need to examine the last term in (H.17)

$$(\mathfrak{R} \mathbf{R}^{-1}) \times (\mathfrak{L} \times \mathfrak{P}) + (\mathfrak{L} \times \mathfrak{P}) \times (\mathfrak{R} \mathbf{R}^{-1}) \stackrel{\text{def}}{=} \mathfrak{C}.$$

With

$$\mathfrak{B} \stackrel{\text{def}}{=} \mathfrak{L} \times \mathfrak{P}$$

we have for the first component \mathbf{C}_1 of \mathfrak{C}

$$\begin{aligned} \mathbf{C}_1 &= ((\mathfrak{R} \mathbf{R}^{-1}) \times \mathfrak{B})_1 + (\mathfrak{B} \times (\mathfrak{R} \mathbf{R}^{-1}))_1 = \\ &= (\mathbf{X}_2 \mathbf{R}^{-1} \mathbf{B}_3 - \mathbf{X}_3 \mathbf{R}^{-1} \mathbf{B}_2) + (\mathbf{B}_2 \mathbf{X}_3 \mathbf{R}^{-1} - \mathbf{B}_3 \mathbf{X}_2 \mathbf{R}^{-1}) = \\ &= [\mathbf{X}_2 \mathbf{R}^{-1}, \mathbf{B}_3] - [\mathbf{X}_3 \mathbf{R}^{-1}, \mathbf{B}_2], \end{aligned}$$

and with (H.13) and (3.27)

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{X}_2 [\mathbf{R}^{-1}, \mathbf{B}_3] + \underbrace{[\mathbf{X}_2, \mathbf{B}_3]}_{i\hbar \frac{\partial \mathbf{B}_3}{\partial \mathbf{P}_2}} \mathbf{R}^{-1} - \mathbf{X}_3 [\mathbf{R}^{-1}, \mathbf{B}_2] - \underbrace{[\mathbf{X}_3, \mathbf{B}_2]}_{i\hbar \frac{\partial \mathbf{B}_2}{\partial \mathbf{P}_3}} \mathbf{R}^{-1} = \\ &= i\hbar \frac{\partial \mathbf{B}_3}{\partial \mathbf{P}_2} \mathbf{R}^{-1} - i\hbar \frac{\partial \mathbf{B}_2}{\partial \mathbf{P}_3} \mathbf{R}^{-1} + \mathbf{X}_2 [\mathbf{R}^{-1}, \mathbf{L}_1 \mathbf{P}_2 - \mathbf{L}_2 \mathbf{P}_1] - \mathbf{X}_3 [\mathbf{R}^{-1}, \mathbf{L}_1 \mathbf{P}_3 - \mathbf{L}_3 \mathbf{P}_1] \\ &\quad - i\hbar \frac{\partial \mathbf{R}^{-1}}{\partial \mathbf{X}_3} \mathbf{X}_2 + i\hbar \frac{\partial \mathbf{R}^{-1}}{\partial \mathbf{X}_2} \mathbf{X}_3, \end{aligned}$$

or

$$\mathbf{C}_1 = i\hbar \frac{\partial \mathbf{B}_3}{\partial \mathbf{P}_2} \mathbf{R}^{-1} - i\hbar \frac{\partial \mathbf{B}_2}{\partial \mathbf{P}_3} \mathbf{R}^{-1}$$

$$+i\hbar\mathbf{L}_1\left(\frac{\partial\mathbf{R}^{-1}}{\partial X_2}X_2+\frac{\partial\mathbf{R}^{-1}}{\partial X_3}X_3\right)-i\hbar(\mathbf{L}_2X_2+\mathbf{L}_3X_3)\frac{\partial\mathbf{R}^{-1}}{\partial X_1}.$$

With $\mathfrak{LR} = \mathbf{L}_1X_1 + \mathbf{L}_2X_2 + \mathbf{L}_3X_3 = \mathbf{0}$ and $\frac{\partial\mathbf{R}^{-1}}{\partial X_k} = -X_k\mathbf{R}^{-3}$ we obtain for the last line

$$i\hbar\mathbf{L}_1(X_1^2 + X_2^2 + X_3^2)\mathbf{R}^{-3} = i\hbar\mathbf{L}_1\mathbf{R}^{-1}.$$

Furthermore, one has

$$\begin{aligned}\frac{\partial\mathbf{B}_3}{\partial\mathbf{P}_2}-\frac{\partial\mathbf{B}_2}{\partial\mathbf{P}_3} &= \frac{\partial}{\partial\mathbf{P}_2}(\mathbf{L}_1\mathbf{P}_2)+\frac{\partial}{\partial\mathbf{P}_3}(\mathbf{L}_1\mathbf{P}_3) = \\ &= 2\mathbf{L}_1+(\mathbf{P}_2\frac{\partial}{\partial\mathbf{P}_2}+\mathbf{P}_3\frac{\partial}{\partial\mathbf{P}_3})\mathbf{L}_1=3\mathbf{L}_1,\end{aligned}$$

thus in total

$$\mathbf{C}_1 = 2i\hbar\mathbf{L}_1.$$

Putting this in (H.17), we have

$$\mathfrak{A} \times \mathfrak{A} = -i\hbar\frac{2}{mZ^2e^4}\mathfrak{L}\left(\frac{1}{2m}\mathfrak{P}^2-e\mathbf{R}^{-1}\right).$$

H.6 Proof of Pauli's Formula (7.18)

$$(7.18) \quad \mathfrak{A}^2 = \frac{2}{mZ^2e^4}\mathbf{E}\left(\mathfrak{L}^2 + \frac{\hbar^2}{4\pi^2}\mathbf{I}\right) + \mathbf{I}.$$

With

$$\mathfrak{A} = \frac{1}{Ze^2m}\frac{1}{2}(\mathfrak{L} \times \mathfrak{P} - i\hbar\mathfrak{P}) + \mathfrak{R}\mathbf{R}^{-1},$$

$$\mathfrak{B} \stackrel{\text{def}}{=} \frac{1}{Ze^2m}\frac{1}{2}(\mathfrak{L} \times \mathfrak{P} - i\hbar\mathfrak{P}),$$

and

$$\mathfrak{A} = \mathfrak{B} + \mathfrak{R}\mathbf{R}^{-1},$$

we have

$$\mathfrak{A} \cdot \mathfrak{A} = \{\mathfrak{B}^2 + \mathfrak{R}\mathbf{R}^{-1}\mathfrak{B} + \mathfrak{B}\mathfrak{R}\mathbf{R}^{-1} + \mathbf{I}\} \quad (\text{H.21})$$

For the first term on the right-hand side of (H.21) we get

$$\begin{aligned}\mathfrak{B}^2 &= \frac{1}{Z^2 e^4 m^2} \frac{1}{4} (\mathfrak{L} \times \mathfrak{P} - i\hbar \mathfrak{P})^2 = \\ &= \frac{1}{Z^2 e^4 m^2} \frac{1}{4} ((\mathfrak{L} \times \mathfrak{P})^2 - i\hbar \mathfrak{P}(\mathfrak{L} \times \mathfrak{P}) - i\hbar(\mathfrak{L} \times \mathfrak{P})\mathfrak{P} - (\hbar)^2 \mathfrak{P}^2). \quad (\text{H.22})\end{aligned}$$

As with formula (H.14), these formulas can be proven.

$$(\mathfrak{L} \times \mathfrak{P})^2 = \mathfrak{P}^2 \mathfrak{L}^2, \quad (\text{H.23})$$

$$\mathfrak{P}(\mathfrak{L} \times \mathfrak{P}) = 2i\hbar \mathfrak{P}^2, \quad (\text{H.24})$$

and

$$(\mathfrak{L} \times \mathfrak{P})\mathfrak{P} = \mathbf{0}. \quad (\text{H.25})$$

With these three formulas, we have

$$\mathfrak{B}^2 = \frac{1}{Z^2 e^4 m^2} \frac{1}{4} (\mathfrak{P}^2 \mathfrak{L}^2 + 2\hbar^2 \mathfrak{P}^2 - (\hbar)^2 \mathfrak{P}^2) = \frac{1}{Z^2 e^4 m^2} \frac{1}{4} \mathfrak{P}^2 (\mathfrak{L}^2 + (\hbar)^2 \mathbf{I}). \quad (\text{H.26})$$

Next we get

$$(\mathfrak{R} \mathbf{R}^{-1}) \cdot (\mathfrak{L} \times \mathfrak{P}) = \begin{pmatrix} X_1 \mathbf{R}^{-1} \\ X_2 \mathbf{R}^{-1} \\ X_3 \mathbf{R}^{-1} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{L}_2 \mathbf{P}_3 - \mathbf{L}_3 \mathbf{P}_2 \\ \mathbf{L}_3 \mathbf{P}_1 - \mathbf{L}_1 \mathbf{P}_3 \\ \mathbf{L}_1 \mathbf{P}_2 - \mathbf{L}_2 \mathbf{P}_1 \end{pmatrix} = 2i\hbar(\mathfrak{R} \cdot \mathfrak{P}) \mathbf{R}^{-1} + \mathbf{L}^2 \mathbf{R}^{-1},$$

and

$$(\mathfrak{L} \times \mathfrak{P}) \cdot (\mathfrak{R} \mathbf{R}^{-1}) = \mathbf{L}^2 \mathbf{R}^{-1}.$$

Thus we have

$$\mathfrak{R} \mathbf{R}^{-1} \mathfrak{B} = \frac{1}{Ze^2 m} \frac{1}{2} (\mathbf{L}^2 \mathbf{R}^{-1} + 2i\hbar \mathfrak{R} \cdot \mathfrak{P}) \mathbf{R}^{-1},$$

and

$$\mathfrak{B} \mathfrak{R} \mathbf{R}^{-1} = \frac{1}{Ze^2 m} \frac{1}{2} (\mathbf{L}^2 \mathbf{R}^{-1} - 2i\hbar \mathfrak{R} \cdot \mathfrak{P}) \mathbf{R}^{-1}.$$

Putting these in (H.21) we obtain the final result

$$\mathfrak{A} \cdot \mathfrak{A} = \frac{2}{m Z^2 e^4} \left(\frac{1}{2m} \mathbf{P}^2 - Ze^2 \mathbf{R}^{-1} \right) (\mathbf{L}^2 + \hbar^2 \mathbf{I}) + \mathbf{I} = \frac{2}{m Z^2 e^4} E (\mathbf{L}^2 + \hbar^2 \mathbf{I}) + \mathbf{I}.$$

Appendix I

Physical Quantities and Units

- Bohr Magneton: $\mu_B = \frac{e\hbar}{2m_e} = 9.27400968 \cdot 10^{-24} \text{ Am}^2$ (or JT^{-1});
- Fine Structure Constant: $\alpha = \frac{e^2}{4\pi\epsilon_0 c \hbar} = \frac{e^2 c \mu_0}{2\hbar} = 7.29735357 \cdot 10^{-3} \approx \frac{1}{137}$;
- Bohr Radius: $a_0 = \frac{4\pi\epsilon_0 \hbar^2}{m_e e^2} = \frac{\hbar}{\alpha m_e c} = 5.2917721092 \cdot 10^{-11} \text{ m} \approx 52.9 \text{ pm};$

Force: $1\text{N}(Newton) = 1\text{m kg/s}^2$;

Work, Energy: $1\text{J}(Joule) = 1\text{N m}$;

Magnetic Flux Density: $1\text{T}(Tesla) = 1\text{N/(Am)}$;

Power: $1\text{W}(Watt) = 1\text{J/s}$;

Voltage: $1\text{V}(Volt) = 1\text{J/C}$;

Charge Quantity: $1\text{C}(Coulomb) = 1\text{A s}$;

Current: $1\text{A}(Ampere) = 1\text{C/s}$.

References

1. G. Aruldas, *Quantum Mechanics* (PHI Learning, New Delhi, 2012)
2. P. Atkins, R. Friedman, *Molecular Quantum Mechanics* (Oxford University Press, Oxford, 2005)
3. L.E. Ballentine, *Quantum Mechanics. A Modern Development* (World Scientific, Singapore, 1998)
4. M. Born, P. Jordan, Zur Quantenmechanik. ZS f. Physik (1925). The English translation of the paper is included as paper 13 in [27]
5. M. Born, P. Jordan, *Elementare Quantenmechanik* (Springer, German, 1930)
6. M. Born, W. Heisenberg, P. Jordan, Zur Quantenmechanik, (II). ZS f. Physik (1925). The English translation of the paper is included as paper 15 in [27]
7. J.W. Brewer, Kronecker products and matrix calculus in system theory. IEEE Trans. Circuits Syst. pp. 772–781 (1978)
8. P.A.M. Dirac, The fundamental equations of quantum mechanics. Proc. R. Soc. London, Ser. A (1925)
9. P.A.M. Dirac, *The Principles of Quantum Mechanics* (Oxford, Oxford, 1967)
10. D. Dubbers, H.-J. Stöckmann, *Quantum Physics, the Bottom-Up Approach* (Springer, Berlin, 2013)
11. S. Gasiorowicz, *Quantum Physics* (Wiley, New York, 2003)
12. H.S. Green, *Matrix Methods in Quantum Mechanics* (Barnes & Noble, New York, 1965)
13. W. Heisenberg, Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. ZS f. Physik (1925). The English translation of the paper is included as paper 12 in [27]
14. P. Jordan, *Anschauliche Quantentheorie* (Springer, German, 1936)
15. G. Ludyk, *Einstein in Matrix Form* (Springer, Berlin, 2013)
16. E.L. Michelsen, *Quirky Quantum Concepts* (Springer, Berlin, 2014)
17. J. Pade, *Quantum Mechanics for Pedestrians (1 and 2)* (Springer, Berlin, 2013)
18. W. Pauli, Über das Wasserstoffspektrum vom Standpunkt der neuen Quantenmechanik. ZS f. Physik (1926). The English translation of the paper is included as paper 16 in [27]
19. M. Razavy, *Heisenberg's Quantum Mechanics* (World Scientific, Singapore, 2011)
20. F. Scheck, *Quantum Physics* (Springer, Berlin, 2013)
21. F. Schwabl, *Quantum Mechanics* (Springer, Berlin, 2007)
22. A. Sommerfeld, *Atomic Structure and Spectral Lines* (Methuen, London, 1923)
23. W.-H. Steeb, Y. Hardy, *Quantum Mechanics Using Computer Algebra* (World Scientific, Singapore, 2010)

24. Th.F. Jordan, *Quantummechanics in Simple Matrix Form* (Dover, 1986)
25. S. Weinberg, *Lectures on Quantum Mechanics* (Cambridge, Cambridge, 2013)
26. S. Weinberg, *Dreams of a Final Theory* (Pantheon Books, New York, 1992)
27. B.L. van der Waerden, *Sources of Quantum Mechanics* (Dover, Amsterdam, 1967)
28. P.A. Schilpp, *Albert Einstein: Philosopher - Scientist* (Kohlhammer, New York, 1949)
29. G.H. Shortley, G.E. Kimball, Analysis of non-commuting vectors with application to quantum mechanics and vector calculus. Proc. N. A. S. (1934)

Index

A

Anderson, 154
Angular frequency, 47
Angular momentum, 1, 8, 59, 79, 94
Angular momentum quantum number, 68, 69
Angular momentum vector, 68
Annihilation operators, 64
Anomalous Zeeman effect, 105, 108, 109
Ansatz, 154
Atomic nucleus, 115
Atomic radius, 9
Atomic spectra, 59
Atomic theory, 77
Average, 44
Average value, 39

B

Balmer, 11, 83
Baryons, 123
 Ω -baryon, 123
Bohr, 1, 81
Bohr magneton, 86
Bohr radius, 103
Born, 23, 26, 38, 48, 59
bosons, 123
Bra, 41
Bra-ket notation, 41
Brackett, 12
Bunsen, 11

C

Clebsch–Gordan coefficient, 98, 104
Commutation relation, 26, 28, 57, 75, 81
131, 138

Commutativity, 63

Commutator, 23
Composite system, 111
Coordinate transformation, 31
Copenhagen interpretation, 175

D

Dash-notation, 151
De Broglie, 129
Degrees of freedom, 28
Density matrix, 39, 41
Determinant, 39
Diagonal elements, 24
Diagonal matrix, 24, 69
Differential equation, 140
Differentiation, 27
Dimension, 57
Dirac, 26
Dirac equation, 150
Dirac matrices, 149
Dirac sea, 153
Dirac spinor, 150
Dirac wave function, 150
 D -line, 87

E

Eigenfunction, 130, 131, 135, 151
Eigenvalue, 135
Eigenvalue equation, 38
Eigenvector, 130
Einstein, 83
Elastic oscillators, 47
Electron, 7
Electron orbit, 15
Energy, 146

Energy quanta, 6
 Energy spectrum, 51
 Ensemble, 39
 Entangled states, 115, 128
 Equation of motion
 relativistic, 145
 Equations of motion, 79
 Equilibrium position, 47
 Event, 141
 Everett, 21
 Expectation value, 39, 43, 44, 112
 Exponential function, 104
 Exponential function of a Pauli matrix, 104

F
 Fermat, 129
 Fermions, 123
 Feynman, 151
 Fourier analysis, 31
 Fourier decomposition, 31
 Fourier series, 19, 31
 Frequency matrix, 55

G
 Gamma Matrices, 154
 Gauge bosons, 123
 Generator, 88
 Gerlach, 34
 Gluons, 123
 Goudsmit, 87, 93
 Green, 59
 Gyromagnetic ratio, 86

H
 Hamilton equations, 33
 Hamilton-Jacobi theory, 131
 Hamiltonian, 2
 Hamiltonian function, 47
 harmonic oscillator, 47, 139
 Heisenberg, 15, 48, 51, 55, 59, 87, 129
 Heisenberg equation of motion, 20, 43
 Heisenberg matrix, 68, 89
 Heisenberg's picture, 134
 Helium-isotopes, 123
 Hermitian matrix, 31, 52, 53
 Hilbert space, 68, 111, 113, 122
 Hydrogen, 103
 Hydrogen atom, 77
 Hydrogen-like atom, 103
 Hydrogen spectrum, 11

I
 Inertial frame, 141
 Inner angular momentum, 87
 Isomorphism, 131

J
 Jordan, 23, 26, 48, 59

K
 Ket, 41
 Kinetic energy, 2
 Kirchhoff, 11
 Klein-Gordon equation, 148
 Kronecker-Delta function, 72
 Kronecker product, 29, 78, 111
 Kronecker sum, 113, 114

L
 Laplace–Runge–Lenz vector, 33, 79
 Larmor, 101
 Larmor interaction, 102
 Leibniz formula, 40
 Lenz-Matrix-Vector, 79
 Leptons, 123
 Likelihood, 19
 Linear operator, 130
 Lithium, 87
 Lorentz, 85
 Lorentz factor, 101
 Lorentz invariant, 144
 Lorentz matrix, 145
 Lorentz transformation, 142

M
 Magnetic dipole moment, 99, 100
 Magnetic field, 34, 64, 85, 99
 Magnetic moment, 98
 Magnetic quantum number, 68
 Manifold, 5
 Mathematical model, 37
 Matrix mechanics, 59
 Matrix vector, 59, 75, 77
 Maupertuis-Euler principle, 129
 Measurement, 38
 Measurements results, 36
 Mesons, 123
 Minkowski matrix, 142
 Mixed state, 41
 Momentum, 144
 Momentum matrix, 90
 Momentum vector, 144

N

Neutron, 115
Normal Zeeman effect, 108
Nucleus, 1, 98

O

Observables, 36
One-electron atom, 86, 87
Operator equation, 131
Orthonormal basis, 112
Oscillation frequency, 3
Oscillator, 3

P

Parallel universe, 21
Partial differential equation, 131
Particle, 111
Paschen, 12
Paschen-Back effect, 109
Pauli, 77, 83, 87
Pauli matrix, 93, 119
Pauli Principle, 116, 125
Periodic function, 31
Permutation, 37
Permutation matrix, 37, 119, 121
Permutation operator, 126, 127
Permutation relationship, 116
Pfund, 12
Photon, 123
Planck, 83
Planck's constant, 1, 59
Positron, 154
Postulates, 35
Potassium, 87
Principal quantum number, 82
Probabilistic interpretation, 38
Probability, 36, 38
Projection matrix, 36, 37
Projection operator, 37
Proton, 115
Pure state, 40, 41

Q

Quantization rule, 1
Quantum communication, 115
Quantum computing, 115
Quantum electrodynamics, 100
Quantum number, 68, 69
Quark, 123

R

Radiation, 7
Relative momentum, 62
Relative position, 62
Rest energy, 146
Rest mass, 146
Ritz, 12
Rotation, 91
Rotator, 5
Rydberg, 12, 83
Rydberg constant, 83

S

Schrödinger, 129
Schrödinger equation, 131, 139, 147
Schrödinger's Cat, 140
Schrödinger's wave equation, 134
Selection condition, 7
Separable states, 114
Silver atoms, 34
Slater determinant, 127
Sodium, 87
Sommerfeld, 1
Space quantization, 69
Spacetime, 141
Spectral line, 1, 87
Spectrum of hydrogen, 83
Spin, 85, 87, 91
Spin angular momentum, 93, 100
Spin down, 34
Spin states, 153
Spin up, 34
Spin-orbit coupling, 98
Spinors, 119
Spring, 47
State matrix, 39
State plane, 5
State representation, 38
State vector, 33
Stationary state, 24
Stern, 34
Stern-Gerlach, 43
Stern-Gerlach apparatus, 37
Stern-Gerlach experiment, 34
Strong magnetic field, 109
Symmetry consideration, 87
Symmetry operation, 91
Symmetric state, 126

T

- Thomas, 101
Thomas half, 102
Thomas precession correction, 101
Three men work, 59
Time evolution, 43
Total angular momentum, 91
Total energy, 2
Total spin-orbit potential, 102
Trace of a matrix, 40
Trajectory, 5
Transition probability, 14
Transposition matrix, 117, 118
Two-particle system, 112

U

- Uhlenbeck, 87, 93
Uncertainty principle, 44
Uncertainty relation, 45

V

- Variance, 44
Velocity vector, 143

W

- Wave, 129
Wave function, 132
Wave vector, 129
Wavelength, 8, 11
Winding number, 7

Z

- Zeeman, 85
Zeeman effect, 85, 105, 109
Zeeman splitting, 86
Zero-point energy, 57