

Notes on Physics

by Jonas Petersen

Til min søn. Denne bog indeholder dybdegående svar til mange af de spørgsmål du stillede som barn. Jeg har skrevet bogen over mange år og gemt i teksten indeholder den meget af din fars sjæl. Med bogen har du muligheden for altid at have en del af din far med hvorend du bevæger dig hen i livet. Far elsker dig – Altid og uanset hvad.

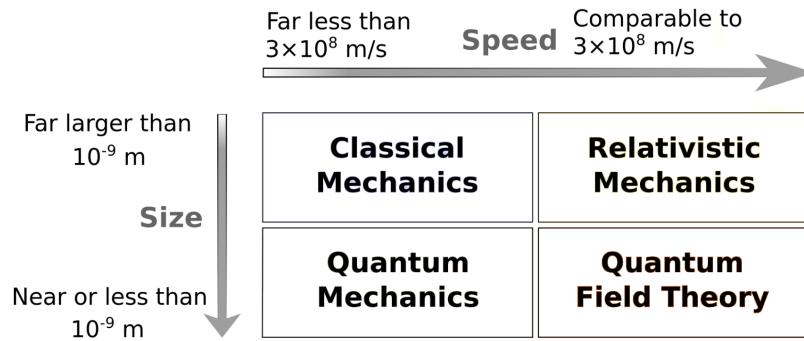
Contents

1 Overture	1
1.1 Notation	1
I Classical Mechanics	7
2 Newtonian Mechanics	9
2.1 Newtons laws	9
2.2 Non-inertial reference frames	11
2.3 Center of mass	12
2.4 Energy	13
2.5 Angular momentum in classical mechanics	16
2.6 Navier-Stokes Equation	24
2.6.1 Computational Fluid Dynamics	27
3 Analytical Mechanics	30
3.1 Lagrangian Mechanics	30
3.2 Hamiltonian Mechanics	35
3.3 Hamilton-Jacobi Mechanics	37
3.3.1 Identifying F_2 with the action	42
4 Thermodynamics	43
4.1 Functions of state	43
4.2 The laws of thermodynamics	44
4.3 Manipulation of thermodynamic potentials	46
4.4 System stability	47
II Quantum Mechanics	49
5 Mathematical setup	51
5.1 The vector space of quantum mechanics	51
5.2 Matrix notation	55
6 Generalization of the classical schemes	62
6.1 Symmetry operators	62
6.2 The Schroedinger Scheme	64
6.3 The Heisenberg Scheme	69
6.4 The Path Integral Scheme	71
7 Angular momentum in quantum mechanics	73
7.1 Eigenvalues and eigenstates of angular momentum	74

8 Approximation methods	85
8.1 Perturbation theory in QM	85
8.2 The WKB approximation	92
III Relativistic Mechanics	118
9 Group Theory and Special Relativity	121
9.1 $SO(3)$	125
9.2 $SU(2)$	126
9.2.1 Relationship between $SO(3)$ and $SU(2)$	126
9.3 The Lorentz group	127
9.4 The Poincare' group	132
10 Classical field theory	133
10.1 The Lagrangian formulation of classical field theory	133
10.2 Irreducible (finite) representations of the Lorentz group	134
10.2.1 Construction a Lagrangian density	138
10.2.2 The Euler-Lagrange equation in field theory	141
10.3 Noethers theorem	144
11 General Relativity	149
11.1 The geodesic equation via the principle of equivalence	149
11.2 The geodesic equation via the principle of general covariance	151
11.3 The geodesic equation via the variational principle	153
11.4 The Einstein equation	156
11.5 Standard Cosmology	158
IV Quantum Field Theory applied to particle physics	160
12 Lagrangian Density of the SM	163
12.1 Canonical Quantization	163
12.2 Path Integral Quantization	170
12.3 Yang-Mills theory	176
12.4 Construction of SM Lagrangian Density	180
13 Observables of Particle Physics	184
13.1 Massive scattering	184
13.2 Massless scattering	186
13.3 $1 \Rightarrow 2$ decay	187
13.4 The Invariant Matrix Element	188
13.5 Effective field theory	203
Appendix A Code for Computational Fluid Dynamics	208
Appendices	
Appendix B Hamiltonian Monte Carlo	223
Appendix C Nested Sampling	227
Bibliography	230

1 | Overture

This book contains the central elements of my physics degree. It is written during my studies and finished after my PhD. Roughly speaking the book is divided into five parts; classical mechanics (large object moving at $v \ll c$), relativistic mechanics (large objects moving at $v \sim c$), quantum mechanics (small objects moving at $v \ll c$), quantum field theory (small objects moving at $v \sim c$) and lastly statistics.



1.1 Notation

Units: Since this book will cover many different fields of physics will vary slightly throughout. As a rule of thumb; In the non-relativistic limit ($v \ll c$) SI units will be employed whereas in the relativistic limit ($v \simeq c$) natural units ($\hbar = k_B = c = 1$) will be utilized.

Vectors: Apart from the statistics part, a three vector will be denoted by a vector arrow, eg. \vec{x} , whereas a four vector will most often have no specific notation, eg. $x = x^\mu \doteq (x^0, x^1, x^2, x^3)^T$. A four vector with index up, eg. x^μ , is a contravariant vector whereas a four vector with index down, eg. x_μ , is a covariant vector (a deeper explanation and definition is found in section 9.3).

Derivatives: The partial derivative is denoted by $\partial_\mu = \frac{\partial}{\partial x^\mu}$ and the (Minkowski space) d'Alembertian is $\square = \partial_\mu \partial^\mu = \partial_0^2 - \vec{\nabla}^2$. The symbol $\check{\partial}_\mu$ is defined by $f \check{\partial}_\mu g = f \partial_\mu g - (d_\mu f)g$.

Metric: $\eta_{\mu\nu}^{(qft)} \doteq \text{diag}(1, -1, -1, -1)$ will be used in quantum field theory¹ and $\eta_{\mu\nu}^{(sr)} \doteq \text{diag}(-1, 1, 1, 1)$ will be used in special relativity². Under normal conditions it will be quite clear which metric is used, and so the superscripts will be suppressed. A generic metric, valid in spaces which differ from Minkowski (special relativity), is denoted by $g_{\mu\nu}$. Often $g_{\mu\nu}$ and $\eta_{\mu\nu}$ will be used interchangeably.

Dirac matrices: Dirac matrices, γ^μ , satisfy

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu}. \quad (1.1)$$

¹Such that $x \cdot x = x_{(0)}x^{(0)} - (x_{(1)}x^{(1)} + x_{(2)}x^{(2)} + x_{(3)}x^{(3)})$.

²Such that $x \cdot x = -x_{(0)}x^{(0)} + (x_{(1)}x^{(1)} + x_{(2)}x^{(2)} + x_{(3)}x^{(3)})$.

Therefore $(\gamma^0)^2 = I \doteq 1$ and $(\gamma^i) = -I \doteq -1$. γ^0 is hermitian whereas γ^i is anti-hermitian, i.e.

$$(\gamma^0)^\dagger = \gamma^0, \quad (\gamma^i)^\dagger = -\gamma^i \Leftrightarrow (\gamma^\mu)^\dagger = \gamma^0 \gamma^\mu \gamma^0, \quad (1.2)$$

where $\dagger = *T$ in matrix notation. The Dirac matrices satisfy the contraction identities

$$\begin{aligned} \gamma^\mu \gamma_\mu &= \eta_{\mu\nu} \gamma^\mu \gamma^\nu = \frac{1}{2} g_{\mu\nu} \{\gamma^\mu, \gamma^\nu\} = \eta_{\mu\nu} \eta^{\mu\nu} = 4, \\ \gamma^\mu \gamma^\nu \gamma_\mu &= -2\gamma^\nu, \\ \gamma^\mu \gamma^\nu \gamma^\alpha \gamma_\mu &= 4\eta^{\nu\alpha}, \\ \gamma^\mu \gamma^\nu \gamma^\alpha \gamma^\beta \gamma_\mu &= -2\gamma^\beta \gamma^\alpha \gamma^\nu, \end{aligned} \quad (1.3)$$

where Einstein summations is used (repeated indices are summed implicitly). The matrix γ^5 is defined as follows

$$\gamma^5 \equiv i\gamma^0 \gamma^1 \gamma^2 \gamma^3, \quad (1.4)$$

where

$$(\gamma^5)^2 = I \doteq 1, \quad (\gamma^5)^\dagger = \gamma^5, \quad \{\gamma^5, \gamma^\mu\} = 0. \quad (1.5)$$

The commutator of the Dirac matrices is defined as follows

$$[\gamma^\mu, \gamma^\nu] = \gamma^\mu \gamma^\nu - \gamma^\nu \gamma^\mu \equiv -2i\sigma^{\mu\nu}, \quad (1.6)$$

where $\sigma^{\mu\nu}$ is a rank 2 tensor. The Dirac matrices obey the following trace identities

$$\begin{aligned} Tr(I) &= Tr(1) = 4, \\ Tr(\gamma^\mu) &= 0, \\ Tr(\gamma^5) &= 0, \\ Tr(\text{odd } \# \text{ of } \gamma\text{'s}) &= 0, \\ Tr(\gamma^\mu \gamma^\nu) &= 4\eta^{\mu\nu}, \\ Tr(\gamma^\mu \gamma^\nu \gamma^\alpha \gamma^\beta \dots) &= Tr(\dots \gamma^\beta \gamma^\alpha \gamma^\nu \gamma^\mu), \\ Tr(\gamma^5 \gamma^\mu \gamma^\nu) &= 0, \\ Tr(\gamma^\mu \gamma^\nu \gamma^\alpha \gamma^\beta) &= 4(\eta^{\mu\nu} \eta^{\alpha\beta} - \eta^{\mu\alpha} \eta^{\nu\beta} + \eta^{\mu\beta} \eta^{\nu\alpha}), \\ Tr(\gamma^5 \gamma^\mu \gamma^\nu \gamma^\alpha \gamma^\beta) &= -4i\varepsilon^{\mu\nu\alpha\beta}, \end{aligned} \quad (1.7)$$

where ε^{\dots} is the Levi-Civita tensor which obeys

$$\begin{aligned} \varepsilon^{0123} &= 1 \\ \varepsilon^{0132} &= -1 \\ \varepsilon^{\mu\nu\alpha\beta} \varepsilon_{\mu\nu\alpha\beta} &= -24, \\ \varepsilon^{\mu\nu\alpha\beta} \varepsilon_{\mu\nu\alpha\sigma} &= -6\delta_\sigma^\beta, \\ \varepsilon^{\mu\nu\alpha\beta} \varepsilon_{\mu\nu\rho\sigma} &= -2(\delta_\rho^\alpha \delta_\sigma^\beta - \delta_\sigma^\alpha \delta_\rho^\beta). \end{aligned} \quad (1.8)$$

The Dirac matrices can be represented by several, equivalent, matrix representations (there is an ambiguity in the definition from the Dirac equation). Two often used representations are the Weyl representation (also called the chiral representation) and the standard representation

$$\begin{aligned} \text{Weyl rep.: } \gamma^0 &\doteq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, & \gamma^i &\doteq \begin{bmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{bmatrix}, & \gamma^5 &\doteq \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \\ \text{Std. rep.: } \gamma^0 &\doteq \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, & \gamma^i &\doteq \begin{bmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{bmatrix}, & \gamma^5 &\doteq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \end{aligned} \quad (1.9)$$

The above matrices are in block diagonal form. The Dirac matrices are 4×4 matrices, so each block is 2×2 . σ^i are called the Pauli matrices and are defined viz

$$\sigma^1 \doteq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 \doteq \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^3 \doteq \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (1.10)$$

The Pauli matrices satisfy

$$\sigma^i \sigma^j = \delta^{ij} + i \varepsilon^{ijk} \sigma^k. \quad (1.11)$$

In relation to the Pauli matrices two 4-vectors are defined

$$\sigma^\mu \equiv (1, \sigma^i), \quad \bar{\sigma}^\mu \equiv (1, -\sigma^i). \quad (1.12)$$

Slash and bar notation: Feynman slash-notation is defined by

$$\not{A} \equiv A_\mu \gamma^\mu = \gamma^\mu A_\mu = \gamma_\mu A^\mu = A^\mu \gamma_\mu. \quad (1.13)$$

Often the partial derivative will appear in slashed notation, i.e. $\not{\partial} = \gamma^\mu \partial_\mu$. The barred notation is defined by

$$\bar{A} = A^\dagger \gamma^0_{Weyl}, \quad (1.14)$$

where it is clear that the γ^0 to be used is only the one from the Weyl representation. The function of this γ^0 is to swap the upper and lower components in a 2×1 matrix and this is what is needed. The γ^0 from the standard representation does not do this. Note however that henceforth the subscript on the γ^0 will be omitted and it is understood that for barred quantities (other than $\bar{\sigma}$ which is poorly defined in this notation) the relevant γ^0 is the one from the Weyl representation.

Dirac Spinors: The Dirac spinors are defined as follows

$$\begin{aligned} u^s(\vec{p}) &= \begin{bmatrix} \sqrt{p \cdot \sigma} \xi^s \\ \sqrt{p \cdot \bar{\sigma}} \xi^s \end{bmatrix}, & \bar{u}^s(\vec{p}) &= [\xi^s \sqrt{p \cdot \sigma} \quad \xi^s \sqrt{p \cdot \bar{\sigma}}] \gamma^0, \\ v^{s'}(\vec{k}) &= \begin{bmatrix} \sqrt{k \cdot \sigma} \eta^{s'} \\ -\sqrt{k \cdot \bar{\sigma}} \eta^{s'} \end{bmatrix}, & \bar{v}^{s'}(\vec{k}) &= [\eta^{s'} \sqrt{k \cdot \sigma} \quad -\eta^{s'} \sqrt{k \cdot \bar{\sigma}}] \gamma^0, \end{aligned} \quad (1.15)$$

where³ $\xi^{\dagger s'} \xi^s = \delta^{ss'}$, $\eta^{\dagger s'} \eta^s = \delta^{ss'}$, $\xi^{\dagger s'} \eta^s = \delta^{ss'}$ and the $\sqrt{p \cdot \sigma}$ -like expressions are to be understood as the square root of the positive eigenvalue of the matrix $p \cdot \sigma$. Hence, the $\sqrt{p \cdot \sigma}$ -like expressions are numbers and can be moved around as such. The $\sqrt{p \cdot \sigma}$ -like expressions obey

$$\begin{aligned} \sqrt{p \cdot \sigma} &= \sqrt{p^0 - \vec{\sigma} \cdot \vec{p}} = \sqrt{\text{eigenvals}(p^0 I - \vec{\sigma} \cdot \vec{p})} \\ &= \begin{bmatrix} \sqrt{p^0 - |\vec{p}|} & 0 \\ 0 & \sqrt{p^0 + |\vec{p}|} \end{bmatrix}, \\ \sqrt{p \cdot \bar{\sigma}} &= \sqrt{p^0 + \vec{\sigma} \cdot \vec{p}} = \sqrt{\text{eigenvals}(p^0 I + \vec{\sigma} \cdot \vec{p})} \\ &= \begin{bmatrix} \sqrt{p^0 + |\vec{p}|} & 0 \\ 0 & \sqrt{p^0 - |\vec{p}|} \end{bmatrix}. \end{aligned} \quad (1.16)$$

³The identity $\xi^{\dagger s'} \eta^s = \delta^{ss'}$ encourages the identification $\xi = \eta$, just as Peskin and Schroeder (1995)[p. 65] has done.

Since there are several eigenvalues which one is taken depends on the vector (ξ, η) which is multiplied (see Lancaster and Blundell 2014[p. 329]). For the Dirac spinors

$$\begin{aligned}
u^{\dagger s}(\vec{p})u^{s'}(\vec{k}) &= \begin{cases} 2p^0\delta^{ss'} & \text{if } p = k \\ \xi^{\dagger s'}[\sqrt{p \cdot \sigma}\sqrt{k \cdot \sigma} + \sqrt{p \cdot \bar{\sigma}}\sqrt{k \cdot \bar{\sigma}}]\xi^{s'} & \text{if } p \neq k \end{cases}, \\
v^{\dagger s}(\vec{p})v^{s'}(\vec{k}) &= \begin{cases} 2p^0\delta^{ss'} & \text{if } p = k \\ \eta^{\dagger s'}[\sqrt{p \cdot \sigma}\sqrt{k \cdot \sigma} + \sqrt{p \cdot \bar{\sigma}}\sqrt{k \cdot \bar{\sigma}}]\eta^{s'} & \text{if } p \neq k \end{cases}, \\
\bar{u}^s(\vec{p})u^{s'}(\vec{k}) &= \begin{cases} 2m\delta^{ss'} & \text{if } p = k \\ \xi^{\dagger s'}[\sqrt{p \cdot \sigma}\sqrt{k \cdot \sigma} + \sqrt{p \cdot \bar{\sigma}}\sqrt{k \cdot \bar{\sigma}}]\xi^{s'} & \text{if } p \neq k \end{cases}, \\
\bar{v}^s(\vec{p})v^{s'}(\vec{k}) &= \begin{cases} -2m\delta^{ss'} & \text{if } p = k \\ -\eta^{\dagger s'}[\sqrt{p \cdot \sigma}\sqrt{k \cdot \sigma} + \sqrt{p \cdot \bar{\sigma}}\sqrt{k \cdot \bar{\sigma}}]\eta^{s'} & \text{if } p \neq k \end{cases}, \\
\bar{u}^s(\vec{p})v^{s'}(\vec{k}) &= \begin{cases} 0 & \text{if } p = k \\ \xi^{\dagger s}[\sqrt{p \cdot \bar{\sigma}}\sqrt{k \cdot \sigma} - \sqrt{p \cdot \sigma}\sqrt{k \cdot \bar{\sigma}}]\eta^{s'} & \text{if } p \neq k \end{cases}, \\
\bar{v}^s(\vec{p})u^{s'}(\vec{k}) &= \begin{cases} 0 & \text{if } p = k \\ \eta^{\dagger s}[\sqrt{p \cdot \bar{\sigma}}\sqrt{k \cdot \sigma} - \sqrt{p \cdot \sigma}\sqrt{k \cdot \bar{\sigma}}]\xi^{s'} & \text{if } p \neq k \end{cases}.
\end{aligned} \tag{1.17}$$

Spin sums

$$\sum_{s=1,2} u^s(\vec{p})\bar{u}^s(\vec{p}) = \not{p} + m, \quad \sum_{s=1,2} v^s(\vec{p})\bar{v}^s(\vec{p}) = \not{p} - m. \tag{1.18}$$

The solutions to Diracs equation should also be solutions to Klein-Gordons equation. The solutions to Klein-Gordons equation are plane waves, so $\psi \propto e^{\pm ip \cdot x}$ for Diracs equation. The " \propto " covers over $u^s(\vec{p}), v^s(\vec{p})$ (either of them). Using solutions on this form in Diracs equation results in a reduced equation from which (see equation (12.27))

$$\begin{aligned}
u^s(\vec{p}) &= \frac{\not{p}}{m}u^s(\vec{p}), \\
\bar{u}^s(\vec{p}) &= \bar{u}^s(\vec{p})\frac{\not{p}}{m}, \\
v^s(\vec{p}) &= -\frac{\not{p}}{m}v^s(\vec{p}), \\
\bar{v}^s(\vec{p}) &= -\bar{v}^s(\vec{p})\frac{\not{p}}{m},
\end{aligned} \tag{1.19}$$

These expressions are formidable for reducing terms on the form $\bar{u}^s(\vec{p})\Gamma v^{s'}(\vec{k})$. The procedure is as follows

$$\begin{aligned}
\bar{u}^s(\vec{p})\Gamma v^{s'}(\vec{k}) &= \frac{1}{2}[\bar{u}^s(\vec{p})\Gamma v^{s'}(\vec{k}) + \bar{u}^s(\vec{p})\Gamma v^{s'}(\vec{k})] \\
&= \frac{1}{2}[\bar{u}^s(\vec{p})\frac{\not{p}}{m}\Gamma v^{s'}(\vec{k}) - \bar{u}^s(\vec{p})\Gamma \frac{\not{k}}{m}v^{s'}(\vec{k})] \\
&= \dots
\end{aligned} \tag{1.20}$$

To proceed use the specific form of Γ alongside commutator relations and so forth.

Fourier transform: The four dimensional Fourier transform is defined by

$$\begin{aligned}
f(x) &= \int \frac{d^4k}{(2\pi)^4} e^{-ik \cdot x} \tilde{f}(k), \\
\tilde{f}(k) &= \int d^4x e^{ik \cdot x} f(x),
\end{aligned} \tag{1.21}$$

where x, k are four vectors of position and momentum, respectively and $\int d^4x = \int dx^0 dx^1 dx^2 dx^3$. Taking the qft metric, the three dimensional Fourier transforms are defined by

$$\begin{aligned} f(\vec{x}) &= \int \frac{d^3k}{(2\pi)^3} e^{+i\vec{k}\cdot\vec{x}} \tilde{f}(\vec{k}), \\ \tilde{f}(\vec{k}) &= \int d^3x e^{-i\vec{k}\cdot\vec{x}} f(\vec{x}). \end{aligned} \quad (1.22)$$

For arbitrary n , the n -dimensional Dirac delta function satisfies

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-i\omega(t-t')} d\omega &= 2\pi\delta(t-t'), \\ \int d^4x e^{i(k+p)\cdot x} &= \int dx^0 e^{-i[-(k^0+p^0)]x^0} \int d^3x e^{-i(\vec{k}+\vec{p})\cdot\vec{x}}, \\ &= (2\pi)\delta[-(k^0+p_0)](2\pi)^3\delta^{(3)}(\vec{k}+\vec{p}), \\ &= (2\pi)^4\delta^{(4)}(k+p), \\ \int d^n x e^{i(k+p)\cdot x} &= (2\pi)^n\delta^{(n)}(k+p). \end{aligned} \quad (1.23)$$

Random variables: The fundamental building blocks for statistics are random variables (usually denoted by X or Y). A random variable is a way to quantify the outcome of a generic experiment. In practice this means mapping the sample space to a line of real numbers. The random variables are manipulated by operators, the most common of which are the expectation operator, \mathbb{E} and the variance operator, Var . The two operators obey the identities

$$\begin{aligned} \mathbb{E}[X + Y] &= \mathbb{E}[X] + \mathbb{E}[Y], \\ \mathbb{E}[aX] &= a\mathbb{E}[X], \\ Var[X] &\equiv \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \\ Var[a + X] &= Var[X], \\ Var[aX] &= a^2Var[X], \\ Var[aX + bY] &= a^2Var[X] + b^2Var[Y] + 2abCov[X, Y], \\ Var[aX - bY] &= a^2Var[X] + b^2Var[Y] - 2abCov[X, Y], \end{aligned} \quad (1.24)$$

where

$$\begin{aligned} Cov[X, Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y], \end{aligned} \quad (1.25)$$

where

$$\begin{aligned} Cov[X, a] &= 0, \\ Cov[X, X] &= Var[X], \\ Cov[X, Y] &= Cov[Y, X], \\ Cov[aX, bY] &= abCov[X, Y], \\ Cov[a + X, b + Y] &= Cov[X, Y], \\ Cov[aX + bY, cW + dV] &= acCov[X, W] + adCov[X, V] + bcCov[Y, W] + bdCov[Y, V]. \end{aligned} \quad (1.26)$$

If two observations are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. In this case it follows that the covariance vanishes.

The random variables can be either discrete or continuous. The random variable is discrete if the outcomes are countable, and continuous otherwise. Both discrete and continuous random variables are described by a probability function (PMF or PDF) and a cumulative function (CMF or CDF).

Discrete Random Variables In the case that the random variable, X , is discrete, the PMF will equal the point probability

$$f(x_i) = \mathbb{P}(X = x_i), \quad (1.27)$$

where

$$f(x_i) > 0 \wedge \sum_{i=1}^n f(x_i) = 1. \quad (1.28)$$

The CMF is an equivalent representation of data which is defined as

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f(x_i), \quad (1.29)$$

where $0 \leq F(x) \leq 1$. In the discrete case the expectation value and variance are defined viz.

$$\mathbb{E}[X] = \sum_{x_i \leq x} x_i f(x_i), \quad \text{Var}[X] = \sum_{x_i \leq x} (x_i - \mu)^2 f(x_i). \quad (1.30)$$

Continuous Random Variables In the case that the random variable, X , is continuous the point probability will vanish since now the integral of the PDF must be unity, i.e.

$$\int_{-\infty}^{\infty} dx f(x) = 1. \quad (1.31)$$

Since the function integrated over an infinite interval must be finite, the value in any given point must vanish. Hence, in this case the PDF does not provide a probability but a probability density instead. The probability is obtained by integrating the PDF

$$\mathbb{P}(A \leq X \leq B) = \int_A^B dx f(x) = F(B) - F(A), \quad (1.32)$$

where F is the CDF defined as

$$F(x) \equiv \int_{-\infty}^x dx' f(x'). \quad (1.33)$$

In the continuous case the expectation value and variance are defined viz.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} dx x f(x), \quad \text{Var}[X] = \int_{-\infty}^{\infty} dx \left(x - \int_{-\infty}^{\infty} dx' x' f(x') \right)^2 f(x). \quad (1.34)$$

Part I

Classical Mechanics

2 | Newtonian Mechanics

Newtonian mechanics is the framework, derived by Isaac Newton in 1687, that describes physics of point particles in the non-relativistic, macroscopic limit (i.e. the limit in which the effect of relativity and quantum mechanics can be neglected). The centerpiece of Newtonian mechanics is Newton's five laws and the laws, the principle of conserved quantities and the choice of reference frame.

2.1 Newton's laws

Armed with the enormous empirical material which is condensed in the laws of Kepler and Galileo, Isaac Newton managed to discover his famous laws of motion alongside the law of gravitational interaction. Newton's five laws of motion can be stated as follows (Knudsen and Hjorth, 2000)

1. *The law of inertia:* A body which is not acted upon by any force will either be at rest or in a state of uniform, linear motion. This law is manifested in the definition of linear momentum. The linear momentum of an object is defined as the mass of the object, m , times the velocity of the object, \vec{v} , i.e.

$$\vec{p} = m\vec{v} = m\dot{\vec{x}}, \quad (2.1)$$

where the dot indicates the time derivative of the position vector. From the definition of linear momentum, the law of inertia can be stated as follows

$$\text{No external force} \Rightarrow \vec{p} = \text{constant vector.} \quad (2.2)$$

In this way the law of inertia coincides with the law of conservation of linear momentum for an isolated particle.

2. *The law of acceleration:* The mass, m , of a body, times its acceleration, \vec{a} , equals the force, \vec{F} , on the body, i.e.

$$\vec{F} = m\vec{a} = m\ddot{\vec{x}}, \quad (2.3)$$

where the dots denote the second time-derivative of the position vector. From the definition of linear momentum, the law of acceleration can be written as follows

$$\vec{F} = \frac{d\vec{p}}{dt}. \quad (2.4)$$

3. *The law of action and reaction:* If body A acts on body B with a force, then body B will act on body A with an equal but oppositely directed force.
4. *The postulate of absolute time:* Absolute, true, and mathematical time, of itself, and from its own nature, flows equally without relation to anything external, and by another name is called duration.
5. *The postulate of absolute space:* Absolute space, in its own nature, without relation to anything external, remains always similar and immovable.

The five laws are, in the form stated above, only valid in what is called an inertial reference frame. An inertial reference frame is defined as a reference frame in which Newtons first law (the law of inertia) holds true. That is to say an inertial reference frame is a reference frame in which if an object is not subject to a resulting force, it will remain at rest or move with a uniform velocity, i.e. it will not accelerate. A non-inertial reference frame is a reference frame in which Newtons first law does not hold. Examples of non-inertial reference frames are eg. rotating reference frames or accelerating reference frames. The reader may be familiar with riding a marry-go-round or accelerating in a car. In both cases the reader might recall being pushed outwards or back into the seat. In both cases, the body is required to either hold onto the marry-go-round or push back into the seat, in order to remain at rest in the car or the marry-go-round. Hence, both of these reference frame are not inertial.

Example 2.1.

An important example of a force is the gravitational force in the Newtonian limit. Newtonian gravity states that the gravitational force between two particles act to attract the two particles. The force is given by

$$\vec{F}_g = -\frac{G_N m_1 m_2}{r^2} \hat{r} = m_1 \vec{g}, \quad (2.5)$$

where $G_N = 6.67 \cdot 10^{-11} \frac{Nm^2}{kg^2}$ is Newtons gravitational constant, m_i are the masses of the two particles and r is the radial vector connecting the two particles. The minus sign indicated the force is attractive.

Example 2.2.

Determine and solve the EOM for a simple pendulum.

The simple pendulum is a mass hanging from a massless string, free to swing (see figure 2.1).

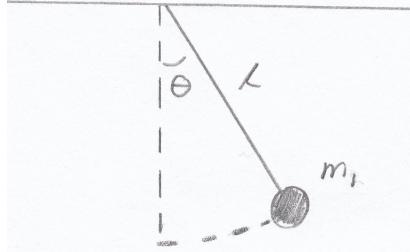


Figure 2.1: The simple pendulum.

From Newtons second law

$$\vec{F}_{res} = m_1 \vec{a} = \vec{F}_g + \vec{F}_s, \quad (2.6)$$

where \vec{F}_s is the force from the string onto the mass, m_1 . m_1 does not move in the radial direction, so \vec{F}_s must cancel the part of \vec{F}_g in the radial direction, i.e.

$$|\vec{F}_s| = -|\vec{F}_g| \cos(\theta). \quad (2.7)$$

Therefore

$$m_1 |\vec{a}| = -|\vec{F}_g| \sin(\theta). \quad (2.8)$$

Now, \vec{a} is the linear acceleration, in terms of the angular acceleration; $|\vec{a}| = \ddot{\theta}l$. Taking the small angle limit ($\sin(\theta) \simeq \theta$), and using that $|\vec{F}_g| = m_1 g$

$$\ddot{\theta} + \frac{g}{l} \theta \simeq 0 \Rightarrow \theta \simeq \theta_0 \cos\left(\sqrt{\frac{g}{l}} t\right). \quad (2.9)$$

2.2 Non-inertial reference frames

When considering non-inertial reference frames it is desired to relate the force in a non-inertial reference frame to the forces in an inertial reference frame. Hence, the procedure of interest is how to transform Newtons second law from an inertial reference frame to a non-inertial reference frame. For this reason, a particle is considered from two different coordinate systems. One coordinate system, I, is an inertial reference frame that describes the location of the particle in terms of the position vector, \vec{x} . The non-inertial reference frame, I', performs a translating and rotating motion with respect to I. The distance from Origo in I to Origo in I' is defined by the vector \vec{x}_0 , while the distance from Origo in I' to the particle is denoted by the vector \vec{x}' (see figure 2.2). The two coordinate system could for example be the heliocentric reference frame as the

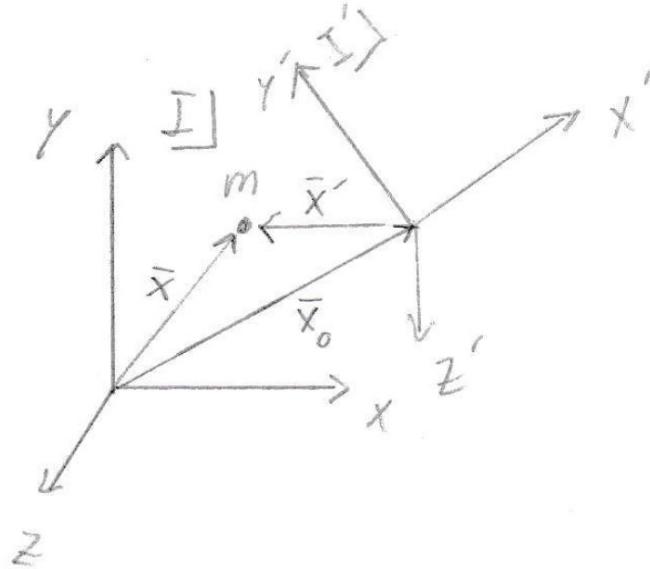


Figure 2.2: The definition of the inertial and non-inertial reference frames, I,I', respectively.

inertial reference frame and the geocentric reference frame as the non-inertial reference frame¹. From the definitions made in figure 2.2 the position vector, \vec{x} , can be expressed in terms of \vec{x}_0 and \vec{x}' viz

$$\vec{x} = \vec{x}_0 + \vec{x}'. \quad (2.10)$$

Equation (2.10) can be used to relate the acceleration on m_1 as seen from I, I' by differentiating it twice with respect to time

$$\ddot{\vec{x}} = \ddot{\vec{x}}_0 + \ddot{\vec{x}'} + \vec{\omega} \times (\vec{\omega} \times \vec{x}') + 2(\vec{\omega} \times \dot{\vec{x}'}) + \dot{\vec{\omega}} \times \vec{x}'. \quad (2.11)$$

By multiplying with the mass of the particle, and isolating $\vec{F}_{res,I'} = m\ddot{\vec{x}'}$ from equation (2.11)

$$\begin{aligned} \vec{F}_{res,I'} &= m\ddot{\vec{x}'} \\ &= m[\ddot{\vec{x}} - \ddot{\vec{x}}_0 - \vec{\omega} \times (\vec{\omega} \times \vec{x}') - 2(\vec{\omega} \times \dot{\vec{x}'}) - \dot{\vec{\omega}} \times \vec{x}'], \end{aligned} \quad (2.12)$$

where $\vec{\omega}$ is the angular velocity of a given rotation around the axis going through the origin in I'. From equation (2.12) it is clear that the resulting force in the reference frame of I' is equal to the resulting force in I subtracted some terms. These subtracted terms are what is called fictitious forces. Fictitious" because the forces are not real in the sense that they appear in an inertial reference frame. The fictitious forces represent

¹Depending on which approximations are made. Often the geocentric reference frame is taken to be an inertial reference frame. This is a valid approximation as long as the physical experiment under consideration does not span a significant portion of the Earth, i.e. as long as the experiment is confined to a smaller region of the Earth.

the movement of the non-inertial reference frame, and as such they do not originate from any other body (as inertial forces do). The fictitious forces are defined as follows

$$\begin{aligned}
\text{Origins; linear acceleration of } I', \text{ wrt. } I: & \quad \ddot{\vec{x}}_0, \\
\text{Centrifugal acceleration:} & \quad \vec{\omega} \times (\vec{\omega} \times \vec{x}'), \\
\text{Coriolis acceleration:} & \quad 2(\vec{\omega} \times \dot{\vec{x}}'), \\
\text{Azimuthal acceleration:} & \quad \dot{\vec{\omega}} \times \vec{x}'.
\end{aligned} \tag{2.13}$$

The linear acceleration term is the one experienced when sitting inside an accelerating car. The centrifugal term is one which is experienced on a merry-go-round. The Coriolis term is more obscure. It is a force experienced by particles which move in a rotating reference frame. For example, one would experience the Coriolis force if one walked around on a rotating merry-go-round. If one walked towards the center of a merry-go-round one would experience the Coriolis force pushing from the side. The azimuthal term, also called the Euler term, is also an obscure term. It is experienced if the angular velocity of the rotating reference frame is changed. For example it would be experienced if the merry-go-round's rotational velocity accelerated or decelerated. If walking towards the center of a merry-go-round that is accelerating in rotational velocity, the azimuthal force would act to the opposite side as that of the Coriolis force. From the above definitions, the resulting force in a non-inertial reference frame can be written viz

$$\vec{F}_{res,I'} = \vec{F}_{res,I} + \vec{F}_{trans} + \vec{F}_{centr} + \vec{F}_{cor} + \vec{F}_{azi}, \tag{2.14}$$

where the fictitious forces are defined to absorb the minus signs.

2.3 Center of mass

Consider a system of point particles observed from some inertial reference frame, I. The particles might be mutually interacting (eg. via the gravitational interaction or the electromagnetic interaction) such that they form a rigid body. A rigid body is defined as a body consisting of many particles which do not change position during movement of the rigid body. The center of mass, with respect to origin in the inertial coordinate system, of both a system of particles and a rigid body is found by summing the positions of the individual particles weighted with their individual masses and divided by the total mass, i.e.

$$\vec{x}_{CM} = \frac{1}{m_{tot}} \sum_i m_i \vec{x}_i, \tag{2.15}$$

where $m_{tot} = \sum_i m_i$. Differentiating equation (2.15) with respect to time, and taking m_{tot} to the left hand side, reveals

$$m_{tot} \vec{v}_{CM} = \sum_i m_i \vec{v}_i = \sum_i \vec{p}_i = \vec{p}. \tag{2.16}$$

Equation (2.16) informs that the total, linear momentum of a system of particles, i.e. \vec{p} , is the sum of the linear momenta of all the constituent particles. Differentiating equation (2.16) once more with respect to time

$$\vec{F}_{res} = \frac{d\vec{p}}{dt} = \sum_i \frac{d\vec{p}_i}{dt} = \sum_i m_i \ddot{\vec{x}}_i. \tag{2.17}$$

The resulting force can be divided into two categories; the external forces and the internal forces. The external forces are forces that originate from bodies outside the system of particles whereas internal forces are forces that originate between constituents of the system of particles. According to Newton's third law (the law of action and reaction); the force from any particle onto another results in an equal force with opposite direction the other way around. Hence, denoting the force between particles i, j as \vec{F}_{ij}

$$\vec{F}_{ij} = -\vec{F}_{ji}. \tag{2.18}$$

Using equation (2.18) in equation (2.17) it is clear that the contribution to the resulting force from the internal forces vanish, and so

$$\vec{F}_{res} = \sum_i \vec{F}_{ext,i}. \quad (2.19)$$

Center of mass theorem: *Hence, the entire mass of a system of particles, rigid or not, behaves as if the entire mass were concentrated in the center of mass and all external forces act in this point (from equation (2.19)).*

Conservation of linear momentum: *If no external forces act on a system of particles the total momentum, \vec{p} , of the system is a constant vector (from equation (2.19), (2.17)).*

2.4 Energy

Consider a particle, observed from an inertial reference frame, moving under the influence of a generic force, \vec{F} , which is taken to be a function of any relevant variables. When the particle is moved, by the force, from point a to point b, the force performs an amount of work, W_{AB} , on the particle, given by

$$W_{ab} = \int_a^b \vec{F} \cdot d\vec{x} = \int \vec{F} \cdot \hat{n} d^3x, \quad (2.20)$$

where \hat{n} is the unit tangent vector along the path of integration, i.e. the path along which which the particle is moving. Using Newtons second law equation (2.20) can be written as follows

$$W_{ab} = \int_a^b m \frac{d\vec{v}}{dt} \cdot d\vec{x} = \int_a^b m\vec{v} \cdot d\vec{v} = \frac{1}{2}mv_b^2 - \frac{1}{2}mv_a^2. \quad (2.21)$$

The quantity $T = \frac{1}{2}mv^2$ is defined as the kinetic energy of the particle. From equation (2.21) it is clear that the work done by all force acting on a particle equals the increase in kinetic energy of the particle, i.e.

$$W = \int \vec{F} \cdot d\vec{x} = \Delta T \Rightarrow dT = \vec{F} \cdot d\vec{x} \Rightarrow \frac{dT}{dt} = \vec{F} \cdot \vec{v}. \quad (2.22)$$

$\vec{F} \cdot \vec{v}$ is defined as the power done by the force \vec{F} . Related to the work integral is the definition of a conservative force and a conservative force field.

Conservative force fields: *A force field is called conservative if the work integral, $W_{ab} = \int_a^b \vec{F} \cdot d\vec{x}$, is independent on the path of integration, and depends only on the initial and final points.*

Central force fields: *A central force field is a field where the force on the particle is always directed towards (or away from) a point, fixed in an inertial reference frame. This point is called the force center. All central force fields are conservative force fields and as such central force fields are a subgroup of conservative force fields.*

A particle in a conservative force field has associated to it a potential energy, $V = V(\vec{x})$. The potential energy is defined with respect to some reference point and the notion of the "total potential energy" is is senseless. Taking $U(P) = 0$, the potential energy of a particle at point a (in a conservative force field), is given by

$$V(\vec{a}) = - \int_P^a \vec{F} \cdot d\vec{x} = -W_{ab} \Rightarrow \vec{F} = -\vec{\nabla}V(\vec{x}), \quad (2.23)$$

where $\vec{\nabla} = [\partial_x \ \partial_y \ \partial_z]^T$ is the gradient operator². From equation (2.23) it is clear that the potential energy is equal to minus the work done in moving a particle between two points. So, one can think of the

²A mathematical operator. Not to be confused with a quantum mechanical operator

potential energy as a means to store work. Consider now two arbitrary points in a conservative force field (a, b) . For these points

$$\int_P^a \vec{F} \cdot d\vec{x} = -V(a), \quad \int_P^b \vec{F} \cdot d\vec{x} = -V(b). \quad (2.24)$$

From equation (2.24)

$$\int_a^b \vec{F} \cdot d\vec{x} = \int_a^P \vec{F} \cdot d\vec{x} + \int_P^b \vec{F} \cdot d\vec{x} = V(a) - V(b). \quad (2.25)$$

Comparing equation (2.25) to equation (2.21)

$$\frac{1}{2}mv_b^2 - \frac{1}{2}mv_a^2 = V(a) - V(b) \Rightarrow \frac{1}{2}mv_b^2 + V(b) = \frac{1}{2}mv_a^2 + V(a). \quad (2.26)$$

From equation (2.26) it is clear that the sum of the kinetic and potential energy does not change as the particle is moved, by a force, through a conservative force field. The sum of the kinetic and potential energy is defined as the mechanical energy, E_{mek} , so

$$\text{In a conservative force field: } E_{mek} = T + V = \text{const.} \quad (2.27)$$

Example 2.3.

A professional skier is skiing downhill along the slope shown in figure 2.3.

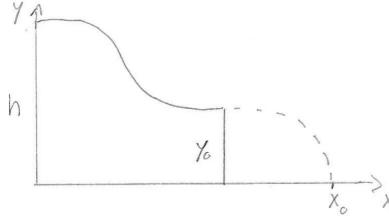


Figure 2.3: The scenario of the ski-jumper.

The skier begins at $h = 10m$ with $v = 0 \frac{m}{s}$. Assuming that there is no friction between the skies and the snow the skier arrives at the horizontal trampoline with height y_0 . Jumping out the skier lands at a distance x_0 from the horizontal trampoline.

1. Determine the value of the trampoline height, y_0 , which maximizes x_0 .

Since the skier is at rest when initiating the descend, from conservation of mechanical energy (no friction forces)

$$E_{mek} = mgh = \frac{1}{2}mv_0^2 + mg y_0 \Rightarrow v_0 = \sqrt{2g(h - y_0)}. \quad (2.28)$$

Next use that the skier exits the jump at a horizontal direction. Subsequent to exiting the jump, the only force the skier is subject to is gravity. Since the gravitational acceleration is constant close to the Earth, the skier will take a time, t_0 , before landing, given by

$$y_0 = \frac{1}{2}gt_0^2 \Rightarrow t_0 = \sqrt{\frac{2y_0}{g}}. \quad (2.29)$$

During this time, the horizontal distance traversed is given by

$$x_0 = v_0 t_0 \Rightarrow x_0 = \sqrt{2g(h - y_0)} \sqrt{\frac{2y_0}{g}} = 2\sqrt{hy_0 - y_0^2}. \quad (2.30)$$

To maximize $x_0(y_0)$, isolate y_0 from $\frac{dx_0(y_0)}{dy_0} = 0$ as follows

$$\frac{dx_0(y_0)}{dy_0} = \frac{h - 2y_0}{\sqrt{hy_0 - y_0^2}} = 0 \Rightarrow y_0 = \frac{h}{2} = 5m. \quad (2.31)$$

$\frac{d^2x_0(y_0)}{dy_0^2} < 0$ is indeed the case, so the point is a local maximum - Just as intended. This means $y_0 = \frac{h}{2} = 5m$ maximizes x_0 .

2. Determine the velocity of the skier when leaving the trampoline (v_0), and the distance in free air (x_0).

$$v_0 = \sqrt{2g(h - \frac{h}{2})} = 9.9 \frac{m}{s}, \quad (2.32)$$

$$x_0 = \sqrt{2h^2 - h^2} = 10m.$$

Example 2.4. —

An experienced mountaineer, which weights $m = 80kg$, is climbing a near vertical wall. The mountaineer is secured against a possible fall by a climbing robe. One end of the robe is attached to the mountaineer while the other is secured by a pole in the rock below the mountaineer. The mountaineer climbs sufficiently high that he has used up the full length of the rope, l_0 , which is still not stretched. This full length of the rope is therefore identical to the difference in height between the mountaineer and the second pole below him. Assume negligible rope weight, and that the rope behaves as an ideal string with elastic constant $K = \frac{a}{l_0}$, with $a = 40000N$.

1. Determine the maximum tension of the rope if the mountaineer falls.

The physical scenario is modeled as shown in figure 2.4.

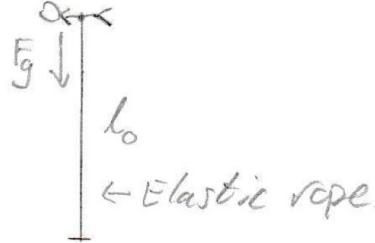


Figure 2.4: The scenario of the mountaineer.

Use that the maximum force from the rope is given by

$$F_{max} = kx_{max}, \quad (2.33)$$

where x_{max} is found by using the conservation of energy. The mountaineer initiates the fall with only potential energy. Subsequent to falling the mountaineer will stretch the rope and at some point, at which time the rope will be stretched maximally, come to rest. When the rope is stretched maximally the tension of the rope will be at its highest, so

$$E_{pot} = mg(2l_0 + x_{max}) = \frac{1}{2}kx_{max}^2$$

$$\Rightarrow x_{max} = \frac{mg + \sqrt{(mg)^2 + 4Kmgl_0}}{k} \quad (2.34)$$

$$\Rightarrow F_{max} = mg + \sqrt{(mg)^2 + 4kmgl_0} \simeq 12015N.$$

2.5 Angular momentum in classical mechanics

Consider a particle of mass m moving with a velocity \vec{v} relative to some inertial reference frame I with origins in the point O. Relative to the point O, the angular momentum is defined as

$$\vec{L} = \vec{x} \times \vec{p} = m\vec{x} \times \vec{v} = mrvsin(\theta), \quad (2.35)$$

where $r = |\vec{x}|$ and θ is the angle between \vec{v} and \vec{x} . The cross product multiplies the radius vector with the part of the velocity that is perpendicular to the radius vector. Differentiating equation (2.35) with respect to time reveals

$$\frac{d\vec{L}}{dt} = \frac{d}{dt}(\vec{x} \times \vec{p}) = \frac{d\vec{x}}{dt} \times \vec{p} + \vec{x} \times \frac{d\vec{p}}{dt}. \quad (2.36)$$

Since $\frac{d\vec{x}}{dt} = \vec{v} \parallel \vec{p}$ the first term in equation (2.36) vanishes and so

$$\frac{d\vec{L}}{dt} \vec{x} \times \frac{d\vec{p}}{dt} = \vec{x} \times \vec{F}. \quad (2.37)$$

The product $\vec{x} \times \vec{F}$ is defined as torque, \vec{N} , so

$$\vec{N} \equiv \vec{x} \times \vec{F} = \frac{d\vec{L}}{dt}. \quad (2.38)$$

The angular momentum theorem: *The rate of change of the angular momentum of a particle around some point, O, equals the torque on the particle, with respect to O.*

Conservation of angular momentum: *If no torque acts on a particle, the angular momentum, around some point O, of that particle, with respect to the point O, is constant in time.*

Example 2.5.

A classic example of a system in which angular momentum is conserved is a central force field, eg. the gravitational field of a point particle.

Example 2.6.

The angular momentum of a closed system, i.e. a system upon which no external forces act, is conserved.

For a system of particles torque is given by

$$\vec{N}_{res} = \sum_i \vec{N}_i = \sum_i \frac{d\vec{L}_i}{dt}, \quad (2.39)$$

where i runs over both external and internal forces. From Newtons third law $\vec{F}_{ij} = -\vec{F}_{ji}$, so for the internal torque

$$\vec{N}_{ij} = \vec{x}_i \times \vec{F}_{ij} + \vec{x}_j \times \vec{F}_{ji} = (\vec{x}_i - \vec{x}_j) \times \vec{F}_{ij} = \vec{0}. \quad (2.40)$$

Hence, analogous to the case of Newtons second law for linear motion in which the resulting force was indifferent to the internal forces; the resulting torque is indifferent to the internal torques, so

$$\vec{N}_{res} = \sum_i \vec{N}_{ext,i}. \quad (2.41)$$

Example 2.7.

Halley's comet is the most famous of all comets. It has a period of 76 years and has been observed regularly since the battle of Hastings in 1066. It has, in the point where it is closest to the Sun, a distance r_0 to the Sun, and a velocity v_0 relative to the heliocentric reference frame.

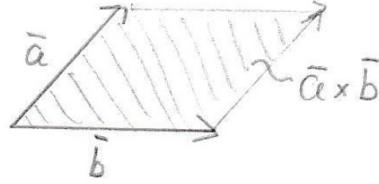


Figure 2.5: The parallelogram of the cross product.

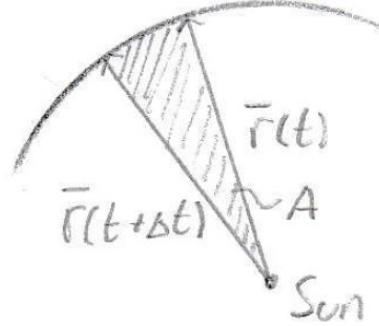


Figure 2.6: The area under consideration.

1. Find a formal expression of the areal velocity of the comet, $\frac{dA}{dt}$, expressed in terms of r_0 and v_0 .

Use that the cross product between two vectors denotes the area of the parallelogram they enclose (see figure 2.5). The vectors relevant for the parallelogram in the physical scenario is $\vec{v}(t)$ and $\vec{r}(t + \Delta t)$ (see figure 2.6). From figure 2.6

$$\Delta A = \frac{\vec{r}(t) \times \vec{r}(t + \Delta t)}{2} \quad (2.42)$$

Dividing equation (2.42) by Δt and taking the limit of $\Delta t \Rightarrow 0$

$$\begin{aligned} \frac{dA}{dt} &= \lim_{\Delta t \rightarrow 0} \left(\frac{\Delta A}{\Delta t} \right) = \lim_{\Delta t \rightarrow 0} \left(\frac{\vec{r}(t) \times \vec{r}(t + \Delta t)}{2\Delta t} \right) \\ &= \lim_{\Delta t \rightarrow 0} \left(\frac{\vec{r}(t) \times (\vec{r}'(t) + \dot{\vec{r}}(t)\Delta t)}{2\Delta t} \right) \\ &= \lim_{\Delta t \rightarrow 0} \left(\frac{\vec{r}(t) \times \dot{\vec{r}}(t) \Delta t}{2} \frac{\Delta t}{\Delta t} \right) = \frac{\vec{r}(t) \times \vec{v}(t)}{2}, \end{aligned} \quad (2.43)$$

where it has been used that $\vec{r}(t) \times \vec{r}(t) = \vec{0}$. Since the area velocity is constant the values of v, r at the closest approach to the Sun can be used. At closest approach $\vec{v}_0 \perp \vec{r}_0$, so

$$\frac{dA}{dt} = \frac{r_0 v_0}{2}. \quad (2.44)$$

Alternatively, angular momentum can be used to obtain the same result. Consider figure 2.7 From figure 2.7

$$dA = \frac{1}{2} r^2 d\theta \Rightarrow \frac{dA}{dt} = \frac{1}{2} r^2 \frac{d\theta}{dt}. \quad (2.45)$$

To determine $\frac{d\theta}{dt}$ use conservation of angular momentum. The equation of motion for the comet is given by

$$m_C \ddot{\vec{r}} = -\frac{G_N m_C m_\odot}{|\vec{r}|^2} \frac{\vec{r}}{|\vec{r}|}, \quad (2.46)$$

where m_C is the mass of the comet and m_\odot is the mass of the Sun. In polar coordinates

$$\ddot{\vec{r}} = \left(|\ddot{\vec{r}}| - |\vec{r}| \left(\frac{d\theta}{dt} \right)^2 \right) \hat{e}_r + \frac{1}{|\vec{r}|} \frac{d}{dt} \left(|\vec{r}|^2 \frac{d\theta}{dt} \right) \hat{e}_\theta. \quad (2.47)$$

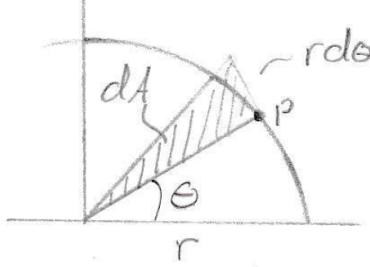


Figure 2.7: The orbit in polar coordinates.

Using equation (2.47) in equation (2.46) results in two equations; a radial and an angular

$$\begin{aligned} \text{Radial: } m_C \left(|\ddot{\vec{r}}| - |\vec{r}| \left(\frac{d\theta}{dt} \right)^2 \right) \hat{e}_r &= -\frac{G_N m_C m_\odot}{|\vec{r}|^2} \hat{e}_r, \\ \text{Angular: } \frac{1}{|\vec{r}|} \frac{d}{dt} \left(|\vec{r}|^2 \frac{d\theta}{dt} \right) \hat{e}_\theta &= 0. \end{aligned} \quad (2.48)$$

From the angular equation

$$\begin{aligned} \frac{d}{dt} \left(|\vec{r}|^2 m_C \dot{\theta} \right) &= 0 \Rightarrow m_C |\vec{r}|^2 \dot{\theta} = \text{const} \equiv L \\ \Rightarrow \dot{\theta} &= \frac{L}{m_C r^2}, \end{aligned} \quad (2.49)$$

where $r = |\vec{r}|$. Using equation (2.49) in equation (2.45)

$$\frac{dA}{dt} = \frac{L}{2m_C} = \frac{m_C \vec{r} \times \vec{v}}{2m_C} = \frac{r_0 v_0}{2}. \quad (2.50)$$

2. Find formal expressions of the angular momentum, L , and the total energy, E_{mek} , of the comet. From these two expressions obtain expressions for the eccentricity, e , the semi-major axis of the comets orbit, a , and the period, T , of the orbital motion. All expressions should be formulated in terms of r_0 , v_0 , m_\odot and G_N .

Begin by considering the radial equation (equation (2.48)). Substituting equation (2.49) into the radial equation of equation (2.48)

$$m_C \ddot{r} - r \frac{L^2}{m_C r^4} = -\frac{G_N m_C m_\odot}{r^2} \Rightarrow \ddot{r} - \frac{L^2}{m_C^2 r^3} = \frac{C}{m_C r^2}, \quad (2.51)$$

where $C \equiv -G_N m_C m_\odot$. Equation (2.51) is a differential equation for $r(t)$, however, in relation to the orbit the quantity of interest is not $r(t)$, but rather $r(\theta)$. To introduce $r(\theta)$ into equation (2.51), use that

$$\begin{aligned} \ddot{r} &= \frac{d}{dt} \left(\frac{dr}{dt} \right) = \frac{d}{dt} \left(\frac{dr}{d\theta} \frac{d\theta}{dt} \right) = \frac{d}{dt} \left(\frac{dr}{d\theta} \frac{L}{m_C r^2} \right) \\ &= \frac{d}{d\theta} \left(\frac{dr}{d\theta} \frac{L}{m_C r^2} \right) \frac{d\theta}{dt} = \frac{L^2}{m_C^2 r^4} \frac{d^2 r}{d\theta^2} + \frac{L^2}{m_C^2 r^2} \left(-2r^{-3} \frac{dr}{d\theta} \right) \frac{dr}{d\theta} \\ &= \frac{L^2}{m_C^2 r^4} \left(\frac{d^2 r}{d\theta^2} - \frac{2}{r} \left(\frac{dr}{d\theta} \right)^2 \right). \end{aligned} \quad (2.52)$$

Next, introduce the variable $u = r^{-1}$, for which

$$\begin{aligned}
\frac{d^2u}{d\theta^2} &= \frac{d}{d\theta}\left(\frac{du}{d\theta}\right) = \frac{d}{d\theta}\left(\frac{du}{dr}\frac{dr}{d\theta}\right) = \frac{d}{d\theta}\left(-r^{-2}\frac{dr}{d\theta}\right) \\
&= \frac{d}{dr}\left(-r^{-2}\right)\left(\frac{dr}{d\theta}\right)^2 + (-r^{-2})\frac{d^2r}{d\theta^2} \\
&= \frac{2}{r^3}\left(\frac{dr}{d\theta}\right)^2 - \frac{1}{r^2}\frac{d^2r}{d\theta^2} \\
&= -\frac{1}{r^2}\left(\frac{d^2r}{d\theta^2} - \frac{2}{r}\left(\frac{dr}{d\theta}\right)^2\right).
\end{aligned} \tag{2.53}$$

Comparing equation (2.52) and (2.53)

$$\ddot{r} = -\frac{L^2}{m_C^2 r^2} \frac{d^2u}{d\theta^2}. \tag{2.54}$$

Using equation (2.54) in equation (2.51), alongside the definition of $u = r^{-1}$

$$-\frac{L^2}{m_C^2} u^2 \frac{d^2u}{d\theta^2} - \frac{L^2}{m_C^2} u^3 = \frac{C}{m_C} u^2 \Rightarrow \frac{d^2u}{d\theta^2} + u = -\frac{C m_C}{L^2}. \tag{2.55}$$

Equation (2.55) is an inhomogeneous, second order, linear, ordinary differential equation. The solution is a sum of *one* solution to the inhomogeneous equation (u_p) and *the* solution to the homogeneous equation (u_h). *One* solution to the inhomogeneous equation is

$$u_p = -\frac{C m_C}{L^2}. \tag{2.56}$$

The solution to the homogeneous equation is found via the conventional means (see Adams and Essex 2009)

$$u_h = A \cos(\theta + \phi_0), \tag{2.57}$$

where ϕ_0 is some phase. Hereby

$$u = B \cos(\theta + \phi_0) - \frac{C m_C}{L^2}. \tag{2.58}$$

By an appropriate placement of the coordinate system, $\phi_0 = 0$. To determine the integration-constant, B , initial conditions must be applied, or alternatively conservation laws. From conservation of energy

$$E_{mek} = \frac{1}{2} m_C v^2 + \frac{C}{r} = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) + \frac{C}{r}, \tag{2.59}$$

where polar coordinates have been used to express v . Using $\theta = \frac{L}{m_C r^2}$, from equation (2.49), the mechanical energy can be written as follows

$$E_{mek} = \frac{L^2}{2m_C r^4} \left(\left(\frac{dr}{d\theta} \right)^2 + r^2 \right) + \frac{C}{r}. \tag{2.60}$$

Using equation (2.58) in equation (2.60)

$$\begin{aligned}
\frac{dr}{d\theta} &= \frac{1}{B} \frac{\sin(\theta)}{\cos(\theta)^2} = B r^2 \sin(\theta) \\
\Rightarrow E_{mek} &= \frac{L^2}{2m_C r^4} \left(B^2 r^4 \sin(\theta)^2 + r^2 \right) + \frac{C}{r}.
\end{aligned} \tag{2.61}$$

Using again equation (2.58) in equation (2.61)

$$E_{mek} = \frac{L^2 B^2}{2m_C} - \frac{C^2 m_C}{2L^2} \Rightarrow B = \sqrt{\left(E_{mek} + \frac{C^2 m_C}{2L^2} \right) \frac{2m_C}{L^2}}. \tag{2.62}$$

Using equation (2.62) in equation (2.58)

$$\frac{1}{r} = \sqrt{\frac{2m_C E_{mek}}{L^2} + \frac{G_N^2 m_\odot^2 m_C^4}{L^4} \cos(\theta)} + \frac{G_N m_C^2 m_\odot}{L^2}. \quad (2.63)$$

Equation (2.63) can be written on the form of an ellipse equation, i.e.

$$\frac{1}{r} = \frac{1}{p}(1 + e \cdot \cos(\theta)), \quad (2.64)$$

where

$$p = \frac{L^2}{G_N m_C^2 m_\odot},$$

$$e = \sqrt{1 + \frac{2E_{mek} L^2}{G_N^2 m_C^3 m_\odot^2}}, \quad (2.65)$$

where e is the eccentricity of the orbit and p is half the length of the chord perpendicular to the axis of the conic section and passing through the focus point. To get some results, use that from equation (2.50)

$$\frac{dA}{dt} = \frac{L}{2m_c} \Rightarrow A = \frac{LT}{2m_C} = \pi ab = \pi a^2 \sqrt{1 - e^2}, \quad (2.66)$$

where a is the length of the semi-major axis, b is the length of the semi-minor axis, $A = \pi ab$ is the area of an ellipse and $b = a\sqrt{1 - e^2}$ is valid for an ellipse. Next use that

$$p = a(1 - e^2) = \frac{L^2}{G_N m_C^2 m_\odot}$$

$$\Rightarrow a = \frac{L^2}{G_N m_C^2 m_\odot} \frac{1}{1 - e^2} = \frac{G_N m_C m_\odot}{2|E_{mek}|}. \quad (2.67)$$

Using equation (2.67) in equation (2.66)

$$A = \frac{LT}{2m_C} = \pi \left(\frac{G_N m_C m_\odot}{2|E_{mek}|} \right)^2 \sqrt{1 - e^2}$$

$$= \pi \left(\frac{G_N m_C m_\odot}{2|E_{mek}|} \right)^2 \sqrt{\frac{2|E_{mek}| L^2}{G_N^2 m_C^3 m_\odot^2}}$$

$$= \pi G_N M_\odot L \sqrt{m_c} (2|E_{mek}|)^{-\frac{3}{2}} = \frac{\pi G_N m_\odot v_0 r_0}{(v_0^2 - \frac{2G_N m_\odot}{r_0})^{\frac{3}{2}}}, \quad (2.68)$$

where L, E_{mek} has been inserted for the last equality. Using E_{mek} in equation (2.67)

$$a = \frac{G_N M_\odot m_C}{2|E_{mek}|} = \frac{1}{\frac{v_0^2}{G_N m_\odot} - \frac{2}{r_0}} \quad (2.69)$$

Lastly, the period of the orbit:

$$T = \frac{2m_C A}{L} = \frac{2\pi G_N m_\odot}{(v_0^2 - \frac{2G_N m_\odot}{r_0})^{\frac{3}{2}}} \quad (2.70)$$

3. Find numerical values for e and T by using $r_0 = 0.5870 \text{ AU}$, $1 \text{ AU} = 1.496 \cdot 10^9 \text{ m}$, $v_0 = 54.53 \frac{\text{km}}{\text{s}}$, $G_N = 6.76 \cdot 10^{-11} \frac{\text{Nm}^2}{\text{kg}^2}$ and $m_\odot = 1.989 \cdot 10^{30} \text{ kg}$.

By inserting the numbers:

$$\begin{aligned}
 a &= \frac{G_N m_{\odot}}{2|\frac{1}{2}v_0^2 - \frac{G_N m_{\odot}}{r_0}|} \simeq 2.69 \cdot 10^9 \text{ km} \simeq 18 \text{ AU} \\
 e &= \sqrt{1 + \frac{v_0^4 r_0^2 - 2G_N m_{\odot} v_0^2 r_0}{G_N^2 m_{\odot}^2}} \simeq 0.967 \\
 T &= \frac{2\pi a^2 \sqrt{1-e^2}}{r_0 v_0} \simeq 76.28 \text{ years}
 \end{aligned} \tag{2.71}$$

Example 2.8.

The Enterprise's last tour into darkness has been completed. After the last deep-space mission the spaceship has re-entered the Solar system. It is now on a circular orbit of radius R_0 around the Earth. The Enterprise has velocity \vec{v}_0 with respect to the reference frame with origin in the center of the Earth and the axis towards what seems to be the fixed stars in the night sky. Consider this reference frame to be inertial. Upon the sudden news that a new super villain is attacking the headquarters of the intergalactic confederation fleet, the Enterprise moves from a circular orbit of radius R_0 to another, elliptical orbit, with perigee, P , at a distance, R , from the center of the Earth. During this maneuverer the Enterprise expels the mass m_1 of fuel while acquiring the velocity $\vec{v}_1 = \vec{v}_0 + \Delta\vec{v}$ with $\Delta\vec{v}$ in the direction of the Earth's center. When the Enterprise passes through the perigee, P , again, it ejects fuel of mass m_2 to acquire the escape velocity tangent to the elliptic orbit. In all the cases considered above, assume, for simplicity, that the fuel is expelled instantaneously and that the modulus of its speed relative to the Enterprise, after been expelled, is v_r . Lieutenant Commander Spock (you) acquires the following data from the command deck: $R = 0.91R_0$ and $v_r = 2\sqrt{\frac{G_N M_E}{R_0}}$. James T. Kirk, the commander of the Enterprise, asks the Lieutenant Commander Spock (you):

- What is the ratio between the mass m_1+m_2 of fuel consumed and the total mass, m_s , of the Enterprise?

The physical scenario is shown in figure 2.8. The goal is to determine: Ratio $1 = \frac{m_1+m_2}{m_s}$. To do so,

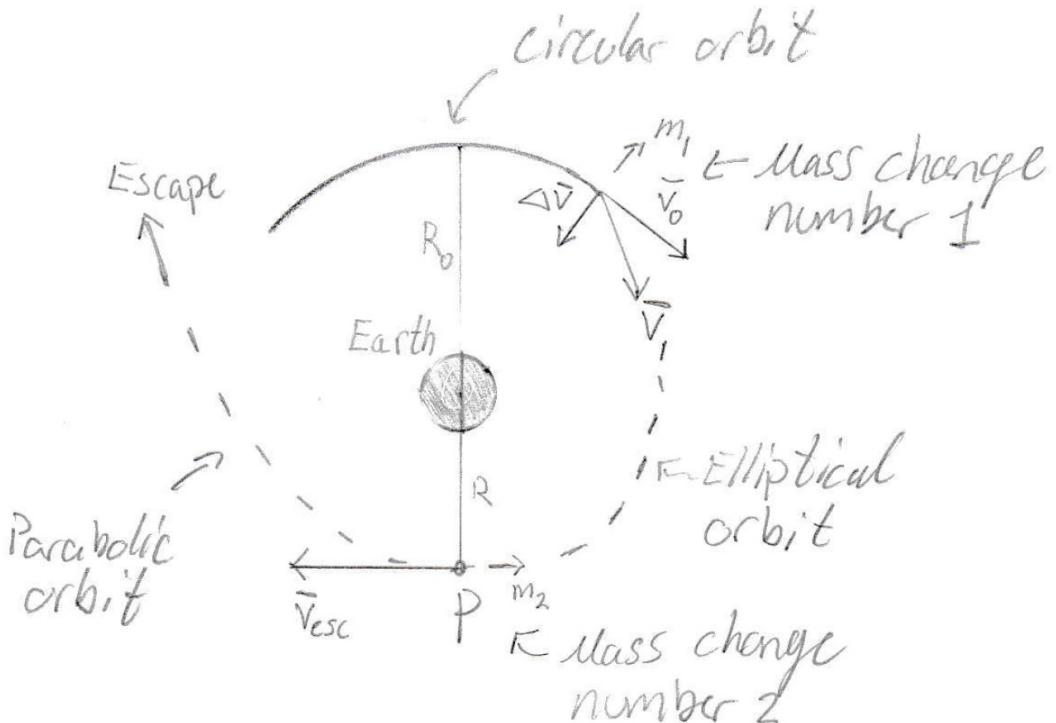


Figure 2.8: The scenario of the Enterprise.

m_1 must be specified in terms of m_2 and m_s must be specified in terms of m_1, m_2 . The plan is to use conservation of linear and angular momentum to do so. Conservation across the first ejection of mass dictates:

$$\begin{aligned} m_s \vec{v}_0 &= (m_s - m_1)(\vec{v}_0 + \Delta \vec{v}_1) + m_1 \vec{u}_1 \\ &= m_s \vec{v}_0 + m_s \Delta \vec{v}_1 + m_1(\vec{u}_1 - \vec{v}_0 - \Delta \vec{v}_1) \\ &= m_s \vec{v}_0 + m_s \Delta \vec{v}_1 + m_1 \vec{v}_r \\ \Rightarrow \Delta \vec{v}_1 &= -\frac{m_1}{m_s} \vec{v}_r \end{aligned} \quad (2.72)$$

where $\Delta \vec{v}_i$ denotes the velocity gained, \vec{u}_i denotes the velocity of the fuel with respect to the (approximated) inertial reference frame of the Earth and $\vec{v}_r = \vec{u} - \vec{v}_0 - \Delta \vec{v}_1$ is the relative velocity between the Enterprise and the fuel after it has been ejected. $\Delta \vec{v}_1 \perp \vec{v}_0$, so $v_1^2 = v_0^2 + \Delta v_1^2 + 2\vec{v}_0 \cdot \Delta \vec{v}_1 = v_0^2 + \Delta v_1^2 = v_0^2 + \left(\frac{m_1}{m_s} v_r\right)^2$. Hence, $v_1(v_0, m_1, m_s)$ has been determined. In order to determine the relation between m_1 and m_s , another (independent) equation for v_1 must be obtained. Such an equation comes from energy conservation between ejection of m_1 and m_2 :

$$\begin{aligned} E_{mek} &= \frac{1}{2}(m_s - m_1)v_1^2 - \frac{G_N(m_s - m_1)m_E}{R_0} \\ &= \frac{1}{2}(m_s - m_1)v_p^2 - \frac{G_N(m_s - m_1)m_E}{0.91R_0} \end{aligned} \quad (2.73)$$

To express the gravitational potential energy in terms of the circular, orbital velocity, v_0 , use that for a circular orbit the eccentricity vanishes, i.e.

$$\begin{aligned} e &= \sqrt{1 + \frac{2E_{mek}L^2}{G_N^2 m_s^3 m_E^2}} = 0 \\ \Rightarrow \frac{2E_{mek}L^2}{G_N^2 m_s^3 m_E^2} &= -1 \\ \Rightarrow E_{mek} &= \frac{G_N^2 m_s^3 m_E^2}{2L^2} = -\frac{G_N^2 m_s m_E^2}{2(R_0 v_0)^2} \end{aligned} \quad (2.74)$$

where $L = v_0 R_0$ for a circular orbit. From this the velocity of the circular orbit can be determined:

$$E_{mek} = \frac{1}{2}m_s v_0^2 - \frac{G_N m_s m_E}{R_0} = -\frac{G_N^2 m_E^2 m_s}{2R_0^2 v_0^2} \Rightarrow v_0 = \sqrt{\frac{G_N m_E}{R_0}} \quad (2.75)$$

Using equation (2.75) in equation (2.73)

$$\begin{aligned} v_p^2 &= v_1^2 + 2v_0^2 \left(\frac{1}{0.91} - 1 \right) \\ &= v_0^2 + \frac{2m_1}{m_s} v_0^2 + 2v_0^2 \left(\frac{1}{0.91} - 1 \right) \\ &= v_0^2 \left[\left(\frac{2m_1}{m_s} \right)^2 + \frac{2}{0.91} - 1 \right]. \end{aligned} \quad (2.76)$$

v_p can be determined from conservation of angular momentum between ejection of m_1 and m_2 . Just after m_1 is ejected

$$\vec{L}_1 = (M - m_1) \vec{R}_0 \times \vec{v}_1. \quad (2.77)$$

The cross product multiplies R_0 with the part of \vec{v}_1 perpendicular to \vec{R}_0 , so

$$L_1 = (M - m_1) R_0 v_0. \quad (2.78)$$

Just before the Enterprise ejects the second mass, m_2 , at perigee

$$L_2 = (M - m_1)Rv_p = 0.91R_0v_p. \quad (2.79)$$

Equating equation (2.78) and (2.79)

$$v_0 = 0.91v_p. \quad (2.80)$$

Using equation (2.80) in equation (2.76) reveals

$$m_s = \frac{2 \cdot 0.91}{1 - 0.91}m_1 \simeq 20.22m_1. \quad (2.81)$$

With equation (2.81) in the bag only a relation between m_1 and m_2 is needed to determine the desired ratio. such an expression is obtained from conservation of momentum across the second mass ejection

$$\begin{aligned} (m_s - m_1)\vec{v}_p &= (m_s - m_1 - m_2)(\vec{v}_p + \Delta\vec{v}_2) + m_2\vec{u}_2 \\ &= (m_s - m_1)(\vec{v}_p + \Delta\vec{v}_2) + m_2(\vec{u}_2 - \vec{v}_p - \Delta\vec{v}_2) \\ &= (m_s - m_1)\vec{v}_{esc} + m_2\vec{v}_r. \end{aligned} \quad (2.82)$$

Since the mass is pushed directly backwards from the Enterprise this time

$$(m_s - m_1)v_p = (m_s - m_1)v_{esc} - m_2v_r, \quad (2.83)$$

where the direction has been extracted from \vec{v}_r . v_{esc} can be specified in terms of v_0 by using that v_{esc} is defined as the velocity which corresponds to $E_{mek} = 0$, i.e.

$$\begin{aligned} E_{mek} &= \frac{1}{2}(m_s - m_1 - m_2)v_{esc}^2 - \frac{G_N(m_s - m_1 - m_2)m_E}{R} = 0 \\ \Rightarrow v_{esc} &= \sqrt{\frac{2}{0.91}}v_0. \end{aligned} \quad (2.84)$$

Using equation (2.84), (2.75) and $v_r = 2v_0$ in equation (2.83)

$$m_s = m_1 + \frac{2}{\sqrt{\frac{2}{0.91} - \frac{1}{0.91}}}m_2 \simeq m_1 + 5.21m_2. \quad (2.85)$$

Equating equation (2.81) and (2.85)

$$m_2 \simeq \frac{1}{2} \left(\frac{2 \cdot 0.91}{1 - 0.91} - 1 \right) \left(\sqrt{\frac{2}{0.91}} - \frac{1}{0.91} \right) m_1 \simeq 3.69m_1. \quad (2.86)$$

Hereby, the ratio is given by

$$\text{Ratio 1} = \frac{m_1 + m_2}{m_s} \simeq 0.23. \quad (2.87)$$

2. Compare the previous ratio with the ratio $\frac{m_3}{m_s}$, where m_3 is the fuel needed for the Enterprise to escape the Earth's gravitational field directly from circular orbit of radius R_0 . m_3 is assumed to be ejected tangentially to the circular orbit.

When escaping directly from circular orbit

$$v_{esc} = \sqrt{\frac{2G_Nm_E}{R_0}} = \sqrt{2}v_0. \quad (2.88)$$

Conservation of linear momentum across the ejection of m_3 dictates

$$\begin{aligned} m_s\vec{v}_0 &= (m_s - m_3)(\vec{v}_0 + \Delta\vec{v}_3) + m_3\vec{u}_3 \\ &= m_s(\vec{v}_0 + \Delta\vec{v}_3) + m_3(\vec{u}_3 - \vec{v}_0 - \Delta\vec{v}_3) \\ &= m_s\vec{v}_{esc} + m_3\vec{v}_r. \end{aligned} \quad (2.89)$$

Since the mass is ejected directly backwards from the Enterprise

$$\begin{aligned}
m_s v_0 &= m_s v_{esc} - m_3 v_r \\
&= m_s \sqrt{\frac{2}{0.91}} v_0 - 2m_3 v_0 \\
\Rightarrow m_s &= \frac{2}{\sqrt{\frac{2}{0.91}} - 1} m_3 \simeq 4.15 m_3.
\end{aligned} \tag{2.90}$$

So, the new ratio is given by

$$\text{Ratio 2} = \frac{m_3}{m_s} \simeq 0.24. \tag{2.91}$$

Comparing equation equation (2.87) and (2.91) reveals a small difference. Escaping directly from circular orbit seems to require more fuel than escaping via an elliptical orbit. The reason for this is that whilst in the elliptical orbit the Enterprise is accelerated by the gravitational field of the planet. This procedure is well known and often used in space-travel. the procedure of using the gravitational field of an object in space to accelerate a spacecraft is called a gravitational assist.

2.6 Navier-Stokes Equation

An important application of Newtonian mechanics is to derive the fundamental equation of fluid dynamics; the Navier-Stokes equation. The Navier-Stokes equation is derived from Cauchy's equation which describes a generic fluid. Cauchy's equation can be derived from Newtons second law, $\vec{F} = m\vec{a}$, and is as such valid only in the non-relativistic limit- Just as Newtons second law. To derive Cauchy's equation the forces on a material particle, flowing along a stream in a fluid, is considered. The material particle is taken to be a small portion of the fluid which contains millions of molecules. Hence, the material particle is a small part of the fluid, but not at the quantum mechanics level (size molecule). Newtons second law for the material particle can be written as follows

$$\vec{F} = m_p \frac{d\vec{v}_p}{dt}, \tag{2.92}$$

where \vec{v}_p is the velocity and m_p is the mass - Both of the material particle. It is convenient to express the acceleration in terms of the velocity field $(v_f(x(t), y(t), z(t), t))$ of the fluid instead of the velocity of the material particle. In terms of the velocity field the derivative of the velocity of the particle is given by

$$\begin{aligned}
\frac{d\vec{v}_p}{dt} &= \frac{\partial \vec{v}_f}{\partial t} + \sum_{i=1}^3 \frac{\partial \vec{v}_f}{\partial x_i} \frac{\partial x_i}{\partial t} \\
&= \frac{\partial \vec{v}_f}{\partial t} + \vec{v}_f \cdot (\vec{\nabla} \cdot \vec{v}_f).
\end{aligned} \tag{2.93}$$

The first term in equation (2.93) accounts for the time-variation in the velocity field at a given point whereas the second term, called the advective acceleration, accounts for how the velocity field changes as a particle travels through the field (as seen from the perspective of the particle). Factorizing \vec{v}_f in equation (2.93) leads to the definition of the material derivative. The material derivative is a differential operator given by

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \vec{v}_f \cdot \vec{\nabla}. \tag{2.94}$$

By using the definition of the material derivative Newtons second law can be written in terms of the velocity field. Dividing through with the volume of the material particle and writing the force as a sum of surface forces (\vec{f}_s) and body forces (\vec{f}_b) results in

$$\rho_p \frac{D\vec{v}_f}{Dt} = \vec{f}_b + \vec{f}_s. \tag{2.95}$$

Among body forces is gravity, electromagnetic forces and fictitious forces. Among surface forces is pressure forces and viscous forces. The surface forces can be specified in terms of the stress tensor. The stress tensor can always be separated into two parts; a pressure part and a shear stress-part viz

$$\sigma_{ij} = -p\delta_{ij} + \tau_{ij}. \quad (2.96)$$

The surface forces are given in terms of the divergence of the stress tensor, i.e.

$$\begin{aligned} \vec{f}_s &= \vec{\nabla} \cdot \sigma \\ &= -\vec{\nabla} p + \sum_{i,j} \partial_j \tau_{ij} \hat{e}_i \\ &= - \begin{bmatrix} \frac{\partial p}{\partial x} \\ \frac{\partial p}{\partial y} \\ \frac{\partial p}{\partial z} \end{bmatrix} + \begin{bmatrix} \sum_{j=x,y,z} \frac{\partial \tau_{xj}}{\partial x_j} \\ \sum_{j=x,y,z} \frac{\partial \tau_{yj}}{\partial x_j} \\ \sum_{j=x,y,z} \frac{\partial \tau_{zj}}{\partial x_j} \end{bmatrix}. \end{aligned} \quad (2.97)$$

Using equation (2.97) in equation (2.95) results in Cauchy's equation

$$\rho_p \frac{D\vec{v}_f}{Dt} = \vec{f}_b - \vec{\nabla} p + \sum_{i,j} \partial_j \tau_{ij} \hat{e}_i. \quad (2.98)$$

Cauchy's equation (equation (2.98)) is valid for all non-relativistic fluids (gases and liquids). The Navier-Stokes equation is obtained from Cauchy's equation by approximating the fluid under consideration as isotropic, Newtonian and with a constant viscosity. Taking the fluid to be isotropic and Newtonian means that the different shear components of the stress tensor can be expressed in terms of velocity gradients since it is changes in the fluid velocity that creates stress in the fluid, i.e.

$$\begin{aligned} \text{Newtonian: } \sigma_{ij} &= \eta \frac{dv_x}{dy}, \\ \text{Newt.+Iso.: } \sigma_{ij} &= -p\delta_{ij} + \eta \left(\partial_i v_{f,j} + \partial_j v_{f,i} \left(\zeta - \frac{2}{3} \vec{\nabla} \cdot \vec{v}_f \right) \right), \end{aligned} \quad (2.99)$$

where "Newt" abbreviates "Newtonian", "Iso" abbreviates "isotropic", η is the dynamic viscosity and ζ is the bulk viscosity. The fact that the viscosity is constant means that it can be taken outside the derivative when the divergence of the stress tensor is considered. Taking the divergence of σ_{ij} from equation (2.99) (the Newtonian + isotropic one) and using it in equation (2.98) Cauchy's equation becomes the Navier-Stokes equation

$$\begin{aligned} \rho_p \frac{D\vec{v}_f}{Dt} &= \vec{f}_b - \vec{\nabla} p + \eta \left(\nabla^2 \vec{v}_f + \frac{1}{3} \vec{\nabla} (\vec{\nabla} \cdot \vec{v}_f) \right) + \zeta \vec{\nabla} (\vec{\nabla} \cdot \vec{v}_f) \\ &= \vec{f}_b - \vec{\nabla} (p - \zeta \vec{\nabla} \cdot \vec{v}_f) + \eta \left(\nabla^2 \vec{v}_f + \frac{1}{3} \vec{\nabla} (\vec{\nabla} \cdot \vec{v}_f) \right) \\ &= \vec{f}_b - \vec{\nabla} p_{mek} + \eta \left(\nabla^2 \vec{v}_f + \frac{1}{3} \vec{\nabla} (\vec{\nabla} \cdot \vec{v}_f) \right) \end{aligned} \quad (2.100)$$

where $p_{mek} \equiv p - \zeta \vec{\nabla} \cdot \vec{v}_f$ has been used. Equation (2.100) can be interpreted as follows The left hand side (the material acceleration) describes the acceleration a material particle experiences in terms of the velocity field. The right hand side describes the forces the material particle is subject to; body forces and surface forces. The body forces, which in most cases are just gravity, act to accelerate the fluid. The surface forces contains the pressure gradient, which denotes a force in the direction of lower pressure. Lastly there are the viscous forces. These create friction in the fluid and damps acceleration of the material particle. The Navier-Stokes equation (equation (2.100)) is not solve-able in the general case, and so approximations obtained from the physical scenario under consideration must be used in order to apply Navier-Stokes equation.

Example 2.9. —

An often used simplification of Navier-Stokes equation is to take the fluid as non-viscous, i.e. $\eta \simeq \zeta \simeq 0$.
Hereby

$$\rho_p \frac{D\vec{v}_f}{Dt} \simeq \vec{f}_b - \vec{\nabla}p. \quad (2.101)$$

Equation (2.101) is called Eulers equation.

Example 2.10. —

Another often used simplification of Navier-Stokes equation is to take the fluid flow as steady, i.e. $\partial_t \vec{v}_f \simeq \vec{0}$.
Taking the fluid to be both non-viscous and the flow as steady means the Navier-Stokes equation simplifies as follows

$$\rho_p \vec{v}_f \cdot (\vec{\nabla} \cdot \vec{v}_f) \simeq \vec{f}_b - \vec{\nabla}p. \quad (2.102)$$

Equation (2.102) is equivalent to Bernoullis equation if gravity is taken as the only body force, i.e. $\vec{f}_b = \vec{f}_g = \rho_p \vec{g}$. To see this use that $\vec{v}_f \cdot (\vec{\nabla} \cdot \vec{v}_f) = \frac{1}{2} \vec{\nabla} \cdot (\vec{v}_f^2) - \vec{v}_f \times (\vec{\nabla} \times \vec{v}_f)$ and recall that Bernoullis equation is along a streamline. A streamline is the line where the velocity field at any given point is tangent to the line. To acquire the equation of motion along the streamline dot equation (2.102), with the vector identity, with an infinitesimal line element, $d\vec{s}$, along the streamline. Since the streamline is tangent to $d\vec{s}$, $d\vec{s}$ is parallel with the velocity field, i.e. $d\vec{s} \parallel \vec{v}_f$. $\vec{v}_f \times (\vec{\nabla} \times \vec{v}_f)$ results in a vector perpendicular to \vec{v}_f , so $d\vec{s} \cdot [\vec{v}_f \times (\vec{\nabla} \times \vec{v}_f)] = 0$.
Hereby

$$\frac{\rho_p}{2} d\vec{s} \cdot \vec{\nabla}(\vec{v}_f^2) \simeq \rho_p \vec{g} - \vec{\nabla}p. \quad (2.103)$$

Use then that

$$d\vec{s} \cdot \vec{\nabla}(\vec{v}_f^2) = [dx \quad dy \quad dz] \begin{bmatrix} \frac{\partial(v_f^2)}{\partial x} \\ \frac{\partial(v_f^2)}{\partial y} \\ \frac{\partial(v_f^2)}{\partial z} \end{bmatrix} = d(v_f^2). \quad (2.104)$$

Similarly to equation (2.104) $d\vec{s} \cdot \vec{\nabla}p = dp$. Using that $\vec{g} = -g\hat{e}_z$; $d\vec{s} \cdot \vec{g} = -dzg$ Navier-Stokes equation reduces to

$$\frac{\rho_p}{2} d(v_f^2) + \rho_p dzg + dp = 0 \Rightarrow \frac{\rho_p}{2} v_f^2 + \rho_p gz + p = const. \quad (2.105)$$

Equation (2.105) is Bernoullis equation. Bernoullis equation represents energy conservation.

Example 2.11. —

Consider an open wine barrel standing up. The barrel has a hole in the side from which wine can flow out of the barrel. Compute the velocity of the escaping wine as a function of the hight of the wine column above the hole of the barrel.

In order to derive the velocity of the fluid as it escapes the barrel, I consider a material particle flowing along a streamline. Taking the flow to be steady and the fluid to be ideal (non-viscous), Bernoullis equation can be used. At the top of the barrel the velocity of the fluid particle will approximately be zero whilst the particle will have a potential energy related to the vertical distance from the particle to the hole in the barrel. I denote the vertical hight of the top of the wine relative to the hole in the barrel to be z . Hereby, at the top of the barrel

$$\rho_p gz + p_{ambient} = const. \quad (2.106)$$

As the particle exits the hole, the potential energy will have been turned into kinetic energy. Therefore

$$\frac{\rho_p v^2}{2} + p_{ambient} = const. \quad (2.107)$$

Equating equation (2.106) and (2.107) reveals $v = \sqrt{2gz}$.

2.6.1 Computational Fluid Dynamics

Few real-world problems are solveable in fluid dynamics due to the complexity of the governing equations. In order to apply the theoretical knowledge to real problems the theoretical equations are therefore discretized and solved numerically as opposed to obtaining an analytical solution. The branch dealing with this procedure is computational fluid dynamics (CFD).

Discretization of Navier-Stokes Equations

The Navier-Stokes equations can only be solved analytically under both simplifying assumptions and settings. In order to apply the Navier-Stokes equations to the diverse set of problems relevant for engineering, the equations must be discretized and solved numerically. In order to do so it is expedient to reformulate the Navier-Stokes equations as follows

$$\rho \frac{\partial \vec{v}}{\partial t} = \vec{f} - \vec{\nabla} p, \quad (2.108)$$

where it is now, and for future reference, understood that $\rho = \rho_p$, $\vec{v} = \vec{v}_f$, $\vec{f} = \vec{f}_{gen}$ and

$$\vec{f} = \vec{f}_b - \rho(\vec{v} \cdot \vec{\nabla})\vec{v} + \eta \left(\nabla^2 \vec{v} + \frac{1}{3} \vec{\nabla}(\vec{\nabla} \cdot \vec{v}) \right) + \zeta \vec{\nabla}(\vec{\nabla} \cdot \vec{v}) \quad (2.109)$$

is a generalized force. The discretization of equations (2.108) and (2.109) has to be performed both in time and space. Since there is only one time-coordinate the discretization of the time derivative proceeds as follows

$$\frac{\partial h(\vec{x}, t)}{\partial t} \simeq \frac{h(\vec{x}, t + \Delta t) - h(\vec{x}, t)}{\Delta t}, \quad (2.110)$$

where h is a generic function used for these definitions. For the spatial derivatives there are more subtleties to consider since there are several coordinates that are to be discretized "together". The "issue" is that when the derivative is discretized it refers to the midway-point between the two extremes. E.g. the time derivative refer to the time stamp $t + \frac{\Delta t}{2}$. This is not an issue since there is only one time coordinate. For the spatial coordinates this is however an issue. To see this consider the naive discretizations of the spatial derivatives

$$\frac{\partial h(x_i, t)}{\partial x_i} \simeq \frac{h(x_i + \Delta x_i, x_{j \neq i}, t) - h(x_i, x_{j \neq i}, t)}{\Delta x_i}. \quad (2.111)$$

These derivatives belong to the points $x_i + \frac{\Delta x_i}{2}$. This is an issue because, as is evident from equation (2.109), it is necessary to consider terms like the divergence $\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}$. In order for the divergence to make sense, the derivatives w.r.t. each spatial component must refer to the *same* point and so by extension the velocity components themselves cannot belong to the same point. Instead the velocity points are placed on different, interwoven grids which together make up what is called a *staggered grid*. A cell on the staggered grid is a region which contain the different points corresponding to the same area (see figure 2.9). Hence $v_x[1, 1, 1]$ actually refer to a (slightly) different spatial location than $v_y[1, 1, 1]$.

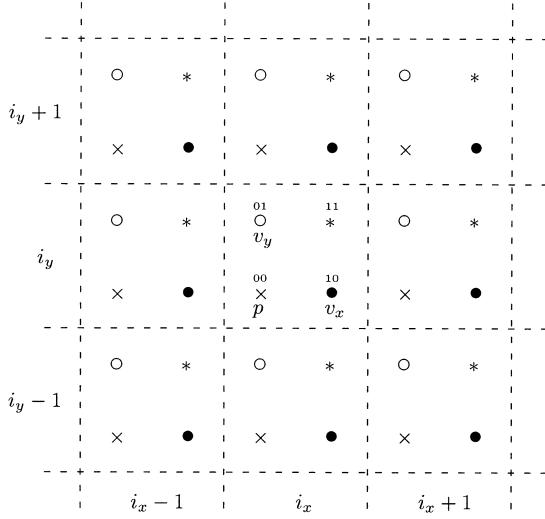


Figure 2.9: The placement of the pressure and velocity variables on a 2D staggered grid. The dashed squares denote the cells. The figure is taken from Lautrup (2011).

With this information in hand equation (2.108) can be discretized as follows

$$\rho \begin{bmatrix} \frac{v_x[i_x, i_y, i_z, t+1] - v_x[i_x, i_y, i_z, t]}{\Delta t} \\ \frac{v_y[i_x, i_y, i_z, t+1] - v_y[i_x, i_y, i_z, t]}{\Delta t} \\ \frac{v_z[i_x, i_y, i_z, t+1] - v_z[i_x, i_y, i_z, t]}{\Delta t} \end{bmatrix} = \begin{bmatrix} f_x[i_x, i_y, i_z, t] - \frac{p[i_x+1, i_y, i_z, t] - p[i_x, i_y, i_z, t]}{\Delta x} \\ f_y[i_x, i_y, i_z, t] - \frac{p[i_x, i_y+1, i_z, t] - p[i_x, i_y, i_z, t]}{\Delta y} \\ f_z[i_x, i_y, i_z, t] - \frac{p[i_x, i_y, i_z+1, t] - p[i_x, i_y, i_z, t]}{\Delta z} \end{bmatrix}. \quad (2.112)$$

Note that the first equation is located on the same grid-point as v_x , the second on the same grid-point as v_y and the third on the same grid-point as v_z . From the derivative of the pressure it is clear that the v_z point must be orthogonal to the pressure point³. This information is important since the different components of the generalized force is forced to be located on the grid-points corresponding to the rest of the given equation. Writing out the generalized force

$$\vec{f} = \begin{bmatrix} f_{b,x} - \rho(v_x \frac{\partial v_x}{\partial x} + v_y \frac{\partial v_x}{\partial y} + v_z \frac{\partial v_x}{\partial z}) + \eta \left[\frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_x}{\partial y^2} + \frac{\partial^2 v_x}{\partial z^2} + (\frac{1}{3} + \frac{\zeta}{\eta}) \left(\frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_y}{\partial x \partial y} + \frac{\partial^2 v_z}{\partial x \partial z} \right) \right] \\ f_{b,y} - \rho(v_x \frac{\partial v_y}{\partial x} + v_y \frac{\partial v_y}{\partial y} + v_z \frac{\partial v_y}{\partial z}) + \eta \left[\frac{\partial^2 v_y}{\partial x^2} + \frac{\partial^2 v_y}{\partial y^2} + \frac{\partial^2 v_y}{\partial z^2} + (\frac{1}{3} + \frac{\zeta}{\eta}) \left(\frac{\partial^2 v_x}{\partial y \partial x} + \frac{\partial^2 v_y}{\partial y^2} + \frac{\partial^2 v_z}{\partial y \partial z} \right) \right] \\ f_{b,z} - \rho(v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z}) + \eta \left[\frac{\partial^2 v_z}{\partial x^2} + \frac{\partial^2 v_z}{\partial y^2} + \frac{\partial^2 v_z}{\partial z^2} + (\frac{1}{3} + \frac{\zeta}{\eta}) \left(\frac{\partial^2 v_x}{\partial z \partial x} + \frac{\partial^2 v_y}{\partial z \partial y} + \frac{\partial^2 v_z}{\partial z^2} \right) \right] \end{bmatrix}. \quad (2.113)$$

In order for the advective term ($\propto \rho$) to belong to the same point as v_x , the derivative must be averaged over the adjacent points, meaning e.g.

$$\begin{aligned} v_x \frac{\partial v_x}{\partial x} &\rightarrow v_x[i_x, i_y, i_z, t] \frac{v_x[i_x+1, i_y, i_z, t] - v_x[i_x-1, i_y, i_z, t]}{2\Delta x}, \\ v_y \frac{\partial v_x}{\partial y} &\rightarrow \langle v_y \rangle \frac{v_x[i_x, i_y+1, i_z, t] - v_x[i_x, i_y-1, i_z, t]}{2\Delta y}, \\ v_z \frac{\partial v_x}{\partial z} &\rightarrow \langle v_z \rangle \frac{v_x[i_x, i_y, i_z+1, t] - v_x[i_x, i_y, i_z-1, t]}{2\Delta z}, \end{aligned} \quad (2.114)$$

where

$$\begin{aligned} \langle v_y \rangle &\equiv \frac{v_y[i_x, i_y, i_z, t] + v_y[i_x, i_y-1, i_z, t] + v_y[i_x+1, i_y, i_z, t] + v_y[i_x+1, i_y-1, i_z, t]}{4}, \\ \langle v_z \rangle &\equiv \frac{v_z[i_x, i_y, i_z, t] + v_z[i_x, i_y, i_z-1, t] + v_z[i_x, i_y+1, i_z, t] + v_z[i_x, i_y+1, i_z-1, t]}{4}. \end{aligned} \quad (2.115)$$

³ v_z is out of the screen on figure 2.9 since $i_z + 1$ minus i_z must land on the pressure point.

The second derivatives are with respect to the same variable are naturally located on the same point e.g.

$$\begin{aligned}\frac{\partial^2 v_x}{\partial x^2} &\rightarrow \frac{v_x[i_x + 1, i_y, i_z, t] + v_x[i_x - 1, i_y, i_z, t] - 2v_x[i_x, i_y, i_z, t]}{\Delta x^2}, \\ \frac{\partial^2 v_x}{\partial y^2} &\rightarrow \frac{v_x[i_x, i_y + 1, i_z, t] + v_x[i_x, i_y - 1, i_z, t] - 2v_x[i_x, i_y, i_z, t]}{\Delta x^2}, \\ \frac{\partial^2 v_x}{\partial z^2} &\rightarrow \frac{v_x[i_x, i_y, i_z + 1, t] + v_x[i_x, i_y, i_z - 1, t] - 2v_x[i_x, i_y, i_z, t]}{\Delta x^2}.\end{aligned}\quad (2.116)$$

For the divergence term it is less simple for the terms with mixed derivatives. Here the average of both derivatives must be used such that

$$\begin{aligned}\frac{\partial^2 v_x}{\partial x \partial y} &\rightarrow \frac{1}{4\Delta x \Delta y} \left(v_x[i_x + 1, i_y + 1, i_z, t] + v_x[i_x - 1, i_y - 1, i_z, t] \right. \\ &\quad \left. - v_x[i_x + 1, i_y - 1, i_z, t] - v_x[i_x - 1, i_y + 1, i_z, t] \right),\end{aligned}\quad (2.117)$$

$$\begin{aligned}\frac{\partial^2 v_x}{\partial x \partial z} &\rightarrow \frac{1}{4\Delta x \Delta z} \left(v_x[i_x + 1, i_y, i_z + 1, t] + v_x[i_x - 1, i_y, i_z - 1, t] \right. \\ &\quad \left. - v_x[i_x + 1, i_y, i_z - 1, t] - v_x[i_x - 1, i_y, i_z + 1, t] \right),\end{aligned}\quad (2.118)$$

The specific approach (read programming) to obtaining a solution from the discretized equations of motion depend on the physical scenario under consideration and potential simplifying assumptions that can be made.

Example 2.12.

The perhaps most used simplification of Navier-Stokes equations is to take the fluid to be incompressible ($\rho \simeq \text{const}$). Taking the density to be constant means the velocity field will be divergence-less. To see this, consider a material particle riding along with the flow in a fluid. At time $t + \delta t$ the particle will be displaced by an amount $\vec{x} + v(\vec{x}, t)\delta t$. The change in mass density during this time

$$\begin{aligned}\delta\rho &= \rho(\vec{x} + v(\vec{x}, t)\delta t, t + \delta t) - \rho(\vec{x}, t) \\ &= v_x \delta t \frac{\partial \rho(\vec{x}, t)}{\partial x} + v_y \delta t \frac{\partial \rho(\vec{x}, t)}{\partial y} + v_z \delta t \frac{\partial \rho(\vec{x}, t)}{\partial z} + \delta t \frac{\partial \rho(\vec{x}, t)}{\partial t} + \mathcal{O}(\delta t^2) \\ &= \left(\frac{\partial \rho}{\partial t} + (\vec{v} \cdot \vec{\nabla})\rho \right) \delta t.\end{aligned}\quad (2.119)$$

Dividing by δt reveals the rate of change of density as seen from the material particle, i.e. the material derivative of the density

$$\frac{D\rho}{Dt} = \frac{\partial \rho}{\partial t} + (\vec{v} \cdot \vec{\nabla})\rho. \quad (2.120)$$

Imposing conservation of mass means

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}(\rho \vec{v}) = 0, \quad (2.121)$$

with $\vec{\nabla}(\rho \vec{v}) = (\vec{v} \cdot \vec{\nabla})\rho + \rho \vec{\nabla} \cdot \vec{v}$. Combining equation (2.120) and (2.121) reveals

$$\frac{D\rho}{Dt} + \rho \vec{\nabla} \cdot \vec{v} = 0. \quad (2.122)$$

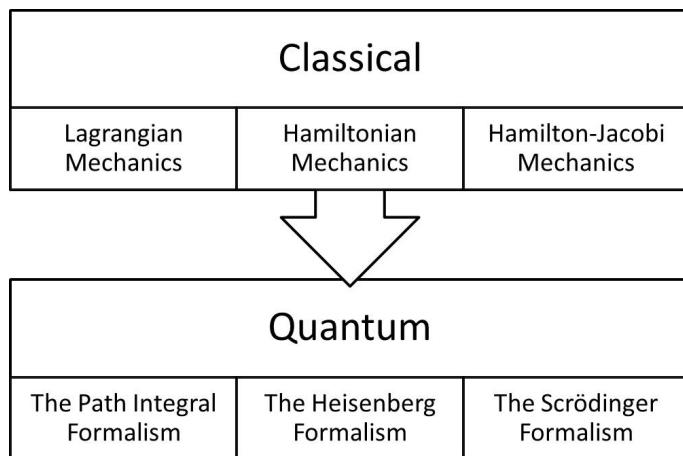
From equations (2.122) and (2.120) it is clear that by taking $\rho = \text{const}$ the material derivative vanish and by extension it is required that the velocity field is divergence-free ($\vec{\nabla} \cdot \vec{v} = 0$). This information can be used to simplify the Navier-Stokes equations. Taking the divergence of equation (2.108)

$$\vec{\nabla}^2 p = \vec{\nabla} \cdot \vec{f} \equiv s, \quad (2.123)$$

where s abbreviates "source".

3 | Analytical Mechanics

Analytical mechanics, derived primarily by Joseph-Louis Lagrange and William Rowan Hamilton in the late 1700's and early 1800's, is a framework equivalent to Newtonian mechanics for conservative forces¹. The framework consists of three equivalent schemes; Lagrangian mechanics, Hamiltonian mechanics and Hamilton-Jacobi mechanics. These schemes provide the foundation for three of the quantum mechanical formalisms. These generalizations will be considered in chapter 6. The schemes of analytical mechanics was



derived with the intention to investigate the fundamental processes in nature that led to Newtons laws. This quest was by and large successful and significant physical insight can be gained by studying the schemes of analytical mechanics.

3.1 Lagrangian Mechanics

The Lagrangian formalism was introduced by Joseph-Luis Lagrange in 1788 (~ 100 years after Newtons laws). Back in the days scientists were concentrated on finding a more general formulation of the principles of physics - Compared to the Newtonian formalism. Newtons laws provide instructions to acquire the laws of motion, but they do not provide an explanation as to *why* the laws are in the way they are. An explanation as to why came from developing Lagrangian mechanics; everything in physics derives from the principle of least action!

The principle of least action

Consider a generic physical system which has an instantaneous configuration described by the values of n generalized coordinates, $\bar{q} \equiv (q_1, q_2, \dots, q_n)$. Because the generalized coordinates determine the configuration of the system, they span what is called *configuration space*, and the physical state of the system is given as a point in this n -dimensional hyperspace (Goldstein et al., 2002). As time progresses, the state of the general physical system will change. For the generalized coordinates, this change of state results in the point tracing

¹The Lagrangian formalism utilizes the potential energy function which is only defined for a conservative force field. Non-conservative forces can be incorporated with much difficulty however.

out a line (or path), called the path of motion, in configuration space. Which path of motion the system will take is described by *Hamiltons principle*, also called the principle of stationary (or least) action.

The principle of stationary action: *The motion of the system from time t_1 to t_2 is such that the action (S) is extremized (minimized most often)²*

The action is defined by the integral (Sannino)

$$S[\bar{q}, \dot{\bar{q}}, t_1, t_2] = \int_{t_1}^{t_2} L(\bar{q}(t), \dot{\bar{q}}(t), t) dt, \quad (3.1)$$

where the square brackets indicate that the action is a *functional* rather than a *function*³. The function L is the *Lagrangian* and is defined as

$$L(\bar{q}(t), \dot{\bar{q}}(t), t) = T(\dot{\bar{q}}(t)) - V(\bar{q}(t), t), \quad (3.2)$$

where T is the kinetic energy of the system and V is the potential energy of the system.

The Euler-Lagrange equation

By minimizing the action of equation (3.1) the master equation of Lagrangian mechanics, the Euler-Lagrange equation, is obtained. Consider a variation of the action:

$$\begin{aligned} \delta S[\bar{q}, \dot{\bar{q}}, t_1, t_2] &= \delta \int_{t_1}^{t_2} L(\bar{q}(t), \dot{\bar{q}}(t), t) dt \\ &= \int_{t_1}^{t_2} \sum_{j=1}^n \left(\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial q_j(t)} \delta q_j(t) + \frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \delta \dot{q}_j(t) \right) dt \\ &= \int \sum_j \frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial q_j(t)} \delta q_j(t) dt + \int \sum_j \frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \delta \dot{q}_j(t) dt, \end{aligned} \quad (3.3)$$

where $\dot{\bar{q}}(t) = (\dot{q}_1(t), \dot{q}_2(t), \dots, \dot{q}_n(t))$ and $\bar{q}(t) = (q_1(t), q_2(t), \dots, q_n(t))$. The last term is evaluated by partial integration

$$\begin{aligned} \int \sum_j \frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \delta \dot{q}_j(t) dt &= \int \sum_j \frac{d}{dt} \left(\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \delta q_j(t) \right) dt \\ &\quad - \int \sum_j \frac{d}{dt} \left(\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \right) \delta q_j(t) dt, \end{aligned} \quad (3.4)$$

where

$$\int_{t_1}^{t_2} \sum_j \frac{d}{dt} \left(\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \delta q_j(t) \right) dt = \sum_j \frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \delta q_j(t) \Big|_{t_1}^{t_2} = 0. \quad (3.5)$$

Equation (3.5) vanish by taking the endpoints of the path to be fixed, i.e. $\delta q_j(t_1) = \delta q_j(t_2) = 0$ is taken as boundary conditions. By using equation (3.5) in equation (3.4) and the result in equation (3.3) the variation in the action can be written as

$$\delta S[\bar{q}, \dot{\bar{q}}, t_1, t_2] = \int \sum_j \left(\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial q_j(t)} - \frac{d}{dt} \left(\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \right) \right) \delta q_j(t) dt. \quad (3.6)$$

²The reason why the system traces the path in configuration space that corresponds to the minimized action can be explained from quantum mechanics and the path integral scheme. In the path integral scheme the transition probability is given in terms of the path integral of $e^{\frac{iS}{\hbar}}$ (SI units). In the classical limit ($\hbar \rightarrow 0$ the exponential will fluctuate wildly and all contributions besides those which corresponds to a minimized action will cancel cf. the random phase principle.)

³A function is a rule that assigns a unique element, $f(x)$, from a set to each value, x , in another set; a function takes in a number and produces a number. A functional, on the other hand, associates a vector valued function with a real number.

To minimize the action (find a minimum in S), the variation (equation (3.6)) must vanish. In order for the variation to vanish for an arbitrary $\delta q_j(t)$, the remaining part of the integrand must always be zero. That is

$$\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial q_j(t)} - \frac{d}{dt} \left(\frac{\partial L(\bar{q}(t), \dot{\bar{q}}(t), t)}{\partial \dot{q}_j(t)} \right) = 0. \quad (3.7)$$

Equation (3.7) is called the *Euler-Lagrange* equation. It is often written in the less detailed form

$$\frac{\partial L}{\partial q_j} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) = 0. \quad (3.8)$$

The Euler-Lagrange equation is a second order differential equation in time, and from it the equations of motion can be determined by specifying, and inserting, the Lagrangian. Thus, the Euler-Lagrange equation is equivalent to Newtonian dynamics, but instead of postulating a series of laws, the equations of motions are derived from the assumption that the path of the system in configuration space is the path of least action. This provides profound, new physical insight into the machinery of nature.

Example 3.1.

Determine and solve the EOM for a movable, simple pendulum.

The movable, simple pendulum consist of two particles; one particle (m_1) is confined to a horizontal movement, whereas another particle (m_2) is attached to the first via a (massless) string, and is free to swing back and forth - See figure 3.1. In terms of the, from the figure, defined coordinate system, the coordinates

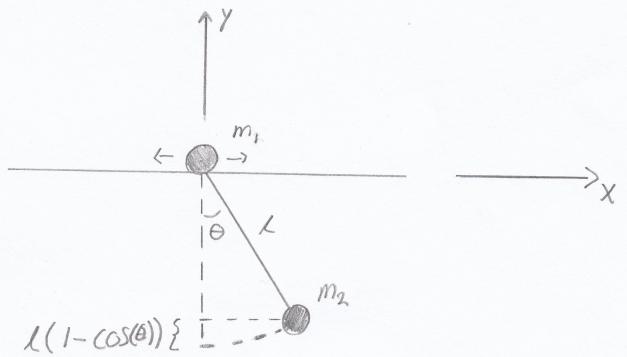


Figure 3.1: The spherical pendulum with spherical coordinates.

are given by

$$x_1 = x_1, \quad y_1 = 0, \quad (3.9)$$

$$x_2 = x_1 + l \sin(\theta), \quad y_2 = -l \cos(\theta), \quad (3.10)$$

where the minus sign in y_2 is there because the second particle is defined to swing at negative y . For the potential energy in the Lagrangian the relevant distance⁴ is the vertical displacement from equilibrium ($l - y_2$), so the Lagrangian is given by

$$L = \frac{1}{2} m_1 \dot{x}_1^2 + \frac{1}{2} m_2 (\dot{x}_2^2 + \dot{y}_2^2) - m_2 g l (1 - \cos(\theta)). \quad (3.11)$$

The velocity of the second particle is

$$\begin{aligned} \dot{x}_2^2 + \dot{y}_2^2 &= \dot{x}_1^2 + 2 \dot{x}_1 l \cos(\theta) \dot{\theta} + l^2 \cos(\theta)^2 \dot{\theta}^2 + l^2 \sin(\theta)^2 \dot{\theta}^2 \\ &= \dot{x}_1^2 + l^2 \dot{\theta}^2 + 2 l \dot{x}_1 \dot{\theta} \cos(\theta). \end{aligned} \quad (3.12)$$

⁴So that the potential energy is positive when the mass is displaced from equilibrium.

So

$$L = \frac{1}{2}m_1\dot{x}_1^2 + \frac{1}{2}m_2(\dot{x}_1^2 + l^2\dot{\theta}^2 + 2l\dot{x}_1\dot{\theta}\cos(\theta)) - m_2gl(1 - \cos(\theta)). \quad (3.13)$$

Using x_1 and θ as generalized coordinates, the EOM can be found from the Euler-Lagrange equation

$$\underbrace{\frac{\partial L}{\partial x_1}}_{=0} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_1}\right) = 0 \Rightarrow \dot{x}_1(m_1 + m_2) + m_2l\cos(\theta)\dot{\theta} = const = p_{x_1}, \quad (3.14)$$

$$0 = \frac{\partial L}{\partial \theta} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) \Rightarrow -m_2lsin(\theta)(\dot{x}_1\dot{\theta} + g) - m_2l(l\ddot{\theta} + \ddot{x}_1\cos(\theta) - \dot{x}_1\dot{\theta}\sin(\theta)) = 0. \quad (3.15)$$

Two of the terms $(\dot{x}_1\dot{\theta})$ cancel, so

$$m_2l^2\ddot{\theta} + m_2glsin(\theta) + m_2l\ddot{x}_1\cos(\theta) = 0. \quad (3.16)$$

Eliminate \ddot{x}_1 from the above equation by using the Euler-Lagrange equation for x_1 . Hereby

$$\begin{aligned} 0 &= \ddot{x}_1(m_1 + m_2) + m_2l(\cos(\theta)\ddot{\theta} - \sin(\theta)\dot{\theta}^2) \\ \Rightarrow \ddot{x}_1 &= \frac{m_2l}{m_1 + m_2}(\dot{\theta}^2\sin(\theta) - \ddot{\theta}\cos(\theta)). \end{aligned} \quad (3.17)$$

So

$$m_2l^2\ddot{\theta} - \frac{m_2^2l^2\cos(\theta)^2}{m_1 + m_2}\ddot{\theta} + \frac{m_2^2l^2\cos(\theta)\sin(\theta)}{m_1 + m_2}\dot{\theta}^2 + m_2lgsin(\theta) = 0. \quad (3.18)$$

This is in general difficult to solve, so I consider the limiting cases

$$m_2 \gg m_1 \wedge \theta = \text{small: } \dot{\theta}^2 + \frac{g}{l} \approx 0 \Rightarrow \theta \approx \sqrt{-\frac{g}{l}}t + A. \quad (3.19)$$

Since $g > 0$ the above EOM is complex. This makes sense physically since if $m_2 \gg m_1$, m_2 will drag m_1 along and no swinging motion at small angles will occur. Hence, $m_2 \gg m_1$ and $\theta = \text{small}$ are incompatible approximations. Instead I consider $m_2 \ll m_1 \wedge \theta = \text{small}$. This corresponds to a small mass swinging from a large mass. In the limit of large m_1 this should simplify to the simple pendulum since at small angles m_1 should not move. Indeed this is what I find

$$m_2 \ll m_1 \wedge \theta = \text{small: } \ddot{\theta} + \frac{g}{l}\theta \approx 0 \Rightarrow \theta \approx \theta_0 \cos\left(\sqrt{\frac{g}{l}}t\right). \quad (3.20)$$

Since the equations of motion falls out directly from the Euler-Lagrange equation, and since the equations of motion in Newtonian mechanics are directly related to the rate of change of momentum, it is no surprise that the momentum, and the rate of change of momentum, is related to the Euler-Lagrange equation. However, since the Euler-Lagrange equation deals with generalized coordinates, the momentum that is referred to is not necessarily the linear momentum. If for example a system of polar coordinates are considered, the momentum derived from the Euler-Lagrange equation is the angular momentum. Thus, the momentum related to the Euler-Lagrange equation is called the canonical momentum conjugate to the generalized coordinate q_j , p_j , and it is defined as

$$p_j = \frac{\partial L}{\partial \dot{q}_j}. \quad (3.21)$$

By using this definition in the Euler-Lagrange-equation, it directly follows that

$$\dot{p}_j = \frac{\partial L}{\partial q_j}. \quad (3.22)$$

Using equation (3.21) and (3.22) in the Euler-Lagrange equation it is clear that it (the Euler-Lagrange equation) stipulates $\dot{p} - \dot{p} = 0$, which is indeed the case. Memo-technically this is a good thing to remember

because one of the quantities are to be differentiated wrt. time, so if the equation must read $\dot{p} - \ddot{p} = 0$ it can be deduced that $\frac{\partial L}{\partial \dot{q}_j}$, which is to be differentiated wrt. time in the Euler-Lagrange equation, must equal p . From equation (3.22) it follows that if the Lagrangian does not depend on a given coordinate, q_j , the corresponding momentum is conserved; $\frac{\partial p_j}{\partial t} = 0 \Rightarrow p_j = \text{conserved}$. An alternative way to formulate the principle of conserved quantities is via Noethers theorem.

Noethers theorem: *Transformations under which the Lagrangian is unchanged have a conserved quantity associated to them.*

Example 3.2.

Consider a translational invariant physical system described by the Lagrangian $L(x_i, \dot{x}_i)$. A constant displacement of x_i will leave the Lagrangian invariant. Hence, the Lagrangian is invariant under the transformation

$$x_i \Rightarrow x_i + \delta x, \quad (3.23)$$

where δx is a constant. The change in the Lagrangian under this transformation can be described as

$$\delta L = \frac{\partial L}{\partial x_i} \delta x_i + \frac{\partial L}{\partial \dot{x}_i} \delta \dot{x}_i = \delta x \frac{\partial L}{\partial x_i} = \delta x \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = 0, \quad (3.24)$$

where $\delta L = 0$ cf. the invariance of L under the transformation and the Euler-Lagrange equation has been used for the last equality. Since $\delta x \neq 0$ in general, $\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = 0$ must hold and so $\frac{\partial L}{\partial \dot{x}_i} = p_i = \text{const}$ w.r.t. time. This example therefore illustrates how invariance of the Lagrangian, under a constant spatial translation, leads to conservation of linear momentum.

Example 3.3.

Consider a Lagrangian that is invariant under the infinitesimal rotation around some vector, $\delta \vec{\phi} (= \epsilon \hat{n})$ i.e. the Lagrangian that is invariant under

$$\delta x_i = (\delta \vec{\phi} \times \vec{r})_i, \quad (3.25)$$

where \vec{r} is the radius vector in spherical coordinates. Then change in the Lagrangian under this transformation can be described viz

$$\begin{aligned} \delta L &= \frac{\partial L}{\partial x_i} \delta x_i + \frac{\partial L}{\partial \dot{x}_i} \delta \dot{x}_i \\ &= \frac{\partial L}{\partial x_i} (\delta \vec{\phi} \times \vec{r})_i + \frac{\partial L}{\partial \dot{x}_i} (\delta \vec{\phi} \times \dot{\vec{r}})_i \\ &= \dot{\vec{p}} \cdot (\delta \vec{\phi} \times \vec{r}) + \vec{p} \cdot (\delta \vec{\phi} \times \dot{\vec{r}}) \\ &= \delta \vec{\phi} \cdot [(\vec{r} \times \dot{\vec{p}}) + (\dot{\vec{r}} \times \vec{p})] \\ &= \delta \vec{\phi} \cdot \frac{d}{dt} (\vec{r} \times \vec{p}) = 0. \end{aligned} \quad (3.26)$$

Since $\delta \vec{\phi} \neq \vec{0}$ in general; $\frac{d}{dt} (\vec{r} \times \vec{p}) = \vec{0}$ must hold and so $\vec{L} = \vec{r} \times \vec{p} = \text{const}$. This example illustrates how invariance of the Lagrangian, under a constant spatial rotation, leads to conservation of angular momentum.

Example 3.4.

Consider the transformation $q \Rightarrow q + \epsilon$, where ϵ is an infinitesimal, constant parameter. Why does this transformation not change the EOM?

Invariance in the EOM means $\delta S = \delta S' = 0$, where S and S' refers to the action before and after transformation. To show that this is indeed the case, consider the Taylor expansion of the transformed Lagrangian

$$\begin{aligned} L(q + \epsilon, \dot{q}) &= L(q, \dot{q}) + \frac{\partial L}{\partial q} \epsilon + \mathcal{O}(\epsilon^2) \\ &= L(q, \dot{q}) + \dot{p} \epsilon, \end{aligned} \quad (3.27)$$

where equation (3.21) has been used for the second equality and $\mathcal{O}(\epsilon^2)$ has been neglected since ϵ is infinitesimal. The corresponding action is given by

$$\begin{aligned} S' &= \int dt L(q + \epsilon, \dot{q}) \\ &= \int dt L(q, \dot{q}) + \int dt \dot{p}\epsilon \\ &= S + \epsilon \left(p(q_t, t_2) - p(q_1, t_1) \right). \end{aligned} \quad (3.28)$$

S and S' differs by a quantity which depends on the endpoints of the path. Since the path endpoints are kept fixed, this vanishes upon variation. Therefore $\delta S = \delta S' = 0$.

The result of example 3.2 can be generalized as follows; it is in general true that Lagrangians, which differ by a time-derivative of a function of coordinates and time, leads to the same EOM. Consider

$$L'(q, \dot{q}, t) = L(q, \dot{q}, t) + \frac{df(q, t)}{dt}, \quad (3.29)$$

where f is a generic function of the coordinates and time. In general L and L' will lead to the same EOM since $\delta S = \delta S' = 0$ cf. the above arguments. Therefore the Lagrangian is only defined up to a total derivative of some generic function of the coordinates and time.

3.2 Hamiltonian Mechanics

The Lagrangian is a function of the generalized coordinates which defines a point in configuration space. Time evolution of the system traces out a path in configuration space. However, the state of the system is defined by $\{q_i\}$ and $\{p_i\}$ in the sense that this information enables a determination of the system state at all times in the future. The pair $\{q_i, p_i\}$ defines a point in $2n$ -dimensional *phase space*⁵. The desire is therefore to find the function (the Hamiltonian) on phase space that will determine the unique evolution of $\{q_i\}$ and $\{p_i\}$ and contain the same information as the Lagrangian. Since $\{q_i, p_i\}$ and $\{q_i, \dot{q}_i\}$ are conjugate sets of variables, the Hamiltonian can be determined from the Lagrangian via a transformation called a *Legendre transformation*. A Legendre transformation converts a function of a set of variables to another, conjugate, set of variables. The Legendre transformation from $L \Rightarrow H$ is done in the following way; take $L = L(q_i, \dot{q}_i)$. Hereby

$$\begin{aligned} dL &= \sum_i \frac{\partial L}{\partial q_i} dq_i + \sum_i \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i \\ &= \sum_i \dot{p}_i dq_i + \sum_i p_i d\dot{q}_i. \end{aligned} \quad (3.30)$$

The variables of the function is expressed as differentials. In order to shift from $(q_i, \dot{q}_i) \Rightarrow (q_i, p_i)$ use that $\sum_i p_i d\dot{q}_i = d(\sum_i p_i \dot{q}_i) - \sum_i \dot{q}_i dp_i$, take the differential $d(\sum_i p_i \dot{q}_i)$ to the left hand side and reverse the signs. Hereby

$$\begin{aligned} dH(q_i, p_q) &\equiv d\left(\sum_i p_i \dot{q}_i - L\right) \\ &= - \sum_i \dot{p}_i dq_i + \sum_i \dot{q}_i dp_i. \end{aligned} \quad (3.31)$$

⁵Since a point in phase space is sufficient to determine the future evolution of the system, paths in phase space can never cross. We say that evolution is governed by a flow in phase space. The movement in phase space can therefore be modeled as the flow of an ideal (incompressible) fluid.

To interpret the Hamiltonian; rewrite the Hamiltonian by identifying the sum as $\sum_{j=1}^n (p_j \dot{q}_j) = 2T$ and using that $L \equiv T - V$, the Hamiltonian can be written as

$$H = T + V. \quad (3.32)$$

Hence, the Hamiltonian can be interpreted as the total energy of the system. The partial derivatives of the Hamiltonian are the Hamilton equations

$$\frac{\partial H}{\partial p_i} = \dot{q}_i, \quad \frac{\partial H}{\partial q_i} = -\dot{p}_i. \quad (3.33)$$

Hamilton's equations are equivalent to the Euler-Lagrange equation. They comprise a set of $2i$ first order PDE's - as opposed to the Euler-Lagrange equation which make up i second order PDE's. From the Hamiltonian the principle of conservation of energy of a closed system can be understood by taking the derivative w.r.t. time of a generic Hamiltonian, $H = H(q_i, p_i, t)$

$$\begin{aligned} \frac{dH}{dt} &= \frac{\partial H}{\partial t} + \sum_i \frac{\partial H}{\partial q_i} \dot{q}_i + \sum_i \frac{\partial H}{\partial p_i} \dot{p}_i \\ &= \frac{\partial H}{\partial t}, \end{aligned} \quad (3.34)$$

where Hamilton's equations have been used to simplify. From equation (3.34) it is clear that if H does not explicitly depend on time, i.e. if the system is not open, the energy of the Hamiltonian (the energy) is conserved.

The Poisson bracket

The Hamilton equations can be used to define a quantity called the Poisson bracket; consider the time derivative of a function, $f = f(q_1, q_2, \dots, q_n, p_1, p_2, \dots, p_n, t)$

$$\begin{aligned} \frac{df}{dt} &= \sum_{j=1}^n \left(\frac{\partial f}{\partial q_j} \dot{q}_j + \frac{\partial f}{\partial p_j} \dot{p}_j \right) + \frac{\partial f}{\partial t} \\ &= \sum_{j=1}^n \left(\frac{\partial f}{\partial q_j} \frac{\partial H}{\partial p_j} - \frac{\partial f}{\partial p_j} \frac{\partial H}{\partial q_j} \right) + \frac{\partial f}{\partial t} \\ &\equiv \{f, H\} + \frac{\partial f}{\partial t}. \end{aligned} \quad (3.35)$$

$\{f, H\}$ is defined as the Poisson bracket between f and H . For two generic functions, the Poisson bracket is given by

$$\begin{aligned} \{f, g\} &= \sum_{j=1}^n \left(\frac{\partial f}{\partial q_j} \frac{\partial g}{\partial p_j} - \frac{\partial f}{\partial p_j} \frac{\partial g}{\partial q_j} \right) \\ &= \frac{\partial f}{\partial q_1} \frac{\partial g}{\partial p_1} - \frac{\partial f}{\partial p_1} \frac{\partial g}{\partial q_1} + \dots \frac{\partial f}{\partial q_n} \frac{\partial g}{\partial p_n} - \frac{\partial f}{\partial p_n} \frac{\partial g}{\partial q_n}. \end{aligned} \quad (3.36)$$

The Poisson bracket has the properties

$$\begin{aligned} \text{Anti-symmetric: } &\Rightarrow \{f, H\} = -\{H, f\}, \\ \text{Linear in arguments: } &\Rightarrow \{f_1 + f_2, H\} = \{f_1, H\} + \{f_2, H\}. \end{aligned} \quad (3.37)$$

An important usage of the Poisson bracket is in relation to the search for constants of motion. A constant of motion is a function of phase space, independent of time, whose value is constant for any particle. Alternatively put; the time derivative, equation $\frac{df}{dt}$, vanishes for a constant of the motion. Considering the case of a

function, $f(q_i, p_i)$, which does not explicitly depend on time, it follows that $f(q_i, p_i)$ is a constant of motion if the Poisson bracket vanish for all points in phase space. That is

$$\{f(q_i, p_i), H\} = 0 \Rightarrow f(q_i, p_i) = \text{constant of motion.} \quad (3.38)$$

Constants of motion play a prominent role in both classical and quantum mechanics, so this tool to find these functions is important⁶. All the familiar conservation laws can be derived from this knowledge.

Example 3.5.

The conservation of energy can be derived by considering the Poisson bracket of Hamiltonian. Using $f = H$ in equation (3.35)

$$\frac{dH}{dt} = \{H, H\} + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t}, \quad (3.39)$$

where $\{f, f\} = 0$ has been used. Equation (3.39) is identical to equation (3.34) and illustrates the same thing - just derived from a different principle. The derivative, in equation (3.39), on the right hand side is taken at a fixed point in phase space while the one on the left hand side is taken by following a particle along as it moves through phase space. The fact that these two derivatives give the same result is quite remarkable. Equation (3.39) states that energy is conserved when the Hamiltonian is time-independent.

Example 3.6.

The conservation of linear momentum can be derived directly from Hamiltons equations as follows

$$\frac{dp_k}{dt} = \{p_k, H\} + \frac{\partial p_k}{\partial t} = -\frac{\partial H}{\partial q_k}. \quad (3.40)$$

From equation (3.40) it is clear that in the case where the Hamiltonian does not depend on the coordinate conjugate to a particular momentum, that momentum is a constant of motion. This is equivalent to what was found in example 2.1.

3.3 Hamilton-Jacobi Mechanics

Consider the Hamiltonian of a generic system, $H(q_i, p_i, t)$. This Hamiltonian can be transformed to another set of coordinate and momenta variables, eg. $H(q_i, p_i, t) \Rightarrow \tilde{H}(\tilde{q}_i, \tilde{p}_i, t)$. Such a transformation will, in general, not conserve the equations of motion. However, a subset of transformations will - This subset of transformations is called *canonical transformations*.

Canonical transformation: *A canonical transformation is a change of canonical coordinates (coordinates which locate the the system in phase space) that preserves the equations of motion (equivalently preserves the form of Hamiltons equations).*

To distinguish a canonical transformation from a generic transformation of the coordinate and momentum variables, such coordinate transformations are denoted viz⁷

$$H(q_i, p_i, t) \Rightarrow K(Q_i, P_i, t), \quad (3.41)$$

where $Q_i = Q_i(q_i, p_i, t)$ and $P_i = P_i(q_i, p_i, t)$ in general. Since the equations of motion are (defined to be)

⁶Think about the importance of conservation of linear momentum, angular momentum, energy and so on.

⁷The Poisson bracket can be used to test whether or not a set of variables are canonical. If they are canonical $\{q, p\} = 1$.

invariant under a canonical transformation the variation of the action must vanish in both cases. Therefore

$$\begin{aligned}
\delta S &= \delta \int dt L \\
&= \int dt \delta \left(\sum_i p_i \dot{q}_i - H(q_i, p_i, t) \right) \\
&= \int dt \delta \left(\sum_i P_i \dot{Q}_i - K(Q_i, P_i, t) \right) \\
&= \delta S' = 0.
\end{aligned} \tag{3.42}$$

Recall that the Lagrangian is only defined up to a time derivative of a generic function of the coordinates and time (see example 3.4). From equation (3.42) it is clear that the integrands must be equal up to such a derivative, i.e.

$$\sum_i p_i \dot{q}_i - H(q_i, p_i, t) = \sum_i P_i \dot{Q}_i - K(Q_i, P_i, t) + \frac{dF(q_i, Q_i, t)}{dt}. \tag{3.43}$$

The function F is called the generating function of the canonical transformation. It is convenient to Legendre transform F to another generating function that is a function of q_i, P_i, t rather than q_i, Q_i, t . From equation (3.43) the differential of F is given by

$$\begin{aligned}
dF(q_i, Q_i, t) &= \sum_i \frac{\partial F(q_i, Q_i, t)}{\partial q_i} dq_i + \sum_i \frac{\partial F(q_i, Q_i, t)}{\partial Q_i} dQ_i + \frac{\partial F(q_i, Q_i, t)}{\partial t} dt \\
&= \sum_i p_i dq_i - \sum_i P_i dQ_i + [K(Q_i, P_i, t) - H(q_i, p_i, t)] dt \\
&= \sum_i p_i dq_i - d \left(\sum_i P_i Q_i \right) + \sum_i Q_i dP_i + [K(Q_i, P_i, t) - H(q_i, p_i, t)] dt.
\end{aligned} \tag{3.44}$$

So, denoting the transformed potential $F_2(q_i, P_i, t)$ reveals

$$\begin{aligned}
dF_2(q_i, P_i, t) &= d(F(q_i, Q_i, t) + \sum_i P_i Q_i) \\
&= \sum_i p_i dq_i + \sum_i Q_i dP_i + [K(Q_i, P_i, t) - H(q_i, p_i, t)] dt.
\end{aligned} \tag{3.45}$$

From which

$$\frac{\partial F_2}{\partial q_j} = p_j, \quad \frac{\partial F_2}{\partial P_j} = Q_j, \quad \frac{\partial F_2}{\partial t} = K - H. \tag{3.46}$$

Example 3.7. —

Consider the canonical transformation defined by the generating function $F_2(q_i, P_i, t) = \sum_i q_i P_i$. What does this transformation say about the transformed Hamiltonian (K)?

Using equation (3.46) it is clear that

$$K(Q_i, P_i, t) = H(q_i, p_i, t) + \frac{\partial F_2(q_i, P_i, t)}{\partial t} = H(q_i, p_i, t), \tag{3.47}$$

where $\frac{\partial F_2(q_i, P_i, t)}{\partial t} = 0$ since $F_2 \neq F(t)$ in this case. Hence, this canonical transformation is the identity transformation.

Using the result from example 3.7 (that $F_2(q_i, P_i, t) = \sum_i q_i P_i$ is the identity transformation) an infinitesimal canonical transformation can be parametrized viz

$$F_2(q_i, P_i, t) = \sum_i q_i P_i + \epsilon G(q_i, P_i, t), \tag{3.48}$$

where ϵ is an infinitesimal parameter and $G(q_i, P_i, t)$ is the generator of the infinitesimal canonical transformation. Equation (3.48) implies that

$$\begin{aligned} p_j &= \frac{\partial F_2}{\partial q_j} = P_j + \epsilon \frac{\partial G}{\partial q_j} \Rightarrow \delta p_j = P_j - p_j = -\epsilon \frac{\partial G}{\partial q_i}, \\ Q_j &= \frac{\partial F_2}{\partial P_j} = q_j + \epsilon \frac{\partial G}{\partial P_j} \Rightarrow \delta q_j = Q_j - q_j = \epsilon \frac{\partial G}{\partial P_j}. \end{aligned} \quad (3.49)$$

Since $P_j = p_j + \mathcal{O}(\epsilon)$; $\frac{\partial G}{\partial P_j} = \frac{\partial G}{\partial p_j} + \mathcal{O}(\epsilon)$ (since $\frac{1}{p_j + \mathcal{O}(\epsilon)} = p_j - \mathcal{O}(\epsilon)$). Hereby

$$\begin{aligned} \delta p_j &= P_j - p_j = -\epsilon \frac{\partial G}{\partial q_i}, \\ \delta q_j &= Q_j - q_j = \epsilon \frac{\partial G}{\partial p_j}. \end{aligned} \quad (3.50)$$

Example 3.8.

Consider the canonical transformation defined by

$$Q_j = q_j(t + \delta t), \quad P_j = p_j(t + \delta t). \quad (3.51)$$

What is the function generating the canonical transformation in the time coordinate?

Take δt to be infinitesimal and Taylor expand Q_j, p_j viz

$$\begin{aligned} Q_j &= q_j(t) + \dot{q}_j \delta t = q_j(t) + \frac{\partial H}{\partial p_j} \delta t, \\ P_j &= p_j(t) + \dot{p}_j \delta t = p_j(t) - \frac{\partial H}{\partial q_j} \delta t, \end{aligned} \quad (3.52)$$

where Hamiltons equations (equation (3.33)) have been applied. By comparing equation (3.52) and (3.50) it is clear that $H \Leftrightarrow G$ and as such the Hamiltonian is the generator of time translations.

Example 3.9.

Consider a generic function of the phase space coordinates, $O(q_i, p_i, t)$. Write the change in the function, δO , under the simultaneous, infinitesimal canonical transformations; $q_j \rightarrow Q_j = q_j + \delta q_j$ and $p_j \rightarrow P_j = p_j + \delta p_j$.

$$\begin{aligned} \delta O(q_i, p_i, t) &= O(Q_i, P_i, t) - O(q_i, p_i, t) \\ &= \sum_j \frac{\partial O}{\partial q_j} \delta q_j + \sum_j \frac{\partial O}{\partial p_j} \delta p_j \\ &= \epsilon \sum_j \left(\frac{\partial O}{\partial q_j} \frac{\partial G}{\partial p_j} - \frac{\partial O}{\partial p_j} \frac{\partial G}{\partial q_i} \right) + \mathcal{O}(\epsilon^2) \\ &= \epsilon \{O, G\} + \mathcal{O}(\epsilon^2). \end{aligned} \quad (3.53)$$

So, the Poisson bracket denotes the change in a phase space function, O , induced by the generator, G .

Hamilton-Jacobi mechanics revolves around the Hamilton-Jacobi equation. The Hamilton-Jacobi equation is obtained from considering the specific canonical transformation for which $K = 0$. Using $K = 0$ in equation (3.46) reveals the Hamilton-Jacobi equation

$$H(q_i, p_i, t) + \frac{\partial F_2(q_i, P_i, t)}{\partial t} = 0. \quad (3.54)$$

From Hamiltons equations, which must also be valid for $K = 0$ since it is a canonical transformation

$$\frac{\partial K}{\partial P_j} = \dot{Q}_j = 0, \quad -\frac{\partial K}{\partial Q_j} = \dot{P}_j = 0. \quad (3.55)$$

Example 3.10.

Equation (3.55) suggests that in the transformed coordinates there is no dynamics, i.e. the system is static. How should this be interpreted? Is the system still dynamical?

For a dynamical system the original coordinates will depend on time. The new coordinates does not need to depend on time if they are such that the time-dependence of the old coordinates are transformed away. The trivial example is that if $q = at$, $Q = q - at = 0$ would be a constant coordinate. From this example it is clear that the system can be dynamical and the transformed coordinates constant at the same time.

Equation (3.55) can be understood as follows; consider a scenario in which the trajectories of a generic system in q_i, p_i, t -space is plotted. The coordinates Q_i, P_i are "tags" for each trajectory in q_i, p_i, t -space. Hence, in this space each line trajectory will have the same values of Q_i, P_i for different values of q_i, p_i, t . Hence, Q_i, P_i are to be understood as constants of the motion. This does not necessarily mean that momentum and position is conserved in these new coordinates. Consider a system for which energy is conserved. Divide this conserved energy with the appropriate natural constants, and a constant of motion with arbitrary units will appear. Hence, in this way the "usual" conserved quantities can be disguised as position and momenta.

Example 3.11.

Consider the Hamiltonian of the harmonic oscillator

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2q^2 = \frac{p^2}{2m} + \frac{1}{2}kq^2. \quad (3.56)$$

Determine $q(t)$, $p(t)$ via Hamilton-Jacobi mechanics.

From equation (3.54)

$$\frac{p^2}{2m} + \frac{1}{2}kq^2 + \frac{\partial F_2}{\partial t} = 0. \quad (3.57)$$

Since $\frac{\partial H}{\partial t} = 0 \Rightarrow H = E = \text{const} \Rightarrow \frac{\partial F_2}{\partial t} = -E \Rightarrow F_2 = F_2(t=0) - Et$

$$\frac{p^2}{2m} + \frac{1}{2}kq^2 - E = 0. \quad (3.58)$$

Next use that, from equation (3.46), $p = \frac{\partial F_2}{\partial q}$. Hereby

$$\frac{1}{2m} \left(\frac{\partial F_2(t=0)}{\partial q} \right)^2 + \frac{1}{2}kq^2 - E = 0. \quad (3.59)$$

From separation of variables

$$dF_2(t=0) = \sqrt{2mE - mkq^2} dq \Rightarrow F_2(t=0) = \sqrt{mk} \int \sqrt{\frac{2E}{k} - q^2} dq + C_0. \quad (3.60)$$

Hence

$$F_2 = \sqrt{mk} \int \sqrt{\frac{2E}{k} - q^2} dq + C_0 - Et. \quad (3.61)$$

The generalized momentum is determined by using equation (3.46) as follows

$$\begin{aligned} p(t) &= \frac{\partial F_2}{\partial q} \\ &= \frac{\partial}{\partial q} \left(\sqrt{mk} \int \sqrt{\frac{2E}{k} - q^2} dq + C_0 - Et \right) \\ &= \sqrt{mk} \sqrt{\frac{2E}{k} - q^2}. \end{aligned} \quad (3.62)$$

The generalized momentum, P , does not immediately appear in F_2 , so the usage of equation (3.46) to determine $q(t)$ is not clear. If P did appear equation (3.46) could be used to find $Q = Q(q, p, t)$ which in turn could be flipped to find $q(t)$. Instead recall that the conserved canonical coordinates represents the constants of the motion. The energy is in this case a constant of motion, so rescale P by the appropriate constant prefactor to make it the energy of the system. Since the prefactor is undefined it is unclear what the result of this will be, however, from equation (3.46), it is clear that by letting $P \Rightarrow E$; $\frac{\partial F_2}{\partial E} = \text{constant of unit time}$. Hereby

$$\frac{\partial F_2}{\partial E} = \sqrt{mk} \int \frac{1}{2\sqrt{\frac{2E}{k}} - q^2} \frac{2}{k} dq - t = \text{const of unit time} = t_0. \quad (3.63)$$

From this

$$t = \sqrt{\frac{m}{k}} \int \frac{1}{2\sqrt{\frac{2E}{k}} - q^2} dq - t_0. \quad (3.64)$$

Evaluate the integral by using $\int \frac{1}{\sqrt{a^2 - x^2}} dx = a \sin(\frac{x}{a})$. Hereby

$$t = \sqrt{\frac{m}{k}} a \sin\left(\sqrt{\frac{k}{2E}} q\right) - t_0. \quad (3.65)$$

Isolating $q(t)$ from t

$$q(t) = \sqrt{\frac{2E}{k}} \sin\left(\sqrt{\frac{k}{m}}(t + t_0)\right). \quad (3.66)$$

Example 3.12.

Consider the Lagrangian

$$L = \frac{1}{2} m \dot{q}^2. \quad (3.67)$$

- Determine $q(t), p(t)$ from Lagrangian mechanics. The Lagrangian is cyclic in q , so the corresponding momentum, p , is conserved. From the Euler-Lagrange equation

$$\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} = 0 \Rightarrow \frac{d}{dt} \left(m \dot{q} \right) \Rightarrow m \ddot{q} \equiv p = \text{const.} \quad (3.68)$$

The differential equation ($\ddot{q} = \frac{p}{m}$) can be solved by separation of variables to find

$$q(t) = \frac{p}{m} t + q_0. \quad (3.69)$$

- Determine $q(t), p(t)$ from Hamiltonian mechanics. Use that $H = L = \frac{p^2}{2m}$ since $V(q) = 0$ in this case. So, from Hamiltons equations

$$\begin{aligned} \frac{\partial H}{\partial q} &= -\dot{p} = 0 \Rightarrow p = \text{const} \\ \frac{\partial H}{\partial p} &= \dot{q} = \frac{p}{m} \Rightarrow q(t) = \frac{p}{m} t + q_0. \end{aligned} \quad (3.70)$$

- Determine $q(t), p(t)$ from Hamiltonian-Jacobi mechanics. Use that $\frac{\partial H}{\partial t} = 0 \Rightarrow H = E = \text{const} = \frac{p^2}{2m} \Rightarrow p = \sqrt{2mE} = \text{const}$. To determine $q(t)$ use that E can be viewed as the rescaled momentum constant, such that

$$\frac{\partial F_2}{\partial E} = \tilde{t}_0. \quad (3.71)$$

F_2 is determined by using that

$$H + \frac{\partial F_2}{\partial t} = E + \frac{\partial F_2}{\partial t} = 0 \Rightarrow F_2(t) = F_2(0) - Et. \quad (3.72)$$

$F_2(0)$ is determined from

$$\frac{\partial F_2}{\partial q} = \frac{\partial F_2(0)}{\partial q} = p = \sqrt{2mE} \Rightarrow F_2(0) = \sqrt{2mE}q + C. \quad (3.73)$$

So

$$\frac{\partial F_2}{\partial E} = \tilde{t}_0 = \frac{1}{2} \frac{2m}{\sqrt{2mE}} q - t \Rightarrow q(t) = \frac{p}{m}t + q_0. \quad (3.74)$$

3.3.1 Identifying F_2 with the action

The generating function F_2 can be identified with a well known quantity. The relation can be realized by considering

$$\frac{dF_2}{dt} = \sum_i \frac{\partial F_2}{\partial q_i} \dot{q}_i + \sum_i \frac{\partial F_2}{\partial P_i} \dot{P}_i + \frac{\partial F_2}{\partial t}. \quad (3.75)$$

Taking the canonical transformation for which $K = 0 \Leftrightarrow \dot{P}_i = 0$ as well as $\frac{\partial F_2}{\partial q_j} = p_j$ and $\frac{\partial F_2}{\partial t} = -H$

$$\frac{dF_2}{dt} = \sum_i p_i \dot{q}_i - H = L \Rightarrow F_2 = \int dt L + const. \quad (3.76)$$

From equation (3.76) it is clear that F_2 corresponds to the action. However, this correspondence must be followed by an asterix since F_2 is a function of time whereas the action is not. In order to identify F_2 with the action one parameterizes the action as a function of the endpoint of the path (recall that $S = S[q_i(t_i), q_f(t_f), t_i, t_f]$). Locking the initial point one can view the action as a function of the endpoint. Hereby the action can be expressed viz

$$S[q_i(t_0), q_f(t), t_0, t] = \int_{t_0}^t dt' L(q_i(t'), \dot{q}_i(t'), t') = \int dt L + S(t_0). \quad (3.77)$$

In the second equality it is used that the upper limit of the integral corresponds to evaluating the integral as an indefinite integral and placing the lower limit in a separate term. From this parametrization is is clear how the action and F_2 are related; $F_2 = S$ if the action is taken as a function of the final point with locked initial point. The time dependence of F_2 translates to the dependence of the upper integration limit of the action integral (t_2) whereas the term from the lower integration limit (t_0) is placed in the "const"-term in equation (3.76).

4 | Thermodynamics

Thermodynamics is the branch of physics which considers heat(Q) and temperature(T) and their relation to work(W) an energy (internal energy, U). It was born in the 19th century as scientists were first discovering how to build and operate steam engines. Thermodynamics deals only with the large scale response of a system which can be observed and measured in experiments. In thermodynamics a system is defined to be whatever part of the universe that is subject to study. Around the system are the surroundings of the system(see figure 4.1).

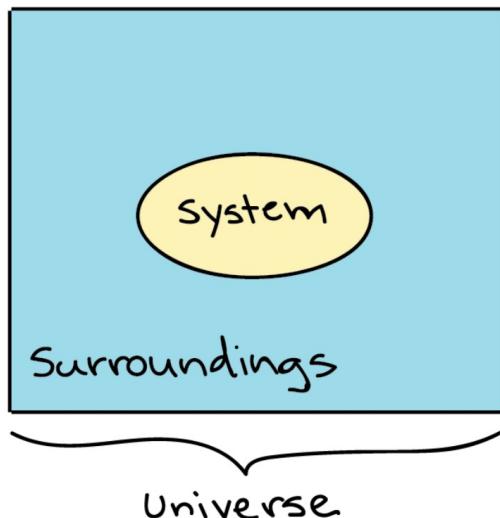


Figure 4.1: The definition of a system and its surroundings in thermodynamics.

The system and the surroundings together make up the universe. In general there are three categorizations of systems in thermodynamics:

1. *Open*: An open system is defined as a system that can exchange both energy and matter with its surroundings.
2. *Closed*: A closed system can exchange neither energy nor matter with its surroundings.
3. *Isolated*: An isolated system can exchange either energy or matter with its surroundings.

The system is said to be in thermodynamic equilibrium if the macroscopic variables characterizing the system (eg. T, P, V, \dots of a gas) is independent of time. Central to thermodynamics is the notion of functions of state, the four laws to thermodynamics and the analysis of system stability.

4.1 Functions of state

As mentioned, a system is in thermodynamic equilibrium if the macroscopic observables are independent of time. The macroscopic observables are functions of state. A function of state is defined as the negative

gradient of some abstract, conservative vector field(just like the case of equation (2.25)). That is, a function of state is defined from the path integral

$$\int_{\vec{x}_i}^{\vec{x}_f} df = f(\vec{x}_f) - f(\vec{x}_i). \quad (4.1)$$

The force field corresponding to the potential function is usually not ascribed any physical interpretation, it is the potential that has physical significance in this case - Since the potential is the physical observable. In general the differential of a function of state is called an exact differential.

Example 4.1.

A conservative vector field has many useful properties, one of which is that the curl of the vector field is 0. This translates to symmetry of the second derivatives of the potential function. Eg. for a potential $U = U(S, V)$

$$\frac{\partial^2 U}{\partial S \partial V} = \frac{\partial^2 U}{\partial V \partial S}, \quad (4.2)$$

where S, V at this point can be considered as nothing more than variables. Later on they can be considered as entropy(S) and volume(V). In thermodynamics, such relations are called Maxwell relations and will prove very useful when the potentials are expressed in terms of state functions.

4.2 The laws of thermodynamics

There are three principal laws of thermodynamics, however, these are supplemented by what is called the zeroth law of thermodynamics to make up four laws in total. Each law leads to the definition of thermodynamic properties which enables one to understand and predict the operation of a physical system.

1. *The zeroth law:* The zeroth law of thermodynamics state that two systems, each separately in thermal equilibrium with a third, are in equilibrium with each other.
2. *The first law:* The first law of thermodynamics state that energy is conserved and heat and work are both forms of energy. The energy under consideration in thermodynamics is the internal energy, U . The internal energy is the sum of all the energy of all the internal degrees of freedom that the system posses. It is a function of state since the energy of a system is independent of how the energy is acquired. The internal energy of a system can be changed by doing work on the system or by being heated. The heat, Q , and work, W , are not functions of state since they describe how energy is delivered to the system. The equation describing the differential change in internal energy is the equation associated to the first law of thermodynamics

$$dU = dQ + dW, \quad (4.3)$$

where the bar-notation denotes inexact differentials. $\bar{d}Q > 0$ is the heat applied to the system and $\bar{d}W > 0$ is the work done on the system. $dQ > 0$.

3. *The second law:* The second law of thermodynamics can be formulated as a statement about the direction of heat flow that occurs as the system approaches equilibrium from an out of equilibrium state; no process is possible whose sole result is the transfer of heat from a colder to a hotter body. The most common manifestation of the second law of thermodynamics is in terms of entropy. Entropy is again defined in terms of the heat change under a reversible process. A reversible process is a process in which the system and surroundings can be restored to the initial state from the final state without producing any changes in the thermodynamic variables of the universe. In practice most processes are reversible¹ if they are made sufficiently slow (quasi-static) such that the system passes through a series of equilibrium states between two macro-states. Any reversible process is quasi-static, but not

¹The macroscopic variables(eg. T, V,P) are only defined in equilibrium, so a quasi-static process has the macroscopic variables defined all the time - Unlike a process which is not quasi-static by fast.

all quasi-static processes are reversible. For example, a quasi-static process involving friction in the mechanics will not be reversible. In general are processes which involve a heat transfer between two systems not reversible. The change of entropy (dS) is defined in terms of the reversible change of heat dQ_{rev} as follows

$$dS = \frac{dQ_{rev}}{T}. \quad (4.4)$$

The entropy is a state function so the differential is exact. Therefore, taking the endpoint to equal the initial point

$$\oint dS = \oint \frac{dQ_{rev}}{T} = 0. \quad (4.5)$$

From Clausius theorem, for a process containing irreversible parts (Blundell et al., 2006)

$$\oint \frac{dQ}{T} \leq 0. \quad (4.6)$$

Taking $a \Rightarrow b$ to be done irreversibly and $b \Rightarrow a$ to be done reversible

$$\int_a^b \frac{dQ}{T} + \int_b^a \frac{dQ_{rev}}{T} \leq 0 \Rightarrow \int_a^b \frac{dQ}{T} \leq \int_a^b \frac{dQ_{rev}}{T}. \quad (4.7)$$

Equation (4.7) holds true however close a, b are as points in state space. So, in general

$$dS = \frac{dQ_{rev}}{T} \geq \frac{dQ}{T}. \quad (4.8)$$

The equality in \geq only holds if the process is reversible. For an isolated system $dQ = 0$, so in general

$$dS \geq 0. \quad (4.9)$$

Equation (4.9) states that any change for a thermally isolated system results in the entropy staying the same (for a reversible process) or increasing (for an irreversible process). This leads to another statement of the second law of thermodynamics; the entropy of an isolated system tends to a maximum. Equation (4.9) is the manifestation of this version of the second law and is often used in analyzing the stability of thermodynamic systems. This version of the second law is also often applied to the universe. Taking the entire universe to be a closed system, it is clear that the entropy of the universe can only stay the same or go up². Taking $dQ_{rev} = TdS$ and $dW = -pdV + \mu dN + \dots$, where p is pressure, V is volume, μ is chemical potential and N is the number of particles, the first law can be written viz

$$dU = TdS - pdV + \mu dN + \dots |_{rev}. \quad (4.10)$$

Equation (4.10) is valid for reversible processes only at first glance. However, since all quantities in equation (4.10) are functions of state (exact differentials), and therefore independent of the path in state space, the equation holds for irreversible processes as well. In irreversible processes $dQ < 0$ and $dW > 0$ such that dU is the same whether the change is reversible or not. Hence, in general

$$dU = TdS - pdV + \mu dN + \dots \quad (4.11)$$

4. *The third law:* The third law of thermodynamics state that the entropy of all systems in internal equilibrium is the same at $T = 0K$ (at absolute zero Kelvin), and may be taken to be zero, i.e.; $S(T = 0K) = 0$.

²This description is on average. In analyzing the movement of individual molecules the entropy can indeed be found to go down.

4.3 Manipulation of thermodynamic potentials

A central part of practical application of thermodynamics is consist of manipulation of the thermodynamic potentials at hand. From the first law of thermodynamics

$$dU = TdS - pdV + \mu dN + \dots \quad (4.12)$$

Note that $U = U(S, V, N, \dots)$. The potential U can be transformed to another potential function, which is a function of different state functions, by performing a Legendre transformation(just as in advanced classical mechanics). The potential $U = U(S, V, N, \dots)$ can be transformed to the potential $F(T, V, N, \dots)$ viz

$$\begin{aligned} dU &= TdS - pdV + \mu dN \dots \\ &= d(TS) - SdT - pdV + \mu dN \dots \end{aligned} \quad (4.13)$$

Taking $d(TS)$ to the left hand side of equation (4.13) reveals

$$\begin{aligned} dF(T, V, N, \dots) &= d(U(S, V, N, \dots) - TS) \\ &= -SdT - pdV + \mu dN \dots \end{aligned} \quad (4.14)$$

From the thermodynamic potentials, the equations of state are defied as the partial derivatives. From equation (4.13)

$$T = \frac{\partial U}{\partial S} \Big|_{V, N, \dots}, \quad -p = \frac{\partial U}{\partial V} \Big|_{S, N, \dots}, \quad \mu = \frac{\partial U}{\partial N} \Big|_{S, V, \dots}, \quad \dots \quad (4.15)$$

Similar equations of state can be derived from F . The equation of equation (4.15) are called equations of state. Equations of state are the first partial derivatives of thermodynamic potentials. Now, from the Maxwell relations

$$\begin{aligned} \frac{\partial^2 U}{\partial S \partial V} &= \frac{\partial^2 U}{\partial V \partial S} \Rightarrow -\frac{\partial p}{\partial S} \Big|_{N, \dots} &= \frac{\partial T}{\partial V} \Big|_{N, \dots}, \\ \frac{\partial^2 U}{\partial N \partial V} &= \frac{\partial^2 U}{\partial V \partial N} \Rightarrow -\frac{\partial p}{\partial N} \Big|_{S, \dots} &= \frac{\partial \mu}{\partial V} \Big|_{S, \dots}, \\ \frac{\partial^2 U}{\partial N \partial S} &= \frac{\partial^2 U}{\partial S \partial N} \Rightarrow \frac{\partial T}{\partial N} \Big|_{V, \dots} &= \frac{\partial \mu}{\partial S} \Big|_{V, \dots}. \end{aligned} \quad (4.16)$$

The Maxwell relations are often supplemented by mathematical tools such as Eulers triangle rule and the reciprocal theorem. Eulers triangle rule

$$\text{Eulers triangle rule: } \frac{\partial S}{\partial T} \Big|_{V, N, \dots} \frac{\partial T}{\partial V} \Big|_{S, N, \dots} \frac{\partial V}{\partial S} \Big|_{T, N, \dots} = -1. \quad (4.17)$$

Note how the derivatives are cyclic, and the relation can be composed from just **one** partial derivative. The reciprocal theorem

$$\text{Reciprocal theorem: } \Rightarrow \frac{\partial S}{\partial V} \Big|_{T, N, \dots} = \frac{1}{\frac{\partial V}{\partial S} \Big|_{T, N, \dots}}. \quad (4.18)$$

Example 4.2.

Eulers triangle rule and the reciprocal theorem are often used in conjunction to rewrite state functions viz

$$\frac{\partial S}{\partial T} \Big|_{V, N, \dots} = -\frac{\partial S}{\partial V} \Big|_{T, N, \dots} \frac{\partial V}{\partial T} \Big|_{S, N, \dots}. \quad (4.19)$$

In this way, a possibly non-calculable differential can be turned into a perhaps calculate-able one. These relations can in turn be combined with the equations of state and the Maxwell relations in order to really make the magic happen.

4.4 System stability

In thermodynamics the thermodynamic potentials are often used to analyze the stability of the system. The condition for stability is that for fixed variables, any change in the thermodynamic potential system must be positive. For example for $F = F(T, V, N)$ when T, V, N are fixed, $\delta F \geq 0$. Another way of stating this is that if you leave a system alone, it is stable if it is in a minimum of the potential. In the minimum, every change in the potential is positive! In equilibrium, the potential must be convex. A convex function has a positive curvature. This is mathematically expressed as

$$d^2F > 0 \quad (4.20)$$

To examine the stability, investigate the fluctuations between two subsystems, when the system is in a stable equilibrium. Take $N = \text{const}$, while allowing T, V to vary. A small perturbation in the potential can then be written as follows

$$\delta F = F_2(T + \Delta T, V + \Delta V, N) + F_1(T - \Delta T, V - \Delta V, N) - 2F(T, M, N) \geq 0, \quad (4.21)$$

where the last term is included to subtract the average offset of the potential. Only the fluctuations around the minimum is of interest, not with respect to the arbitrary reference value that has been chosen. Next, Taylor expand F_1 and F_2

$$\begin{aligned} F_2 &= F(T, V, N) + \frac{\partial F}{\partial T} \Big|_{V,N} \Delta T + \frac{\partial F}{\partial V} \Big|_{T,N} \Delta V \\ &\quad + \frac{1}{2} \left(\frac{\partial^2 F}{\partial T^2} \Big|_{V,N} (\Delta T)^2 + \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} (\Delta V)^2 + 2 \frac{\partial^2 F}{\partial T \partial V} \Big|_N \Delta M \Delta T \right) + \dots, \\ F_1 &= F(T, V, N) - \frac{\partial F}{\partial T} \Big|_{V,N} \Delta T - \frac{\partial F}{\partial V} \Big|_{T,N} \Delta V \\ &\quad + \frac{1}{2} \left(\frac{\partial^2 F}{\partial T^2} \Big|_{V,N} (\Delta T)^2 + \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} (\Delta V)^2 + 2 \frac{\partial^2 F}{\partial T \partial V} \Big|_N \Delta V \Delta T \right) + \dots, \end{aligned} \quad (4.22)$$

where the terms of higher order is neglected in the usual analysis of stability. By inserting equation (4.22) in equation (4.21), the first order terms, along with the zero order terms, disappear and leave only the second order terms. Let the changes go to infinitesimal sizes ($\Delta \Rightarrow d$) to find

$$\delta F \cong \frac{\partial^2 F}{\partial T^2} \Big|_{V,N} (dT)^2 + \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} (dV)^2 + 2 \frac{\partial^2 F}{\partial T \partial V} \Big|_N dV dT = d^2F > 0. \quad (4.23)$$

The " \cong " in equation (4.23) come from neglecting higher order terms from the Taylor series. Equation (4.23) can be written in matrix form

$$d^2F = [dT \quad dV] \begin{bmatrix} \frac{\partial^2 F}{\partial T^2} \Big|_{V,N} & \frac{\partial^2 F}{\partial T \partial V} \Big|_N \\ \frac{\partial^2 F}{\partial T \partial V} \Big|_N & \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} \end{bmatrix} \begin{bmatrix} dT \\ dV \end{bmatrix} > 0. \quad (4.24)$$

Since the matrix is symmetric, it can be diagonalized with the eigenvalues in the diagonal. That is, equation (4.24) can be written on the form

$$d^2F = [dT \quad dV] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} dT \\ dV \end{bmatrix} = \lambda_1(dT)^2 + \lambda_2(dV)^2 > 0. \quad (4.25)$$

For this expression to hold for every possible eigenvalue, the eigenvalues must be positive. Since they must be positive, the determinant and trace of the diagonalized matrix must be positive as well. That is

$$\begin{aligned} \det \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} &= \lambda_1 \lambda_2 > 0, \\ \text{tr} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} &= \lambda_1 + \lambda_2 > 0. \end{aligned} \quad (4.26)$$

Note that the determinant can be positive if both eigenvalues are negative, and the trace can be positive if the smallest eigenvalue is negative. If both conditions are fulfilled, then both eigenvalues must be positive. A diagonalization is a linear transformation. Under such a transformation both the determinant and trace are conserved quantities. That means the trace and determinant of the diagonalized matrix can be determined directly from the undiagonalized matrix. No diagonalization have to be performed in practice. The determinant and trace of the "middle" matrix in equation 4.24, are in this case found to be:

$$\begin{aligned} \det \begin{bmatrix} \frac{\partial^2 F}{\partial T^2} \Big|_{V,N} & \frac{\partial^2 F}{\partial T \partial V} \Big|_N \\ \frac{\partial^2 F}{\partial T \partial V} \Big|_N & \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} \end{bmatrix} &= \frac{\partial^2 F}{\partial T^2} \Big|_{V,N} \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} - \left(\frac{\partial^2 F}{\partial T \partial V} \Big|_N \right)^2 > 0 \\ \text{tr} \begin{bmatrix} \frac{\partial^2 F}{\partial T^2} \Big|_{V,N} & \frac{\partial^2 F}{\partial T \partial V} \Big|_N \\ \frac{\partial^2 F}{\partial T \partial V} \Big|_N & \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} \end{bmatrix} &= \frac{\partial^2 F}{\partial T^2} \Big|_{V,N} + \frac{\partial^2 F}{\partial V^2} \Big|_{T,N} > 0 \end{aligned} \quad (4.27)$$

After this expression is reached, the expressions for the derivatives must be analyzed in order to interpret the physical content of the stability of the system³. The second derivatives couple to the heat capacities, so the stability of the system couples to the allowed signs of the heat capacities⁴. Further more, the general derivatives can be analyzed by determining whether a thermodynamic variable is extensive or intensive. Thermodynamic potentials are convex functions of their extensive variables, and concave functions of their intensive variables. For example the temperature is an intensive variable; it does not change if the system size is altered. From this it can be concluded that $\frac{\partial^2 F}{\partial T^2} \Big|_{M,N} \leq 0$. Likewise for other thermodynamic variables and potentials⁵.

³Note that 4.27 is derived on the basis that a perturbation in V and T is performed whilst N is held constant. If these conditions are not met, the equations does not hold(!).

⁴Often they must be positive in a stable system

⁵That goes for the evaluation of the second derivatives as well as the stability treatment of the potential.

Part II

Quantum Mechanics

5 | Mathematical setup

Quantum mechanics is the science of small systems in which the particles move at non-relativistic speeds. The involved speeds need to be non-relativistic (i.e. relatively low) since if this is not the case, the energies involved will be so large that particles can be created and annihilated. Since conservation of particle number is central in QM it cannot accommodate a variable number of particles without problems. This is the reason for introducing QFT later on.

A quantum mechanical system is defined as a system which is disturbed by the act of (a very careful) measurement¹. In the words of Paul Dirac (Dirac, 1930)

“... We have to assume that there is a limit to the finiteness of our powers of observation and the smallness of the accompanying disturbance - a limit which is inherent in the nature of things and can never be surpassed by improved technique or increased skill on the part of the observer.”

Because measurements disturb the systems in quantum mechanics the QM system is fundamentally different from a system of classical mechanics. In classical mechanics the physical state of a system can be specified by specifying the system particles' positions and momenta (q_i, p_i) at a given time. Because the act of measurement disturbs a quantum mechanical system, the data that can be assigned to a quantum mechanical system depends on whether or not measurements of the different quantities disturb each other. In the words of Paul Dirac (Dirac, 1930, p.11)

“The limitation in the power of observation puts a limitation on the number of data that can be assigned to a state. Thus a state of an atomic system must be specified by fewer or more indeterminate data...”

“A state of a system may be defined as an undisturbed motion that is restricted by as many conditions or data as are theoretically possible without mutual interference or contradiction.”

5.1 The vector space of quantum mechanics

Because of this difference (from classical mechanics) the quantum mechanical state of a system is represented as a vector, called the *state vector*, in a complex vector space of N dimensions. If the possible results of a measurement make up an infinite set, as is the case when measuring position or momentum, the vector space is of infinite dimension and called a *Hilbert space*². I shall use the notation of Paul Dirac, and call a vector in the complex vector space (of any dimension) a *ket*. I will refer to general vectors representing a general physical state as $|\alpha\rangle, |\beta\rangle, |\gamma\rangle, \dots$. Initially I will take the state vector to be independent of time. This provides the opportunity to introduce symmetry operators and derive the Schrödinger equation later on³. For two arbitrary vectors, $|\alpha\rangle$ and $|\beta\rangle$, in the complex vector space the following axioms apply (Sakurai, 1994).

¹There exist several interpretations of quantum mechanics, including The Copenhagen interpretation and Bohmian mechanics (pilot wave theory). The former assume probability waves are the true nature of quantum mechanics and upon measurement the probability wave collapses into a particle. The latter assume there is a real particle that is guided around by real waves of unknown nature. The two interpretations yield the same predictions, but have different interpretations. Only the Copenhagen interpretation, however, have a relativistic version in QFT. This is because QFT requires all possible particle paths to be equally real, which is not the case in Bohmian mechanics where there is a real particle that takes a single path.

²However, physicists often denote the vector space of quantum mechanics a Hilbert space even if the dimensions are finite!

³How to describe the time-dependence of the state vector is, as so many other things in QM, a matter of taste.

1. Two vectors can be added to form a new vector

$$|\alpha\rangle + |\beta\rangle = |\gamma\rangle. \quad (5.1)$$

2. Vector addition is cumulative

$$|\alpha\rangle + |\beta\rangle = |\beta\rangle + |\alpha\rangle. \quad (5.2)$$

3. Vector addition is associative

$$(|\alpha\rangle + |\beta\rangle) + |\gamma\rangle = |\beta\rangle + (|\alpha\rangle + |\gamma\rangle). \quad (5.3)$$

4. There is a unique vector which can be added, to another vector, which results in the first vector(null vector)

$$|0\rangle + |\alpha\rangle = |\alpha\rangle. \quad (5.4)$$

5. Given any ket, there is a unique ket that when added results in zero

$$|\alpha\rangle + (-|\alpha\rangle) = 0. \quad (5.5)$$

6. The distributive property holds (for a complex number c)

$$c(|\alpha\rangle + |\beta\rangle) = c|\alpha\rangle + c|\beta\rangle. \quad (5.6)$$

7. Vectors can be multiplied with a complex number to form a new vector

$$c|\alpha\rangle = |\alpha'\rangle. \quad (5.7)$$

Axioms 6 and 7 together are called linearity. Ordinary 3-vectors would satisfy these axioms, except for one thing: Axiom 7 allows a vector to be multiplied with a complex number, whereas ordinary 3-vectors are only allowed to be multiplied with a real number. If axiom 7 is applied to the state vector, it would change the length of the vector, but not the physical state the vector is referring to - it is only the direction, not the magnitude, of the state vector that is of importance (regarding the physical state of the system).

Just like complex numbers have dual versions in the form of complex conjugate numbers, so have the ket vector space a dual version that is essentially the complex conjugate version. This dual vector space contains vectors called bra's. For each ket there exists a bra and there is a one-to-one correspondence between the vectors in the two vector spaces, i.e. for each ket $|\alpha\rangle$ there exists a bra $\langle\alpha|$. The above axioms containing kets apply for bra's if all kets are replaced with bra's.

Because the act of measurement disturbs the system, measurements in quantum mechanics are represented by operators, which are applied to the state vector. Operators representing a measurable physical quantity are called *observables*, and could be position (x, y, z) , momentum(p), energy(H) and so on⁴. Mathematically a general operator, A , operates on the ket to its right to form a new ket

$$A \cdot (|\alpha\rangle) = A|\alpha\rangle = |\alpha'\rangle, \quad (5.8)$$

where $|\alpha'\rangle$ just denotes that the result of an operator operating on a ket is just another ket. Likewise, an operator operates on a bra from the right side to form a new bra

$$(\langle\alpha|) \cdot A = \langle\alpha| A = \langle\alpha'|. \quad (5.9)$$

The ket $A|\alpha\rangle$ and bra $\langle\alpha|A$ are, in general, not dual to each other. The operator operating on a dual bra to create a bra dual to $A|\alpha\rangle$ is denoted as the *Hermitian adjoint of operator* of an operator. It is defined as

$$A|\alpha\rangle \iff \langle\alpha|A^\dagger \quad (5.10)$$

With the Hermitian adjoint operator comes the following axioms.

⁴A common notation for operators is a hat. However, this notation is absent in more advanced texts and so it is a good exercise to get used to it from the beginning.

1. The Hermitian adjoint of a ket is a bra

$$(|\alpha\rangle)^\dagger = \langle\alpha|. \quad (5.11)$$

2. Considering an operator sandwiched by a ket and bra the operator will initially act on the ket to the right, however by taking the Hermitian adjoint of the operator the operator can be made to act on the bra to the right. That is⁵

$$\langle\beta| (A|\alpha\rangle) = (\langle\beta| A^\dagger)|\alpha\rangle. \quad (5.12)$$

3. The Hermitian adjoint of an operator that is already the Hermitian adjoint of an operator is the operator itself

$$(A^\dagger)^\dagger = A. \quad (5.13)$$

4. The Hermitian adjoint is distributive in the way that

$$(AB)^\dagger = B^\dagger A^\dagger, \quad (5.14)$$

where B is just another operator.

5. The Hermitian adjoint is also distributive in the way that

$$(A + B)^\dagger = A^\dagger + B^\dagger. \quad (5.15)$$

6. The Hermitian adjoint of a number is the complex conjugate of the number

$$c^\dagger = c^*. \quad (5.16)$$

7. Taking the Hermitian adjoint of two operators reverses the direction they act but also the order

$$\langle\beta| AB|\alpha\rangle = (\langle\beta| A^\dagger)B|\alpha\rangle = ((\langle\beta| A^\dagger)B^\dagger)|\alpha\rangle. \quad (5.17)$$

If an operator is equal to the Hermitian adjoint of itself it is said to be *Hermitian*. That is an operator is Hermitian if

$$A = A^\dagger. \quad (5.18)$$

This is the case for all operators corresponding to a measurable physical quantity, i.e. for observables. The reason for this is that Hermitian operators have real eigenvalues, and the eigenvalues are the possible outcomes of a measurement. Since a measurement must be a real number, the operators corresponding to something measurable must be Hermitian. For general operators, A , B and C , the following axioms apply.

1. Addition of operators are commutative

$$A + B = B + A. \quad (5.19)$$

2. Addition of operators are associative

$$A + (B + C) = (A + B) + C. \quad (5.20)$$

3. The majority of operators are linear⁶

$$A(c_\alpha|\alpha\rangle + c_\beta|\beta\rangle) = c_\alpha A|\alpha\rangle + c_\beta A|\beta\rangle. \quad (5.21)$$

⁵The Hermitian adjoint is defined as the operator giving the complex conjugate eigenvalue if operating on a dual eigenvector to the left. If one made the operator operate to the right the outcome would in general not be the complex conjugate eigenvector. So $A^\dagger|a'\rangle \neq a'^*|a'\rangle$.

⁶With the exception of the time-reversal operator, all operators in is linear.

4. Multiplication of operators are in general *not* commutative

$$AB \neq BA. \quad (5.22)$$

5. Multiplication of operators are associative

$$A(BC) = (AB)C = ABC. \quad (5.23)$$

In general when a linear operator acts on a vector, it will change the direction of the vector. However, for a given operator, there will be a set of vectors that will have the same direction, but possibly a different magnitude, after the operator acts upon them - these vectors are called *eigenvectors*, and the constant representing the possible change in length of the eigenvector is called the *eigenvalue*. It is said that the eigenvectors corresponds to an eigenvalue, and therefore these vectors are written in a special way

$$A|a'\rangle = a'|a'\rangle \quad \text{and} \quad \langle a'|A^\dagger = \langle a'|a'^*, \quad (5.24)$$

where the primed notation denote that there can be more, and that the vector and constant belong together in sets. In general when considering a ket that is an eigenvector of some observable I will indicate it with a label (subscript, superscript or primed notation). The eigenvalue is in general a complex number, but for the physically measurable quantities (hermitian operators) the eigenvalues are, as mentioned, what can be the outcome of an experiment, and must be *real*.

With both kets, bras and operators defined, I move on to the inner product. Between the two vector spaces, the ket and bra space, an inner product, which results is a complex number in general, is defined as $\langle \alpha | \beta \rangle$ ⁷. For the inner product applies the following axioms.

1. Inner products are linear

$$\langle \gamma | (|\alpha\rangle + |\beta\rangle) = \langle \gamma | \alpha \rangle + \langle \gamma | \beta \rangle. \quad (5.25)$$

2. Interchanging bras and kets correspond to complex conjugation

$$\langle \alpha | \beta \rangle = \langle \beta | \alpha \rangle^*. \quad (5.26)$$

3. The inner product between corresponding bras and kets is a real, positive number

$$\langle \alpha | \alpha \rangle = \Re \geq 0. \quad (5.27)$$

where the equality in axiom 3 only holds if $|\alpha\rangle$ is a null ket and the part about $\langle \alpha | \alpha \rangle$ being real in axiom 3 follows from axiom 2. Two kets are said to be orthogonal if the inner product between the one as a ket and the other as a bra is zero, i.e.

$$\langle \alpha | \beta \rangle = 0. \quad (5.28)$$

This analog to 3-space, where the inner product between orthogonal vectors, eg. representing two of the axis of 3-space, is zero. From the inner product a normalized ket can be constructed

$$|\tilde{\alpha}\rangle = \left(\frac{1}{\sqrt{\langle \alpha | \alpha \rangle}} \right) |\alpha\rangle. \quad (5.29)$$

The normalized ket has the property

$$\langle \tilde{\alpha} | \tilde{\alpha} \rangle = 1. \quad (5.30)$$

Because $|\alpha\rangle$ and $c|\alpha\rangle$ represent the same physical state, physical states are by convention required to be normalized⁸. In practice when evaluating the inner product the identity operator is used to project the state vector onto a set of basis vectors. The identity operator is defined from the completeness relation (defined

⁷The definition of the inner product is the reason for the names of the ket and bra. When put together to an inner product they form a bracket.

⁸Although obviously an un-normalized state vector corresponds to the same physical.

in equation (5.31) and (5.32)), and state that the sum of outer products (eg. $|a'\rangle\langle a'|$) between kets of a complete, orthogonal set is 1. The completeness relation exists both in a discrete and continuous version

$$\sum_{a'} |a'\rangle\langle a'| = 1, \quad (5.31)$$

$$\int_{-\infty}^{\infty} |x\rangle\langle x| dx = 1, \quad (5.32)$$

where a' is just an example, and it could be another complete, orthogonal, discrete set of vectors as well (eg. $|n\rangle$) and 1 is the identity operator; it has the property of $1|\alpha\rangle = |\alpha\rangle$ for any $|\alpha\rangle$. Equation (5.32) can be used to write the inner product⁹ as

$$\langle\alpha|\alpha\rangle = \int_{-\infty}^{\infty} \langle\alpha|x\rangle\langle x|\alpha\rangle dx, \quad (5.33)$$

where $\langle x|\alpha\rangle$ is a complex function called the wave function¹⁰. The wave function is a central element in QM since it describes the state vector in the often utilized position basis. One finds a complete orthogonal set of kets as the eigenkets of *any* arbitrary observable! It is a fundamental theorem of quantum mechanics that the eigenvectors of Hermitian operators make up a complete, orthogonal ($\langle\alpha|\beta\rangle = \delta_{\alpha\beta}$) set. The complete, orthogonal set of vectors span the vector space in which the operator is defined. Hence, as an example, the complete set from the spin operator span a different vector space than the complete set from the Hamiltonian operator for the harmonic oscillator. This might seem like a mathematical subtlety, however it becomes relevant when the complete, orthogonal set is used to expand a state vector upon.

5.2 Matrix notation

So far I have used the notation of Paul Dirac for the vectors in the complex vector space. However, as is clear from the axioms regarding both operators and kets, the operators and kets obey axioms that are similar to that of linear algebra. This means that operators¹¹, with discrete eigenvalues, obeying the axioms can be regarded as matrices and kets and bras as vectors. In the words of Paul Dirac (Dirac, 1930, p.69)

"We can summarize our results for the case when there is only one ξ and it has discrete eigenvalues as follows

- (i) *Any linear operator is represented by a matrix.*
- (ii) *The unit operator is represented by the unit matrix.*
- (iii) *A real linear operator is represented by a Hermitian matrix.*
- (iv) *ξ and functions of ξ are represented by diagonal matrices.*
- (v) *The matrix representing the product of two linear operators is the product of the matrices representing the two factors."*

In order to develop the matrix notation I consider the operator B . I use the completeness relation twice, such that

$$B = \sum_{a'} \sum_{a''} |a''\rangle\langle a''| B |a'\rangle\langle a'|, \quad (5.34)$$

where $\{|a'\rangle\}$ is a complete orthogonal set, which are in general *not* eigenvectors of B , and $|a''\rangle$ and $|a'\rangle$ are just kets from $\{|a'\rangle\}$. The elements (numbers) $\langle a''|B|a'\rangle$ determine the operator (B) completely. They can be arranged in a matrix to represent B in matrix representation. They are in general calculated by using the completeness relation twice

$$\begin{aligned} \langle a''|B|a'\rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle a''|x'\rangle\langle x'|B|x\rangle\langle x|a'\rangle dx dx' \\ &= \int_{-\infty}^{\infty} \psi_{a'}^*(x)b(x)\psi_{a'}(x)dx, \end{aligned} \quad (5.35)$$

⁹Between to identical state vectors in this case - it could be two arbitrary kets in general.

¹⁰If the variables are defined, the inner product becomes a complex number in general. You can view it as $\langle x=1|\alpha\rangle = \text{number}$ or $\psi(x=1) = \text{number}$ - the two are equivalent.

¹¹Any *linear* operator can be represented by a matrix. Almost all operators are linear, but not all!

where¹² I have used that $B = B(x)$, and thereby the eigenvalues of the operator can be inserted in its place; $B(x)|x\rangle = b(x)|x\rangle$. The eigenvalue of the operator is the "scalar" operator in the position basis. This is analog to the known relation $V(x)|x\rangle = V(x)|x\rangle$. By using $B(x)|x\rangle = b(x)|x\rangle$ I find that $\langle x'|B|x\rangle = b(x)\delta(x' - x)$ - which reduces the double integral to a single integral. To interpret the physical significance of the matrix elements, $\langle i|B|j\rangle$, I cite from Dirac (1930, p.45)

"In classical mechanics an observable always, as we say, 'has a value' for any particular state of the system. What is there in quantum mechanics corresponding to this? If we take any observable ξ and any two states x and y , corresponding to the vectors $|x\rangle$ and $|y\rangle$, then we can form the number $\langle x|\xi|y\rangle$. This number is not very closely analogous to the value which an observable can 'have' in the classical theory, for the reasons, namely, (i) it refers to two states of the system, while the classical value always refers to one, (ii) it is in general not a real number, and (iii) it is not uniquely determined by the observables and the states, since the vectors $|x\rangle$ and $|y\rangle$ contain arbitrary numerical factors. Even if we impose on $|x\rangle$ and $|y\rangle$ the condition that they shall be normalized, there will still be an undetermined factor of modulus unity in $\langle x|\xi|y\rangle$. These three reasons cease to apply, however, if we take the two states to be identical and $|y\rangle$ to be the conjugate imaginary vector to $|x\rangle$. The number that we then get, namely $\langle x|\xi|x\rangle$ is necessarily real, and also it is uniquely determined when $|x\rangle$ is normalized, since we multiply $|x\rangle$ by the numerical factor e^{ic} , c being some real number, we must multiply $|x\rangle$ by e^{-ic} and $\langle x|\xi|x\rangle$ will be unaltered.

(...) We therefore make the general assumption that if the measurement of the observable ξ for the system in the state corresponding to $|x\rangle$ is made a large number of times, the average of all the results obtained will be $\langle x|\xi|x\rangle$, provided $|x\rangle$ is normalized. If $|x\rangle$ is not normalized, as is necessarily the case if the state x is an eigenstate of some observable belonging to an eigenvalue in a range, the assumption becomes that the average result of a measurement of ξ is proportional to $\langle x|\xi|x\rangle$."

To recap; the general element $\langle i|B|j\rangle$ does not hold any special position whilst $\langle i|B|i\rangle$ denotes the expectation value of measuring the operator B . In matrix notation, and in the basis of $|a'\rangle$, the operator, B , can be represented by

$$B \doteq \begin{bmatrix} \langle a'|B|a'\rangle & \langle a'|B|a''\rangle & \dots \\ \langle a''|B|a'\rangle & \langle a''|B|a''\rangle & \dots \\ \vdots & & \ddots \end{bmatrix}, \quad (5.36)$$

where $|a'\rangle$, as states before, in general are not the eigenvectors of B . In the special case where the basis kets are eigenvectors of the operator, equation (5.34) reduces to

$$A = \sum_{a'} \sum_{a''} |a''\rangle \langle a''| A |a'\rangle \langle a'| = \sum_{a'} a' |a'\rangle \langle a'|, \quad (5.37)$$

where I have used that the square matrix $\langle a''| A |a'\rangle$ is diagonal in its eigenbasis. Therefore

$$\langle a''| A |a'\rangle = \langle a'| A |a'\rangle \delta_{a'a''} = a' \delta_{a'a''}. \quad (5.38)$$

Equation (5.37) is used to express an operator when the eigenvalues and eigenvectors are known. Going back to the general case where the basis is not the eigenbasis of the operator; I use the information (that B can be represented as a matrix) to define kets and bras as vectors. I use equation (5.8) with B instead of A , and the general kets $|\gamma\rangle$ and $|\alpha\rangle$ instead of the eigenkets. I write

$$|\gamma\rangle = B |\alpha\rangle. \quad (5.39)$$

I now sandwich with $\langle a'|$ from the left and insert the completeness relation

$$\langle a'|\gamma\rangle = \sum_{a''} \langle a'| B |a''\rangle \langle a''|\alpha\rangle. \quad (5.40)$$

¹² $\psi_{a'}(x)$ is completely analog to $\psi_n(x)$.

Equation (5.40) can be considered as a linear transformation. The matrix is represented by the elements $\langle a' | B | a'' \rangle$ whereas the kets are represented by the vector elements $|\alpha\rangle \doteq \langle a''|\alpha\rangle$ and $|\gamma\rangle \doteq \langle a'|\gamma\rangle$. Thus, I can represent the kets in vector notation as

$$|\alpha^j\rangle \doteq \begin{bmatrix} \langle a'|\alpha^j\rangle \\ \langle a''|\alpha^j\rangle \\ \langle a'''|\alpha^j\rangle \\ \vdots \\ \langle a^i|\alpha^j\rangle \end{bmatrix}, \quad |\gamma^j\rangle \doteq \begin{bmatrix} \langle a'|\gamma^j\rangle \\ \langle a''|\gamma^j\rangle \\ \langle a'''|\gamma^j\rangle \\ \vdots \\ \langle a^i|\gamma^j\rangle \end{bmatrix}, \quad (5.41)$$

where the superscript "j" denotes that there are different eigenvectors of B - each corresponding to an eigenvalue. As for the bra's; in matrix notation they are represented as the complex conjugate transposed of the kets. Thus, the inner product can be written in matrix form as

$$\langle a|b\rangle = [a_1^* \ a_2^* \ a_3^* \ \dots \ a_i^*] \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_i \end{bmatrix} = a_1^*b_1 + \dots a_i^*b_i. \quad (5.42)$$

So, what is the purpose of going to a matrix notation instead of the notation of Dirac? Well, first of all; most people are more used to linear algebra, so it is more familiar to them, but the main reason is because the eigenvalue problem, given eg. by equation (5.24), can be solved quite neatly by using linear algebra. To illustrate this I consider the task of finding the eigenvalues and eigenvectors of an operator, B . The eigenvectors and eigenvalues of B are defined from the eigenvalue equation

$$B|b'\rangle = b'|b'\rangle, \quad (5.43)$$

where b' and $|b'\rangle$ are to be determined. I sandwich equation (5.43) from the left with $\langle a''|$, which I assume is from the *known* set $\{|a'\rangle\}$ (else I would not be able to calculate the matrix elements), and use the completeness relation¹³

$$\sum_{a'} \langle a''|B|a'\rangle \langle a'|b'\rangle = b' \langle a''|b'\rangle. \quad (5.44)$$

Equation (5.44) can be written in matrix notation as

$$\begin{bmatrix} \langle a'|B|a'\rangle & \langle a'|B|a''\rangle & \dots \\ \langle a''|B|a'\rangle & \langle a''|B|a''\rangle & \dots \\ \vdots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \langle a'|b'\rangle \\ \langle a''|b'\rangle \\ \vdots \end{bmatrix} = b' \begin{bmatrix} \langle a'|b'\rangle \\ \langle a''|b'\rangle \\ \vdots \end{bmatrix}. \quad (5.45)$$

From Cramers rule, equation (5.45) only has non-trivial for $\langle a^{(k)}|b^{(l)}\rangle$ when

$$\det(B - \lambda I) = 0, \quad (5.46)$$

where I is the identity matrix. This corresponds to diagonalizing¹⁴ B . The eigenvalues of B are the eigenvalues of B and the eigenvectors of B make up the coefficients $\langle a^{(k)}|b^{(l)}\rangle$, which are the elements of the eigenvectors of B . Thus

$$\lambda_i = b^i = \text{eigenvalues of } B, \quad (5.47)$$

$$|b'\rangle \doteq \begin{bmatrix} \langle a'|b'\rangle \\ \langle a''|b'\rangle \\ \vdots \end{bmatrix} = \text{the } b'\text{'th eigenvector of } B. \quad (5.48)$$

¹³I take $|a'\rangle$ to (still) be the eigenkets of an observable, and therefore make up a complete set of orthogonal vectors. Thus, i can use the completeness relation.

¹⁴When diagonalization of an operator is mentioned the operator is always implicitly thought of in matrix notation.

The elements, $\langle a^{(k)} | b^{(l)} \rangle$, of the eigenvectors, $|b^{(l)}\rangle$, can be arranged into a matrix called the *transformation matrix* (U) , which can be used to transform an operator, or state vector, from the basis of $|a'\rangle$ to the basis of $|b'\rangle$. I will come back to how the eigenvectors fit into the transformation matrix, but before I do so, let me make a formal introduction of the transformation operator! A transformation of basis becomes relevant when one considers two non-commuting observables. These does not have simultaneous eigenvectors, but since they are both observables, their eigenvectors clearly make a complete, orthogonal set and span the same vector space. Thus, one might be interested in changing from representing a given state vector, say $|\alpha\rangle$, in one basis to the other. To this end the transformation operator is used. To be more explicit, I consider the eigenkets of two non-commuting observables, A and B ($[A, B] \neq 0$); $\{|a'\rangle\}$ and $\{|b'\rangle\}$. I define the operator corresponding to the transformation matrix as the operator responsible for a basis transformation

$$U |a^{(l)}\rangle = |b^{(l)}\rangle. \quad (5.49)$$

The obvious representation of U is¹⁵

$$U = \sum_k |b^{(k)}\rangle \langle a^{(k)}|. \quad (5.50)$$

From equation (5.50) it is clear that U is a unitary operator ($UU^\dagger = 1$). I can represent U in matrix form by multiplying $\langle a^{(k)}|$ from the left

$$\langle a^{(k)}| U |a^{(l)}\rangle = \langle a^{(k)} | b^{(l)}\rangle. \quad (5.51)$$

From equation (5.51) it is clear that the elements of the transformation matrix is the elements of the eigenvectors of B . In matrix notation

$$U = \begin{bmatrix} \langle a^1 | b^1 \rangle & \langle a^1 | b^2 \rangle & \dots \\ \langle a^2 | b^1 \rangle & \langle a^2 | b^2 \rangle & \\ \vdots & & \ddots \end{bmatrix} = \begin{bmatrix} |b^1\rangle & |b^2\rangle & \dots \end{bmatrix}, \quad (5.52)$$

where the last "matrix" in equation (5.52) is to be interpreted such that the eigenvectors of B are listed from left to right as column vectors in the matrix U . So, say I have an operator C in the basis $|a'\rangle$ under consideration. I want to transform the basis to that of $|b'\rangle$. If I have $|b'\rangle$ in the basis of $|a'\rangle$ I just list the kets, $|b^1\rangle, |b^2\rangle, |b^3\rangle, \dots$ as columns in a matrix. That matrix is then the transformation matrix, U , responsible for transforming operators in the basis of $|a'\rangle$ to that of $|b'\rangle$! Now, in this case the new basis, $|b'\rangle$, is not the eigenbasis of the operator C , but it could have been - it applies in that case too.

To show how U transforms operators and kets to different basis', I consider the operator, C , expressed in the basis of $|a'\rangle$ - just like in equation (5.34) with $B \Rightarrow C$. I sandwich it with kets from the basis $|b'\rangle$

$$\begin{aligned} \overbrace{\langle b'' | C | b' \rangle}^{C'} &= \sum_{a'} \sum_{a''} \langle b'' | a'' \rangle \langle a'' | C | a' \rangle \langle a' | b' \rangle \\ &\Rightarrow C' = U^\dagger C U. \\ &= \sum_{a'} \sum_{a''} \underbrace{\langle a'' | U^\dagger | a'' \rangle}_{U^\dagger} \underbrace{\langle a'' | C | a' \rangle}_{C} \underbrace{\langle a' | U | a' \rangle}_U \end{aligned} \quad (5.53)$$

The above transformation transforms the operator C from the basis of $|a'\rangle$ to the basis of $|b'\rangle$ ($C \Rightarrow C'$). Equation (5.53) may seem complicated, but in matrix notation it is just the linear transformation of C from one basis to another, given by $C' = U^\dagger C U$. where the hermitian adjoint of the transformation matrix is the complex conjugate transposed; $U^\dagger = (U^*)^T$.

To illustrate how U performs a basis change for a general ket, I consider the case where I would like to

¹⁵Because it operates on $|a^{(l)}\rangle$ with $\langle a^{(k)}|$. Since there kets are orthogonal the inner product is 1 when $a^{(l)} = a^{(k)}$ and 0 otherwise. In the case where it is 1, the corresponding ket $|b^{(k)}\rangle = |b^{(l)}\rangle$ gets to "survive" - the rest of the b's don't.

change the basis to that of $\{|b'\rangle\}$, from the basis of $|a'\rangle$. I am therefore interested in finding the expansion coefficients in the basis; $\langle b^{(k)}|\alpha\rangle$. I find these by multiplying $\langle b^{(k)}|$ from the left and using equation (5.51)

$$\begin{aligned}\langle b^{(k)}|\alpha\rangle &= \sum_l \langle b^{(k)}|a^{(l)}\rangle \langle a^{(l)}|\alpha\rangle \\ &= \sum_l \langle a^{(k)}|U^\dagger|a^{(l)}\rangle \langle a^{(l)}|\alpha\rangle \\ &\doteq \begin{bmatrix} \langle b^1|a^1\rangle & \langle b^1|a^2\rangle & \langle b^1|a^3\rangle & \dots \\ \langle b^2|a^1\rangle & \langle b^2|a^2\rangle & \langle b^2|a^3\rangle & \dots \\ \vdots & & \ddots & \end{bmatrix} \begin{bmatrix} \langle a^1|\alpha\rangle \\ \langle a^2|\alpha\rangle \\ \vdots \end{bmatrix} = \begin{bmatrix} \langle b^1|\alpha\rangle \\ \langle b^2|\alpha\rangle \\ \vdots \end{bmatrix}.\end{aligned}\quad (5.54)$$

As is clear from equation (5.54); U^\dagger transforms the basis vectors from one basis to another. Thus, U is central to making a basis transformation of both operators and basis vectors. To recap in an intuitive notation

$$\text{Operator: } A_{\text{new}} = U^\dagger A_{\text{old}} U, \quad (5.55)$$

$$\text{Basis vectors: } \langle \text{new}|\alpha\rangle = U^\dagger \langle \text{old}|\alpha\rangle, \quad (5.56)$$

where A is just some general operator and the column vectors in U is the eigenvectors of the *new* basis, expressed in the *old* basis.

Example 5.1.

An operator, A , is given by

$$A = 2|1\rangle\langle 1| + 3|2\rangle\langle 2| + |1\rangle\langle 2| + |2\rangle\langle 1|. \quad (5.57)$$

where $|1\rangle$ and $|2\rangle$ are normalized and orthogonal.

1. Write A as a 2×2 matrix.

I take $|1\rangle$ and $|2\rangle$ as a complete, orthogonal basis and use equation (5.34) to represent the operator as a matrix

$$A \doteq \begin{bmatrix} \langle 1|A|1\rangle & \langle 1|A|2\rangle \\ \langle 2|A|1\rangle & \langle 2|A|2\rangle \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}. \quad (5.58)$$

2. Find the eigenvalues of $[A]$ and show they are real.

I use equations (5.43) through (5.45) to write

$$A|a'\rangle = a'|a'\rangle. \quad (5.59)$$

$$\sum_{b'} \langle b''|A|b'\rangle \langle b'|a'\rangle = a' \langle b''|a'\rangle. \quad (5.60)$$

Because the basis only contains $|1\rangle$ and $|2\rangle$

$$\begin{bmatrix} \langle 1|A|1\rangle & \langle 1|A|2\rangle \\ \langle 2|A|1\rangle & \langle 2|A|2\rangle \end{bmatrix} \begin{bmatrix} \langle 1|a'\rangle \\ \langle 2|a'\rangle \end{bmatrix} = \lambda \begin{bmatrix} \langle 1|a'\rangle \\ \langle 2|a'\rangle \end{bmatrix}. \quad (5.61)$$

Hereafter I use equation (5.46) to write

$$\det \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 3 - \lambda \end{bmatrix} = \lambda^2 - 5\lambda + 5 = 0. \quad (5.62)$$

Equation (5.62) reveals a second order polynomial. Defining $a = 1$, $b = -5$ and $c = 5$ the solutions are on the form

$$\lambda_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{5 \pm \sqrt{5}}{2}. \quad (5.63)$$

It is clear that both eigenvalues are real. Thereby A can be an observable since the eigenvalues, which are the possible results of measurement, is real - as they must be in order to be physically measurable.

3. Find the eigenkets and show they are orthogonal.

The eigenkets are found by finding the eigenvectors of A and hereafter using equation (5.48). By inserting λ_1 in equation (5.61) I find

$$\begin{bmatrix} 2 - \frac{5+\sqrt{5}}{2} & 1 \\ 1 & 3 - \frac{5+\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} \langle 1|a' \rangle \\ \langle 2|a' \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (5.64)$$

Taking just the first equation

$$(2 - \frac{5+\sqrt{5}}{2}) \langle 1|a' \rangle + \langle 2|a' \rangle = 0. \quad (5.65)$$

The other equation is not linearly independent, so it cannot be used. From equation (5.65) I can find the non-normalized eigenvector corresponding to λ_1 . To normalize it I require

$$|\langle 1|a' \rangle|^2 + |\langle 2|a' \rangle|^2 = 1, \quad (5.66)$$

where in general $\langle 1|a' \rangle$ can be a complex number. However, since the elements of A are all real numbers, the coefficients are also real. Thus, I can use¹⁶

$$\langle 1|a' \rangle^2 + \langle 2|a' \rangle^2 = 1. \quad (5.67)$$

Using normality, I find the eigenvector corresponding to λ_1 . Thus, if λ_1 is obtained from measuring A , the system is in the eigenstate denoted by

$$\lambda_1 = a' \Rightarrow |a'\rangle = \begin{bmatrix} \langle 1|a' \rangle \\ \langle 2|a' \rangle \end{bmatrix} = \begin{bmatrix} 0.526 \\ 0.851 \end{bmatrix} = 0.526|1\rangle + 0.851|2\rangle. \quad (5.68)$$

Likewise for λ_2 I find

$$\lambda_2 = a'' \Rightarrow |a''\rangle = \begin{bmatrix} \langle 1|a'' \rangle \\ \langle 2|a'' \rangle \end{bmatrix} = \begin{bmatrix} 0.851 \\ -0.526 \end{bmatrix} = 0.851|1\rangle - 0.526|2\rangle. \quad (5.69)$$

That the eigenvectors are orthogonal can be found from the inner product

$$\begin{aligned} \langle a'|a'' \rangle &= [0.526 \quad 0.851] \begin{bmatrix} 0.851 \\ -0.526 \end{bmatrix} \\ &= 0.526 \cdot 0.851 - 0.851 \cdot 0.526 = 0. \end{aligned} \quad (5.70)$$

Example 5.2.

Consider A from the previous example. A system is prepared in $|1\rangle$ before a measurement of A is performed.

1. What are the possibilities of measuring A ?

The possible outcomes of measuring an observable is the eigenvalues of the operator. thus, the possible values that can be measured is λ_i . For the same reason the eigenvalues of the operator must be real in order for the operator to be an observable!

2. What is the probability of measuring either outcome?

The probability of measuring λ_i , i.e. finding the system in a given state, is given as the squared amplitude of the inner product between the eigenstate and the initial state. The initial state goes to the right in the inner product, that is

$$\text{Probability of measuring } \lambda_i = |\langle a^i|1\rangle|^2. \quad (5.71)$$

¹⁶If the numbers could be imaginary, I would define: $\langle 1|a' \rangle = \text{real}$ and $\langle 1|a' \rangle = Ae^{i\theta}$. This would result in more complex criterion's for normality.

For example

$$\text{Probability of measuring } \lambda_1 = a' = |\langle a' | 1 \rangle|^2 = 0.526^2 = 0.276, \quad (5.72)$$

$$\text{Probability of measuring } \lambda_2 = a'' = |\langle a'' | 1 \rangle|^2 = 0.851^2 = 0.724, \quad (5.73)$$

where of course $|\langle a' | 1 \rangle|^2 + |\langle a'' | 1 \rangle|^2 = 1$ must hold, since the system must be in either of the possible states.

3. *What is the ket of the system in each case?*

The ket of the system is the ket corresponding to the eigenvalue that is measured. That is

$$\lambda_1 = a' \Rightarrow |a'\rangle, \quad (5.74)$$

$$\lambda_2 = a'' \Rightarrow |a''\rangle. \quad (5.75)$$

6 | Generalization of the classical schemes

With the mathematical setup of QM in the bag it is time to consider the generalization of the classical schemes to QM. To do so I will consider the time evolution in quantum mechanics. It turns out that there are different, equivalent, interpretations of time evolution in QM. Time evolution is represented by symmetry operator.

6.1 Symmetry operators

So far I have considered the mathematical setup of QM; I have introduced operators, vector spaces, kets, bras, commutators, wave functions and a useful way to view it all; in terms of matrices and vectors. The next step is to consider how to make things move in space and time - this is done by a special class of operators; symmetry operators!

The space and time evolution of the state vector is a special case of transformations that can be performed on the state of a system. This is also true in classical mechanics where the motion of the system is represented as a transformation of the phase space position. Transformations that do not change the properties of the system is called a symmetry transformation. A symmetry is a transformation on the space of physical configurations, so in quantum mechanics, it must be represented by a map from the Hilbert space onto itself. The transformation should map states to physically equivalent states, so the inner product between a pair of transformed states must be the same as the inner product between the original states. Because the inner product contains information about the probabilities of measurement, this can be restated as requiring conservation of probability. Conservation of probability leads to requiring the operator to be unitary. To be specific I define an infinitesimal symmetry operator ($\mathcal{S}(\epsilon)$), that differs infinitesimally from the unitary operator, to be given by

$$\mathcal{S}(\epsilon) = 1 - \frac{i\epsilon}{\hbar} G. \quad (6.1)$$

where ϵ is the infinitesimal transformation, the 1 comes from the fact that the transformation is infinitesimal and G is the Hermitian operator generating the symmetry in question. If a transformation is zero, then the original state must be recovered. Thus, an infinitesimal transformation must only differ slightly from the identity operator. Conservation of probability leads to

$$\langle \beta | \alpha \rangle = \langle \beta' | \alpha' \rangle = \langle \beta | \mathcal{S}(\epsilon)^\dagger \mathcal{S}(\epsilon) | \alpha \rangle \Rightarrow \mathcal{S}^\dagger(\epsilon) \mathcal{S}(\epsilon) = 1. \quad (6.2)$$

Which illustrates the requirement of the symmetry operator to be unitary. The group of symmetries where the symmetries are continuous is called the *Lie group*¹. This group contain, among other symmetries, spatial, time and rotational translation.

1. Translation operator:

The infinitesimal translation operator is defined as the operator translating the state of a system by an infinitesimal amount. That is

$$\mathcal{T}(d\vec{x}') |\vec{x}'\rangle = |\vec{x}' + d\vec{x}'\rangle. \quad (6.3)$$

¹See chapter 9

The generator of translation is linear momentum, and the infinitesimal transformation is in this case spatial translation. Thereby, from equation (6.1)

$$\mathcal{T}(d\vec{x}') = 1 - \frac{id\vec{x}'}{\hbar}\vec{p}. \quad (6.4)$$

The operator responsible for finite translations can be found by compounding infinitesimal translations

$$\mathcal{T}(\Delta\vec{x}') = \lim_{N \rightarrow \infty} \left(1 - \frac{i\Delta\vec{x}'}{\hbar N}\vec{p} \right)^N = e^{-\frac{i\Delta\vec{x}'\vec{p}}{\hbar}}, \quad (6.5)$$

where the exponential is understood to be a function of the vector operator \vec{p} . The exponential applied to a state will generate a finite translation of the state. That is

$$\begin{aligned} \mathcal{T}(\Delta\vec{x}') |\vec{x}'\rangle &= e^{-\frac{i\Delta\vec{x}'\vec{p}}{\hbar}} |\vec{x}'\rangle \\ &= \left(1 - \frac{i\vec{p}}{\hbar} \Delta\vec{x}' + \frac{(-i)^2}{2} \left(\frac{\vec{p}\Delta\vec{x}'}{\hbar} \right)^2 + \dots \right) |\vec{x}'\rangle \\ &= |\vec{x}' + \Delta\vec{x}'\rangle. \end{aligned} \quad (6.6)$$

Performing two subsequent translations corresponds to one translation of the length equal to the sum of the two translations. Hence

$$\mathcal{T}(d\vec{x}'')\mathcal{T}(d\vec{x}') = \mathcal{T}(d\vec{x}' + d\vec{x}''). \quad (6.7)$$

2. Rotational translation operator:

The rotational translation operator is defined as the operator rotating some state an infinitesimal angle around some axis. The generator of rotational translation is the angular momentum operator. By a completely analog procedure the infinitesimal rotational operator is given by

$$\mathcal{D}(\hat{e}_n, d\phi) = 1 - \frac{i\hat{e}_n \cdot \vec{J}}{\hbar} d\phi, \quad (6.8)$$

where \vec{J} is the angular momentum vector operator and \hat{e}_n denotes a unit vector in the \vec{n} -direction. For a finite rotational translation around the z -axis

$$\mathcal{D}_z(\Delta\phi) = \lim_{N \rightarrow \infty} \left(1 - \frac{i\Delta\phi}{\hbar N} J_z \right)^N = e^{-\frac{i\Delta\phi J_z}{\hbar}}. \quad (6.9)$$

3. Time translation operator:

The time translation operator can be defined in complete analogy to the translation operator; the time translation operator can be defined as the operator translating the state of a system by an infinitesimal amount in time. That is

$$\mathcal{U}(dt) |\alpha, t_0\rangle = |\alpha, t_0 + dt\rangle. \quad (6.10)$$

The generator of time translations is the Hamiltonian operator. The Hamiltonian operator describes the energy of the system and it can be found by quantizing the classical Hamiltonian. By using a generic, quantized Hamiltonian operator as the generator of time translations

$$\mathcal{U}(dt) = 1 - \frac{idt}{\hbar} H. \quad (6.11)$$

The specific form of the time evolution operator depends on the time dependence of the Hamiltonian operator. In analog to the translation operator the time evolution operator applied twice to the same state will be the same as applying a single time translation of the summed duration once. Hence

$$\mathcal{U}(t_2, t_1)\mathcal{U}(t_1, t_0) = \mathcal{U}(t_2, t_0). \quad (6.12)$$

From equation (6.10) it is clear that the ket, $|\alpha, t_0 + dt\rangle$, acquires a time dependence (on dt) from the time evolution operator. This interpretation of kets as time dependent, and operators as time-independent, is characteristic of one of the generalized schemes from classical mechanics. This scheme is called the Schroedinger scheme. Another scheme is defined by making another interpretation; consider a generic matrix element in which the states are time-dependent

$$\langle a(t)|C|b(t)\rangle = \langle a|U^\dagger CU|b\rangle. \quad (6.13)$$

From equation (6.13) the time evolution operators can be interpreted as acting upon the states - this is the Schroedinger picture. However, it is equally valid to interpret the time evolution operators as time-evolving the operator, i.e. C in this case, and leaving the kets time-independent. This gives rise to a second scheme of QM called the Heisenberg scheme. In this scheme the operators depend on time whereas the kets do not. These two different approaches will lead to two different equations of motion; one for the time-dependent state vector (the Schroedinger equation) and one for the time-dependent operator (the Heisenberg equation). These two equations define each scheme.

6.2 The Schroedinger Scheme

As mentioned the Schroedinger scheme is defined by interpreting the kets as time-dependent and the operators as time-independent. That is

$$|\alpha(t)\rangle_s \equiv |\alpha(t)\rangle = U(t, t_0)|\alpha\rangle \quad \wedge \quad A^s \neq A(t). \quad (6.14)$$

The index "s" denotes that the quantity belongs to the Schroedinger picture. The sub/super-scripts denoting the scheme are usually omitted in place of an explicit denotation of time-dependence of the kets and operators involved. To derive the EOM, the Schroedinger equation (SE), for the time-dependence of the states, which characterize the time-evolution of the system, I begin with the time evolution operator. I start out by exploiting the identity of the time evolution operator given by equation (6.12). By letting $t_1 = t$ and $t_2 = t + dt$

$$U(t + dt, t_0) = U(t + dt, t)U(t, t_0) = \left(1 - i\frac{H}{\hbar}dt\right)U(t, t_0). \quad (6.15)$$

By subtracting $U(t, t_0)$ the LHS is going to express the change in the time evolution operator ($U(t + dt, t_0) - U(t, t_0)$). Taking the differential of this and dividing dt to the LHS along with $i\hbar$

$$i\hbar \frac{\partial U(t, t_0)}{\partial t} = HU(t, t_0). \quad (6.16)$$

This equation is the SE equation for the time evolution operator. From equation (6.16) the specific form of the time evolution operator can be determined, depending on how the Hamiltonian operator depends on time. The simplest case, in which the Hamiltonian is independent of time, is the one encountered most often. Here equation (6.16) reduces to a first order differential equation from which the time evolution operator can be found to be

$$U(t, t_0) = e^{-\frac{iH(t-t_0)}{\hbar}}. \quad (6.17)$$

If the Hamiltonian depends on time the time evolution operator will take a different form.

From equation (6.16) the SE is found by multiplying both sides with $|\alpha\rangle$

$$i\hbar \frac{\partial |\alpha(t)\rangle}{\partial t} = H|\alpha(t)\rangle. \quad (6.18)$$

As is evident from equation (6.18) the time-evolution of the state vector is determined by the Hamiltonian operator. The Hamiltonian operator for a single particle is determined by quantizing the classical Hamiltonian via what is called *canonical quantization*. Canonical quantization consists of taking the dynamical variables of

classical mechanics, i.e. q, p , promoting them to operators and imposing the canonical commutator relation. The commutator is a generalization of the Poisson bracket cf.

$$\{A, B\} = \sum_i \frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} \Rightarrow \frac{1}{i\hbar} [A, B] = AB - BA. \quad (6.19)$$

The commutator denotes whether or not two operators interfere with one another; if two operators commute, i.e. $[A, B] = 0$, they do not interfere and they therefore have simultaneous eigenstates and can be measured simultaneously.

Example 6.1. —

Let A and B be observables. Suppose the simultaneous eigenkets of A and B , $\{|a', b'\rangle\}$ form a complete orthonormal set of base kets. Can it always be concluded that $[A, B] = 0$? Prove your answer.

If two observables have simultaneous eigenstates, their commutator will always be zero - that comes from the definition! So the answer is yes. This can be proved by first expanding the commutator on two completeness relations

$$[A, B] = \sum_{a', b'} \sum_{a'', b''} |a', b'\rangle \langle a', b'| [A, B] |a'', b''\rangle \langle a'', b''|. \quad (6.20)$$

If this is to be zero, then the following must be true

$$[A, B] |a'', b''\rangle = AB |a'', b''\rangle - BA |a'', b''\rangle = (a''b'' - b''a'') |a'', b''\rangle = 0. \quad (6.21)$$

This is always true since the last part is just numbers, and numbers commute. Thereby $[A, B] = 0$ for observables with simultaneous eigenstates.

Example 6.2. —

Two Hermitian operators anti-commute

$$\{A, B\} = AB + BA = 0. \quad (6.22)$$

Is it possible to have a simultaneous eigenket of A and B ?

No, it is not possible since it is the commutator, not the anti-commutator, that has to be zero. The one being zero eliminates the possibility of the other being zero in general.

Example 6.3. —

Consider a three dimensional ket space. If a certain set of orthogonal kets - say, $|1\rangle, |2\rangle, |3\rangle$ - are used as the base kets, the operators A and B are represented by

$$A \doteq \begin{bmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -a \end{bmatrix}, \quad B \doteq \begin{bmatrix} b & 0 & 0 \\ 0 & 0 & -ib \\ 0 & ib & 0 \end{bmatrix}, \quad (6.23)$$

where a and b are real.

1. Obviously A exhibits a degenerate spectrum. Does B also?

I set up the eigenvalue equation, and use Cramer's rule. Thereby

$$\det \begin{bmatrix} b - \lambda & 0 & 0 \\ 0 & -\lambda & -ib \\ 0 & ib & -\lambda \end{bmatrix} = (b - \lambda)(b - \lambda)(b + \lambda) = 0 \quad (6.24)$$

$$\Rightarrow \lambda_i = b, b, -b.$$

Thus, B has twofold degenerate eigenvalues - like A .

2. Show that A and B commute.

To this end I use the matrix representation

$$AB \doteq \begin{bmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -a \end{bmatrix} \begin{bmatrix} b - \lambda & 0 & 0 \\ 0 & -\lambda & -ib \\ 0 & ib & -\lambda \end{bmatrix} \doteq BA, \quad (6.25)$$

since $AB = BA$ the commutator; $[A, B] = AB - BA = 0$.

3. Find a new set of orthonormal kets that are simultaneous eigenkets of both A and B .

Since $[A, B] = 0$ the eigenvectors of B will do. I set up the eigenvalue equation

$$\begin{bmatrix} b - \lambda_i & 0 & 0 \\ 0 & -\lambda_i & -ib \\ 0 & ib & -\lambda_i \end{bmatrix} \begin{bmatrix} \langle 1|b_i \rangle \\ \langle 2|b_i \rangle \\ \langle 3|b_i \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (6.26)$$

The eigenvector corresponding to the first eigenvalue, $\lambda_1 = b$, result in the first equation, the only one involving $\langle 1|b_1 \rangle$. Thus, $\langle 1|b_1 \rangle$ is arbitrary². I at first set $\langle 1|b_1 \rangle = 1$. From normalization the other coefficients must be zero. Thereby

$$\lambda_1 = b \Rightarrow |b_1\rangle = \begin{bmatrix} \langle 1|b_1 \rangle \\ \langle 2|b_1 \rangle \\ \langle 3|b_1 \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = |1\rangle. \quad (6.27)$$

For the case of $\lambda_1 = b$ I denote the coefficients $\langle i|b_2 \rangle$. Before I put the first coefficient to 1, now I put it to zero; $\langle 1|b_2 \rangle = 0$. Hereby I can set $\langle 2|b_2 \rangle = 1$ and find $\langle 3|b_2 \rangle = i$. After normalization

$$\lambda_2 = b \Rightarrow |b_2\rangle = \begin{bmatrix} \langle 1|b_2 \rangle \\ \langle 2|b_2 \rangle \\ \langle 3|b_2 \rangle \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ i \end{bmatrix} = \frac{1}{\sqrt{2}}(|2\rangle + i|3\rangle). \quad (6.28)$$

For the case of $\lambda_1 = -b$ I find equation one revealing $b\langle 1|b_3 \rangle = 0$, and so $\langle 1|b_3 \rangle = 0$. I hereafter set $\langle 2|b_3 \rangle = 1$ and find $\langle 3|b_3 \rangle = -i$. After normalization

$$\lambda_3 = -b \Rightarrow |b_3\rangle = \begin{bmatrix} \langle 1|b_3 \rangle \\ \langle 2|b_3 \rangle \\ \langle 3|b_3 \rangle \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix} = \frac{1}{\sqrt{2}}(|2\rangle - i|3\rangle). \quad (6.29)$$

Two of the eigenvectors are at the moment characterized by the same eigenvalue. In order to make them distinguishable, I bring in a commuting observables eigenvalues; I bring in A 's eigenvalues. Since $[A, B] = 0$ the eigenvectors of B (and A) are simultaneous eigenvectors, and so the eigenvectors of B correspond to an eigenvalue of A also. I find out which corresponds to which

$$\begin{bmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -a \end{bmatrix} \begin{bmatrix} \langle 1|b_i \rangle \\ \langle 2|b_i \rangle \\ \langle 3|b_i \rangle \end{bmatrix} = \lambda_i^A \begin{bmatrix} \langle 1|b_i \rangle \\ \langle 2|b_i \rangle \\ \langle 3|b_i \rangle \end{bmatrix}. \quad (6.30)$$

Hereby

$$\begin{aligned} |b_1\rangle : \quad & \begin{bmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -a \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow |b_1\rangle = |b, a\rangle, \\ |b_2\rangle : \quad & \begin{bmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -a \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ i \end{bmatrix} = -a \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ i \end{bmatrix} \Rightarrow |b_2\rangle = |b, -a\rangle, \\ |b_3\rangle : \quad & \begin{bmatrix} a & 0 & 0 \\ 0 & -a & 0 \\ 0 & 0 & -a \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix} = -a \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -i \end{bmatrix} \Rightarrow |b_3\rangle = |-b, -a\rangle. \end{aligned} \quad (6.31)$$

²This is a result of the degenerate eigenvalues.

As it is evident, the three eigenvectors are now characterized by a unique set of eigenvalues of commuting operators.

Note that the commutator does not immediately look like the Poisson bracket. However, the commutator is derived from the desire to have a quantum mechanical quantity that obeys the same identities as the Poisson bracket. The canonical commutator relations are then defined as the QM generalization of the classical Poisson bracket between the generalized position and momenta, i.e.

$$\{q, p\} = 1 \Rightarrow \frac{1}{i\hbar}[q, p] = qp - pq = 1. \quad (6.32)$$

In this way the classical Hamiltonian function can be generalized to a QM operator cf.

$$H(q, p) = \frac{p^2}{2m} + V(q) \Rightarrow H(q, p) = \frac{p^2}{2m} + V(q). \quad (6.33)$$

In equation (6.33) there is no *apparent* difference in going from a classical to a quantum mechanical Hamiltonian. In reality there is however a very important difference; the difference lies in interpreting the momentum and position as operators rather than dynamical variables. From the Hamiltonian operator the time development of the state vector can be determined by specifying the Hamiltonian operator (i.e. specifying the potential specific to the physical system) and expanding the state vector in terms of eigenvectors of the Hamiltonian cf.

$$|\alpha(t)\rangle = \sum_n |n\rangle \langle n|\alpha(t)\rangle = \sum_n c_n(t) |n\rangle. \quad (6.34)$$

Hereafter the state vector can be projected into a given basis (most often the position basis) and the SE can be solved by means depending on the potential of the Hamiltonian operator. Most often the Hamiltonian will be independent of time. In this case the Schroedinger equation can be solved via separation of variables. In this case the Schroedinger equation can be separated into a time-dependent and time-independent equation cf.

$$\begin{aligned} \langle \vec{x} | \alpha(t) \rangle &= \sum_n c_n(t) \langle \vec{x} | n \rangle \\ &= \sum_n c_n(t) \psi_n(\vec{x}). \end{aligned} \quad (6.35)$$

Considering only the n'th functions, and using equation (6.35) in the Schrödinger equation

$$i\hbar \psi_n(\vec{x}) \frac{\partial c_n(t)}{\partial t} = c_n(t) H(\vec{x}, \vec{p}) \psi_n(\vec{x}) = c_n(t) E_n \psi_n(\vec{x}), \quad (6.36)$$

where, if the Hamiltonian depends on time $c_n(t)$ cannot be moved "through" the Hamiltonian. From the above two ODE's are obtained

$$\begin{aligned} \text{Time-Independent Schr. eq.: } &H(\vec{x}, \vec{p}) \psi_n(\vec{x}) = E_n \psi_n(\vec{x}), \\ \text{Time-dependent Schr.r eq.: } &i\hbar \frac{\partial c_n(t)}{\partial t} = c_n(t) E_n. \end{aligned} \quad (6.37)$$

By solving equation (6.37), and using equation (6.35), the time-evolution of the state vector in the position representation can be determined. What varies between physical systems is the potential of the Hamiltonian, so the time-dependent Schrödinger equation will be the same for all systems³. Equation (6.37) is a first order ODE and has a solution on the form

$$c_n(t) = c_n(t=0) e^{-\frac{iE_n}{\hbar}t} = c_n e^{-\frac{iE_n}{\hbar}t}. \quad (6.38)$$

³All systems in which the Hamiltonian does not depend on time that is.

Hence, the expansion of the state vector can, for a time-independent Hamiltonian⁴, be written as

$$|\alpha(t)\rangle = \sum_n c_n e^{-\frac{iE_n}{\hbar}t} |n\rangle, \quad (6.39)$$

where the coefficients c_n are independent of time. The time-independent Schroedinger equation, equation (6.37), is more complicated to solve in general. Unless the potential is sufficiently simple, the differential equation cannot be solved exactly and approximation methods must be used. See appendix 8.1 for perturbation theory in QM. After determining the state coefficients for the state vector, the probability for a given outcome of a measurement can be determined cf.

$$P_a = |\langle a|\alpha(t)\rangle|^2, \quad (6.40)$$

where a is one of the n states which the state vector is expanded upon.

Example 6.4.

Let $|a'\rangle$ and $|a''\rangle$ be eigenstates of a Hermitian operator A with eigenvalues a' and a'' , respectively ($a' \neq a''$). The Hamiltonian operator is given by

$$H = \delta |a'\rangle \langle a''| + \delta |a''\rangle \langle a'|, \quad (6.41)$$

where δ is a real number.

1. Clearly $|a'\rangle$ and $|a''\rangle$ is not eigenstates of the Hamiltonian (in that case the Hamiltonian would be diagonal). Write down the eigenstates of the Hamiltonian. What are their energy eigenvalues?

I write up the Hamiltonian on matrix form

$$H \doteq \begin{bmatrix} 0 & \delta \\ \delta & 0 \end{bmatrix}. \quad (6.42)$$

I diagonalize the matrix to find the eigenvalues and eigenvectors of the Hamiltonian:

$$\det \begin{bmatrix} -\lambda & \delta \\ \delta & -\lambda \end{bmatrix} = \lambda^2 - \delta^2 = 0 \Rightarrow \lambda_n = E_n = \pm\delta. \quad (6.43)$$

I determine the corresponding eigenvectors by inserting the eigenvalues one after the other into the eigenvalue equation

$$\begin{bmatrix} 0 - \lambda_n & \delta \\ \delta & -0 - \lambda_n \end{bmatrix} \begin{bmatrix} \langle a'|E_n\rangle \\ \langle a''|E_n\rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (6.44)$$

I use only one of the linearly dependent equations from each eigenvalue along with a normalization requirement ($(\langle 1|E_n\rangle)^2 + (\langle 2|E_n\rangle)^2 = 1$), to find the eigenvectors. I find

$$\lambda_1 = E_1 = \delta \Rightarrow |E_1\rangle = \begin{bmatrix} \langle a'|E_1\rangle \\ \langle a''|E_1\rangle \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|a'\rangle + |a''\rangle), \quad (6.45)$$

$$\lambda_2 = E_2 = -\delta \Rightarrow |E_2\rangle = \begin{bmatrix} \langle a'|E_2\rangle \\ \langle a''|E_2\rangle \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|a'\rangle - |a''\rangle). \quad (6.46)$$

2. Suppose the system is known to be in $|a'\rangle$ at $t = 0$. Write down the state vector in the Schroedinger picture for $t > 0$.

Since the Hamiltonians eigenvectors make up a complete orthogonal set I can develop the ket on them. Cf. equation (6.39)

$$\begin{aligned} |\alpha\rangle &= c_1 e^{-\frac{i\delta}{\hbar}t} |E_1\rangle + c_2 e^{\frac{i\delta}{\hbar}t} |E_2\rangle \\ &= \frac{1}{\sqrt{2}}(c_1 e^{-\frac{i\delta}{\hbar}t} |a'\rangle + c_1 e^{-\frac{i\delta}{\hbar}t} |a''\rangle + c_2 e^{\frac{i\delta}{\hbar}t} |a'\rangle - c_2 e^{\frac{i\delta}{\hbar}t} |a''\rangle) \\ &= \frac{1}{\sqrt{2}} \left((c_1 e^{-\frac{i\delta}{\hbar}t} + c_2 e^{\frac{i\delta}{\hbar}t}) |a'\rangle + (c_1 e^{-\frac{i\delta}{\hbar}t} - c_2 e^{\frac{i\delta}{\hbar}t}) |a''\rangle \right). \end{aligned} \quad (6.47)$$

⁴When the Hamiltonian depends on time perturbation theory is used. In perturbation theory a similar expression is used, however $c_n = c_n(t)$. This corresponds to pulling $e^{-\frac{iE_n}{\hbar}t}$ out of $c_n(t)$.

Now, to find the time independent coefficients.

$$t = 0 : \quad |\alpha(0)\rangle = |a'\rangle \Rightarrow c_1 = c_2. \quad (6.48)$$

From requiring normalization also I find $c_i = \frac{1}{\sqrt{2}}$. Thereby

$$|\alpha\rangle = \cos\left(\frac{\delta}{\hbar}t\right)|a'\rangle - i\sin\left(\frac{\delta}{\hbar}t\right)|a''\rangle \doteq \begin{bmatrix} \cos\left(\frac{\delta}{\hbar}t\right) \\ -i\sin\left(\frac{\delta}{\hbar}t\right) \end{bmatrix}. \quad (6.49)$$

Test by insertion into Schroedinger equation

$$i\hbar \frac{\partial}{\partial t} \begin{bmatrix} \cos\left(\frac{\delta}{\hbar}t\right) \\ -i\sin\left(\frac{\delta}{\hbar}t\right) \end{bmatrix} = \begin{bmatrix} 0 & \delta \\ \delta & 0 \end{bmatrix} \begin{bmatrix} \cos\left(\frac{\delta}{\hbar}t\right) \\ -i\sin\left(\frac{\delta}{\hbar}t\right) \end{bmatrix}. \quad (6.50)$$

Which checks out.

3. *What is the probability for finding the system in $|a''\rangle$ for $t > 0$ if the system is known to be in $|a'\rangle$ at $t = 0$?*

$$P = |\langle a''|\alpha\rangle|^2 = \sin^2\left(\frac{\delta}{\hbar}t\right). \quad (6.51)$$

4. *Can you think of a physical situation corresponding to this problem?*

The Hamiltonian is similar to S_x if $\delta = \frac{\hbar}{2}$. Therefore the state vector can be thought of as describing the time evolution of a state initially in the eigenstate of $|S_x; +\rangle$.

Relation to classical schemes

In relation to the classical schemes the Schroedinger scheme is the QM generalization of the HJ scheme. The generalization proceeds cf.

$$\frac{\partial S}{\partial t} + H\left(q, \frac{\partial q}{\partial t}\right) = 0 \Rightarrow i\hbar \frac{\partial \psi}{\partial t} = H(x, -i\hbar\nabla)\psi, \quad (6.52)$$

where the wave function is projected into the position basis, hence the position dependence of the Hamiltonian. The operator sign is omitted from the Hamiltonian because a QM operator is defined to act on a quantum state, i.e. a ket, whereas when the state vector is projected into the position basis, and the Hamiltonian likewise, the Hamiltonian acts on the wave function, and is as such not an operator any more. This is however debatable and interpreted differently in different books.

6.3 The Heisenberg Scheme

As mentioned the Heisenberg scheme is defined by interpreting the operators as time-dependent and the kets as time-independent. That is

$$|\alpha\rangle_h \neq |\alpha(t)\rangle \quad \wedge \quad A^h \equiv U^\dagger A^s U = A(t). \quad (6.53)$$

The index "h" denotes that the quantity belongs to the Heisenberg picture. Like for the Schroedinger picture the sub/super-scripts denoting the scheme are usually omitted in place of an explicit denotation of time-dependence of the kets and operators involved. The EOM for the time-dependence of the operator is found by generalizing equation (3.35) cf. the generalization of the Poisson bracket to QM

$$\frac{dA(t)}{dt} = \frac{1}{i\hbar}[A(t), H] + \frac{\partial A(t)}{\partial t}. \quad (6.54)$$

Often the operator has no explicit time dependence and the last term in equation (6.54) vanishes.

Example 6.5.

Consider an electron in a uniform, time-independent magnetic field of Strength B in the z -direction. I take the Hamiltonian to be given by

$$H = \frac{|q_e|B}{m_e c} S_z = \omega S_z, \quad (6.55)$$

where ω is the precession angular frequency around the z -axis. Write, and solve, the Heisenberg equations of motion for the time-dependent operators $S_i^h(t)$.

I consider Heisenberg's equation of motion

$$\frac{dS_i^h(t)}{dt} = \frac{1}{i\hbar} [S_i^h(t), H]. \quad (6.56)$$

Heisenberg's equation is applied by evaluating the commutator in the Schroedinger picture. Since in this case $H \propto S_z$ the relevant commutator relations are $[S_i, S_z]$. Using the commutation relations from angular momentum, which spin is just a special case of

$$[S_i, S_j] = i\hbar\epsilon_{ijk}S_k \Rightarrow [S_x, S_z] = -i\hbar S_y, [S_y, S_z] = i\hbar S_x, [S_z, S_z] = 0. \quad (6.57)$$

Using these I find three differential equations, two of which is coupled

$$\frac{dS_x^H(t)}{dt} = \frac{1}{i\hbar} [S_x, H] = -\omega S_y^h(t), \quad (6.58)$$

$$\frac{dS_y^H(t)}{dt} = \frac{1}{i\hbar} [S_y, H] = \omega S_x^h(t), \quad (6.59)$$

$$\frac{dS_z^H(t)}{dt} = \frac{1}{i\hbar} [S_z, H] = 0. \quad (6.60)$$

From equation (6.60) I immediately find $S_z^h(t) = \text{const}$. Since at $t = 0$ $S_i^h(t) = S_i$ (where no superscript is the Scrodinger picture) the operator S_z in the Heisenberg picture must equal the operator in the Schroedinger picture

$$S_z^h(t) = S_z. \quad (6.61)$$

In the case of the other two operators, I isolate $S_y^h(t)$ from equation (6.58) and insert it into equation (6.59) to obtain

$$\frac{d^2 S_x^h(t)}{dt^2} = -\omega^2 S_x^h(t). \quad (6.62)$$

This is a second order Ode with $a = 1, b = 0, c = \omega^2$ resulting in $D = b^2 - 4ac < 0$ and a general solution on the form

$$\frac{-b \pm \sqrt{D}}{2a} = k \pm iw \Rightarrow y(t) = Ae^{kt}\cos(wt) + Be^{kt}\sin(wt), \quad (6.63)$$

where in this particular case $b = 0 \Rightarrow k = 0$ and $w = \omega$. Thereby the solution is on the form

$$S_x^h(t) = A\cos(\omega t) + B\sin(\omega t). \quad (6.64)$$

Likewise for $S_y^H(t)$

$$S_y^h(t) = C\cos(\omega t) + D\sin(\omega t). \quad (6.65)$$

All I need to do now is to determine the coefficients A, B, C, D . From using that at $t = 0$ $S_i^h(t) = S_i$ I find $A = S_x$ and $C = S_y$. The remaining two coefficients are found by inserting the solutions into equation (6.58) and equation (6.59). From insertion into equation (6.58), and isolating B , I find

$$B = (A - D)\tan(\omega t) - C. \quad (6.66)$$

From insertion into equation (6.59), and isolating B , I find

$$B = (D - A) \frac{1}{\tan(\omega t)} - C. \quad (6.67)$$

Setting the two expressions equal reveals that $B = A$. Using this in either expression for B reveals $B = -C$. Thereby

$$S_x^h(t) = S_x \cos(\omega t) - S_y \sin(\omega t), \quad (6.68)$$

$$S_y^h(t) = S_y \cos(\omega t) + S_x \sin(\omega t). \quad (6.69)$$

Relation to classical schemes

The QM analog of the Hamiltonian scheme is the Heisenberg picture. The generalization proceeds cf.

$$\frac{dA}{dt} = \{A, H\} + \frac{\partial A}{\partial t} \Rightarrow \frac{dA^h}{dt} = \frac{1}{i\hbar} [A^h, H] + \frac{\partial A^h}{\partial t}. \quad (6.70)$$

6.4 The Path Integral Scheme

The path integral scheme is defined somewhat different as opposed to the Schroedinger and Heisenberg schemes. Instead of considering the time evolution of the states or operators, the path integral scheme formulates the transition probability in terms of a path integral. To derive the path integral I begin by considering the transition amplitude, \mathcal{A} , for the system to evolve between the state $|q_i\rangle$ at time t_i to the state $|q_f\rangle$ at time t_f

$$\mathcal{A} = \underbrace{\langle q_f | e^{iH(t_f - t_i)} | q_i \rangle}_{\text{Schroedinger scheme}} = \underbrace{\langle q_f, t_f | q_i, t_i \rangle}_{\text{Heisenberg scheme}}. \quad (6.71)$$

At any fixed time, the Heisenberg states make a complete, orthogonal set, so the completeness relation can be expressed cf.

$$1 = \int dq |q, t\rangle \langle q, t|. \quad (6.72)$$

Inserting $N - 1$ of these completeness relations into \mathcal{A} with small time separations ($t_{i+1} - t_i = \varepsilon$)

$$\begin{aligned} \mathcal{A} &= \langle q_f, t_f | q_i, t_i \rangle \\ &= \int dq_1 \int dq_2 \cdots \int dq_{N-1} \prod_{i=0}^{N-1} \langle q_{i+1}, t_{i+1} | q_i, t_i \rangle. \end{aligned} \quad (6.73)$$

The bracket can then be rewritten

$$\begin{aligned} \langle q_{i+1}, t_{i+1} | q_i, t_i \rangle &= \langle q_{i+1} | e^{-iHt_{i+1}} e^{iHt_i} | q_i \rangle \\ &= \langle q_{i+1} | e^{-iH\varepsilon} | q_i \rangle \\ &= \int dp_i \langle q_{i+1} | p_i \rangle \langle p_i | e^{-iH\varepsilon} | q_i \rangle \\ &= \int dp_i e^{iq_{i+1}p_i} \langle p_i | 1 - iH\varepsilon + \mathcal{O}(\varepsilon^2) | q_i \rangle \\ &= \int dp_i e^{iq_{i+1}p_i} e^{iH(q_i, p_i)\varepsilon} \langle p_i | q_i \rangle \\ &= \int dp_i e^{-i\varepsilon(H(q_i, p_i) - p_i \frac{q_{i+1} - q_i}{\varepsilon})}. \end{aligned} \quad (6.74)$$

Using this in \mathcal{A}

$$\mathcal{A} = \int dp_0 \int dq_1 \int dp_1 \int dq_2 \int dp_2 \cdots \int dq_{N-1} \int dp_{N-1} e^{-i\varepsilon \sum_{i=0}^{N-1} (H(q_i, p_i) - p_i \frac{q_{i+1} - q_i}{\varepsilon})}. \quad (6.75)$$

Taking now the limit $\varepsilon \Rightarrow 0$ the integral runs over an infinite number of interaction variables and defines a functional integral. Further more, the $\frac{q_{i+1}-q_i}{\varepsilon}$ -term becomes a derivative and $\sum \Rightarrow \int$. Defining $\lim_{\varepsilon \Rightarrow 0} (N(\varepsilon) \int dp_0 \int dq_1 \int dp_1 \int dq_2 \int dp_2 \cdots \int dq_{N-1} \int dp_{N-1}) = \int Dq \int Dp$, where $N(\varepsilon)$ is a normalization constant, the transition amplitude can be written as

$$\mathcal{A} = \int Dq \int Dp e^{i \int dt' (p\dot{q} - H(q, p))}. \quad (6.76)$$

Taking $H(q, p) = \frac{p^2}{2m} - V(q)$ means the integral over p is a set of Gaussian integrals. Absorbing these integrals into the normalization constant the transition amplitude can be written as

$$\mathcal{A} = \int Dq e^{iS}. \quad (6.77)$$

Equation (6.77) is the final expression for the transition amplitude in terms of a path integral. The physical interpretation of equation (6.77) can be understood in terms of the classical scheme.

Relation to classical schemes

The QM analog of Lagrangian mechanics is the path integral scheme. This generalization to QM is less straightforward than the other schemes. The central element of Lagrangian mechanics is the Euler-Lagrange equation. In deriving the Euler-Lagrange equation it is used that the path a system takes through configuration space is the one that minimizes the action. In QM this is however not true, and *all* paths of motion contribute. To "taste" this strange fact, consider the double split experiment; particles, one at a time, is shot through a sheet containing holes. If not detected, the registration of the different particle's locations will reveal an interference pattern. This illustrate that each particle takes all paths through the sheet, and the different paths interfere with each other. No matter how many holes there is in the sheet, poke another one and the interference pattern will change so that it stems from the new set of holes. Poke another one and the pattern will change again. Poke infinitely many and there will be no sheet and the particle is free to take *any* path, and all these possible paths will interfere. In QM each path is associated with a phase ($e^{i\frac{S}{\hbar}}$), and because random phases cancel, many contributions cancel. It is only when the phase is stationary (the action is minimum) that the nearby paths contribute. This idea is in QM generalized to the path integral formalism which is a way of calculating probability amplitudes. The analogy

$$\frac{\partial L}{\partial q} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) = 0 \Rightarrow \mathcal{A} \equiv \langle q_f, t_f | q_i, t_i \rangle = \int Dq e^{i\frac{S}{\hbar}}. \quad (6.78)$$

7 | Angular momentum in quantum mechanics

As illustrated in the previous chapter, angular momentum is the generator of rotations. The angular momentum operator (\vec{J}) is defined as the sum of two sources; orbital angular momentum ($\vec{L} = \vec{x} \times \vec{p}$) and spin angular momentum (\vec{S}). Orbital angular momentum is the quantity known from classical mechanics. It originates from movements external to some particle, eg. an electron around a nucleus. Spin angular momentum originates due to intrinsic properties of a particle.

Example 7.1.

Represent the components of the spin operator, S_i , in matrix notation. Use the basis of S_z .

Consider a spin half system. This system has a two-dimensional Hilbert space; the system can be in spin up or spin down. To represent the components of the spin operator in matrix notation, a basis(a complete set of orthogonal basis vectors) must be chosen. In the case of a spin half system the system can either be in spin-up-state, or spin-down-state, and so the complete set of basis vectors will just consist of two vectors in this case. There are several sets of vectors that make up a complete, orthogonal set for this system. Since the spin operator is an observable, the eigenvectors of S_i each make up such a set. By convention the eigenvectors of S_z is taken as the basis. In this basis set the completeness relation can be written as follows¹

$$|S_z; +\rangle \langle S_z; +| + |S_z; -\rangle \langle S_z; -| = 1. \quad (7.1)$$

To express the operator S_z in terms of its eigenvectors, use the eigenvalues ($\pm \frac{\hbar}{2}$) and the basis from equation 7.1 in equation 5.37. Hereby

$$S_z = \frac{\hbar}{2} \left(|S_z; +\rangle \langle S_z; +| - |S_z; -\rangle \langle S_z; -| \right) \Rightarrow S_z \doteq \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \frac{\hbar}{2} \sigma^3, \quad (7.2)$$

where σ^i are the Pauli matrices defined in section 1.1. The other operators are expressed in terms of the eigenvectors of S_z ² by considering the Stern-Gerlach experiment (Sakurai, 1994, p.26). Here it is used that since the operators does not commute, the probability of the system to end up in $|S_z; \pm\rangle$ given the system is in $|S_x; +\rangle$, is 50/50. That is

$$|\langle S_z; \pm | S_x; +\rangle| = \frac{1}{2} \Rightarrow |S_x; +\rangle = \frac{1}{\sqrt{2}} (e^{i\alpha} |S_z; +\rangle + e^{i\beta} |S_z; -\rangle), \quad (7.3)$$

where the exponential terms represent the possible phase difference. Since the phase of the two are relative to each other, the phase of one of the coefficients can always be set to zero (i.e. $\alpha = 0$ and $\beta \Rightarrow \beta'$). Hereby

$$|S_x; +\rangle = \frac{1}{\sqrt{2}} (|S_z; +\rangle + e^{i\beta'} |S_z; -\rangle). \quad (7.4)$$

¹where $|S_z; +\rangle$ denotes that the ket represents the spin up state for the operator S_z . In reality the spin state is defined in terms of $|s, m\rangle$, where s denotes the spin angular momentum and m denotes the spin state. Since $m = \pm 1$ often \pm is written instead of $m = \pm 1$. In the case of an electron (spin half) the spin is trivial $s = \frac{1}{2}$ and is often omitted.

²Remember that the eigenvectors of S_z make up a complete set so they can also describe the other spin operators.

The same procedure can be repeated for $|S_x; +\rangle$ and $|S_y; \pm\rangle$. The phase is found by determining the inner product; $\langle S_y; \pm | S_x; \pm \rangle = \frac{1}{\sqrt{2}}$. Hereby, the eigenvectors of all the spin operators, in the basis of the eigenvectors of S_z , can be determined viz

$$\begin{aligned} |S_z; \pm\rangle &= |S_z; \pm\rangle \Rightarrow |S_z; +\rangle \doteq \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad |S_z; -\rangle \doteq \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \\ |S_x; \pm\rangle &= \frac{1}{\sqrt{2}}(|S_z; +\rangle \pm |S_z; -\rangle) \doteq \frac{1}{\sqrt{2}} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \pm \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \\ |S_y; \pm\rangle &= \frac{1}{\sqrt{2}}(|S_z; +\rangle \pm i |S_z; -\rangle) \doteq \frac{1}{\sqrt{2}} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \pm i \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right). \end{aligned} \quad (7.5)$$

Now that the eigenstates of the operators are determined, use these, along with the eigenvalues of the operators, to define the S_y, S_x in the basis of S_z viz

$$\begin{aligned} S_x &= \frac{\hbar}{2} \left(|S_z; +\rangle \langle S_z; -| + |S_z; -\rangle \langle S_z; +| \right) \Rightarrow S_x = \frac{\hbar}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \frac{\hbar}{2} \sigma^1, \\ S_y &= -\frac{i\hbar}{2} \left(|S_z; +\rangle \langle S_z; -| - |S_z; -\rangle \langle S_z; +| \right) \Rightarrow S_y = \frac{\hbar}{2} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} = \frac{\hbar}{2} \sigma^2. \end{aligned} \quad (7.6)$$

For two particles, of angular momenta \vec{J}_i , the total angular momentum is given as follows

$$\begin{aligned} \vec{J} &= \vec{J}_1 \otimes I_2 + \vec{J}_2 \otimes I_1 \\ &= (\vec{L}_1 \otimes I_{S_1} + \vec{S}_1 \otimes I_{L_1}) \otimes I_2 + (\vec{L}_2 \otimes I_{S_2} + \vec{S}_2 \otimes I_{L_2}), \end{aligned} \quad (7.7)$$

where I_S denotes the identity operator in spin space, I_L denotes the identity operator in position space and $I_{1,2}$ denotes the identity operators in the spaces of the two angular momenta corresponding to the two particles. The tensor product makes sure that the operators are of the same dimension and as such addable. If the angular momentum of more than two particles is considered, the angular momentum of the third particle can be added in the same way as the angular momenta, $\vec{J}_{1,2}$, are added. Often the tensor products are omitted and the angular momentum of a particle is just written as

$$\vec{J} = \vec{J}_1 + \vec{J}_2 = \vec{L}_1 + \vec{S}_1 + \vec{L}_2 + \vec{S}_2, \quad (7.8)$$

where the identity operators are understood. The angular momentum operator obeys the commutator relations given as follows

$$[J_{1i}, J_{2j}] = 0, \quad [J_i, J_j] = i\hbar\varepsilon_{ijk}J_k = i\hbar\varepsilon_{ijk}(L_k + S_k). \quad (7.9)$$

From equation (7.9) it is clear that angular momentum operators of different particles commute whereas the component of the same angular momentum operator does not commute. Writing out the commutator

$$\begin{aligned} [J_i, J_j] &= J_i J_j - J_j J_i \\ &= (L_i + S_i)(L_j + S_j) - (L_j + S_j)(L_i + S_i) \\ &= L_i L_j + L_i S_j + S_i L_j + S_i S_j - L_j L_i - L_j S_i - S_j L_i - S_j S_i \\ &= [L_i, L_j] + [L_i, S_j] + [S_i, L_j] + [S_i, S_j]. \end{aligned} \quad (7.10)$$

The orbital and angular momentum commutes, so $[L_i, S_j] = 0$, $[S_i, L_j] = 0$. By combining this with equation (7.9) and (7.10)

$$[L_i, L_j] = i\hbar\varepsilon_{ijk}L_k, \quad [S_i, S_j] = i\hbar\varepsilon_{ijk}S_k. \quad (7.11)$$

7.1 Eigenvalues and eigenstates of angular momentum

In order to determine the eigenstates of the angular momentum operator the notion of a complete set of commuting observables must be discussed. A complete set of commuting observables are defined viz

Complete set of commuting observables: A complete set of commuting observables is the set of observables (Hermitian operators) that is needed to uniquely label each eigenvector in a complete set.

As an example, take A to be an observable with eigenstates $|a'\rangle$. If the eigenvalues of A , i.e. a' , are non-degenerate, then all the eigenvectors, i.e. $|a'\rangle$, are uniquely specified. This can be generalized viz If the eigenvalues of an observable are non-degenerate, the eigenvectors of the observable are uniquely specified and the observable is said to, by itself, form a complete set of commuting observables. Conversely; if some of the eigenvalues were degenerate, all the eigenvectors would not be uniquely specified. In this case, one must find other observables (eg. B) that commute with A ($[A, B] = 0$). Commuting observables have simultaneous eigenstates, and so it is possible to label the states by eigenvalues from both observables. In this way, some of the states with the same value of a' may become uniquely specified by being assigned with a unique value of b' . If in this way all degenerate states become non-degenerate A, B is said to make up a complete set of commuting observables. If some degenerate states remain, more commuting observables must be introduced until no degenerate states remain. In terms of angular momentum eigenstates three observables are needed to uniquely specify each eigenstate. Several sets of observables that will do the job. For a system described by $\vec{J} = \vec{J}_1 + \vec{J}_2$, the following two sets are, by convention, used:

1. Set 1: Simultaneous eigenvectors of $\vec{J}_1^2, \vec{J}_2^2, J_{1z}$ and J_{2z} ; $| \dots \rangle = |j_1, j_2, m_1, m_2\rangle$.
2. Set 2: Simultaneous eigenvectors of $\vec{J}^2, \vec{J}_1^2, J_z$ and J_z ; $| \dots \rangle = |j_1, j_2, j, m\rangle$.

where J_z is the z -component of \vec{J}^2 , j_i belongs to the eigenvalues of \vec{J}_i^2 , j belongs to the eigenvalue of \vec{J}^2 , m_i belongs to the eigenvalues of J_{iz} and m belongs to the eigenvalue of J_z . The eigenvalues of the operators belonging to the two sets are determined by introducing the ladder operators defined by

$$J_{\pm} \equiv J_x \pm iJ_y \quad (7.12)$$

The ladder operators obey

$$[J_+, J_-] = 2\hbar J_z, \quad [J_z, J_{\pm}] = \pm\hbar J_{\pm} \quad \text{and} \quad [\vec{J}^2, J_{\pm}] = 0. \quad (7.13)$$

Take the eigenvalues of \vec{J}^2, J_z to be denoted by a, b respectively , i.e.

$$\vec{J}^2 |a, b\rangle = a |a, b\rangle, \quad J_z |a, b\rangle = b |a, b\rangle. \quad (7.14)$$

The ladder operators then act viz

$$J_z(J_{\pm} |a, b\rangle) = ([J_z, J_{\pm}] + J_{\pm} J_z) |a, b\rangle = (b \pm \hbar) J_{\pm} |a, b\rangle. \quad (7.15)$$

From equation (7.15) it is clear that $J_{\pm} |a, b\rangle$ is an eigenvector of J_z with eigenvalue $(b \pm \hbar)$. Hence by applying J_{\pm} to a J_z eigenstate the eigenstate remains the same and the eigenvalue is changed by $\pm\hbar$. For \vec{J}^2

$$\vec{J}^2(J_{\pm} |a, b\rangle) = J_{\pm} \vec{J}^2 |a, b\rangle = a(J_{\pm} |a, b\rangle). \quad (7.16)$$

From equation (7.16) it is clear that $J_{\pm} |a, b\rangle$ is an eigenvector of \vec{J}^2 with eigenvalue a . Hence, the effect of J_{\pm} onto $|a, b\rangle$ can be written viz

$$J_{\pm} |a, b\rangle = c_{\pm} |a, b \pm \hbar\rangle, \quad (7.17)$$

where c_{\pm} is a proportionality constant to be determined by normalization later on. To determine the possible values of a, b , use that

$$\begin{aligned} \langle a, b | \vec{J}^2 - J_z^2 |a, b\rangle &= \langle a, b | (J_x^2 + J_y^2 + J_z^2) - J_z^2 |a, b\rangle \\ &= \langle a, b | J_x^2 + J_y^2 |a, b\rangle \\ &= \frac{1}{2} \langle a, b | J_+ J_- + J_- J_+ |a, b\rangle \\ &= \frac{1}{2} \langle a, b | J_+ J_+^\dagger + J_+^\dagger J_+ |a, b\rangle, \end{aligned} \quad (7.18)$$

where it has been used that $J_- = J_+^\dagger$. Both $\langle a, b | J_+ J_+^\dagger | a, b \rangle \geq 0$ and $\langle a, b | J_+^\dagger J_+ | a, b \rangle \geq 0$ and so

$$\langle a, b | \vec{J}^2 - J_z^2 | a, b \rangle \geq 0 \Rightarrow a \geq b^2. \quad (7.19)$$

Since $a \geq b^2$ there must be a maximum value for b . Acting with J_+ on this maximum value will annihilate the state, i.e.

$$J_+ |a, b_{max}\rangle = 0 \Rightarrow J_- J_+ |a, b_{max}\rangle = 0. \quad (7.20)$$

$J_- J_+ = \vec{J}^2 - J_z^2 - \hbar J_z$, so equation (7.20) implies

$$(\vec{J}^2 - J_z^2 - \hbar J_z) |a, b_{max}\rangle = 0 \Rightarrow a - b_{max}^2 - b_{max} \hbar = 0 \Leftrightarrow a = b_{max}(b_{max} + \hbar). \quad (7.21)$$

Similarly, $J_- |a, b_{min}\rangle = 0$ leads to $a = b_{min}(b_{min} - \hbar)$. By comparison; $b_{max} = -b_{min}$ and $-b_{max} \leq b \leq b_{max}$. $|a, b_{max}\rangle$ must be reachable by applying J_+ repeatedly to $|a, b_{min}\rangle$. Therefore $b_{max} = b_{min} + n\hbar$, where n is an integer denoting the number of times J_+ has been applied in going from the ground state to the highest state. Since $b_{min} = -b_{max}$

$$b_{max} = \frac{n\hbar}{2}. \quad (7.22)$$

Conventionally $j \equiv \frac{b_{max}}{\hbar}$ is used instead of b_{max} . Hereby

$$a = \hbar^2 j(j+1). \quad (7.23)$$

Defining $b = m\hbar$; $m = -j, \dots, j$. Hereby, the eigenvalues and eigenkets of the two sets are given viz

1. *Set 1:* Simultaneous eigenvectors of \vec{J}_1^2 , \vec{J}_2^2 , J_{1z} and J_{2z} ;

$$\begin{aligned} \vec{J}_1^2 |j_1, j_2, m_1, m_2\rangle &= j_1(j_1+1)\hbar^2 |j_1, j_2, m_1, m_2\rangle, \\ \vec{J}_2^2 |j_1, j_2, m_1, m_2\rangle &= j_2(j_2+1)\hbar^2 |j_1, j_2, m_1, m_2\rangle, \\ J_{1z} |j_1, j_2, m_1, m_2\rangle &= m_1 \hbar |j_1, j_2, m_1, m_2\rangle, \\ J_{2z} |j_1, j_2, m_1, m_2\rangle &= m_2 \hbar |j_1, j_2, m_1, m_2\rangle. \end{aligned} \quad (7.24)$$

2. *Set 2:* Simultaneous eigenvectors of \vec{J}^2 , \vec{J}_1^2 , J_2 and J_z

$$\begin{aligned} \vec{J}_1^2 |j_1, j_2, j, m\rangle &= j_1(j_1+1)\hbar^2 |j_1, j_2, j, m\rangle, \\ \vec{J}_2^2 |j_1, j_2, j, m\rangle &= j_2(j_2+1)\hbar^2 |j_1, j_2, j, m\rangle, \\ \vec{J}^2 |j_1, j_2, j, m\rangle &= j(j+1)\hbar^2 \hbar |j_1, j_2, j, m\rangle, \\ J_z |j_1, j_2, j, m\rangle &= m \hbar |j_1, j_2, j, m\rangle. \end{aligned} \quad (7.25)$$

The two basis sets are related via a set of coefficients denoted by Clebsch-Gordon coefficients. These coefficients will be the subject of example 8.4.

Example 7.2. —

1. If two particles have spin $\frac{1}{2}$, the direct product spin space of the composite system is four dimensional. A particular basis is spanned by the eigenvectors of $S_z^{(1)}$ and $S_z^{(2)}$

$$\{|+, +\rangle, |+, -\rangle, |-, +\rangle, |-, -\rangle\}. \quad (7.26)$$

In this particular basis, find the elements of the 4×4 matrices S_z and \vec{S}^2 - where $\vec{S} = \vec{S}_1 + \vec{S}_2 = S_x^2 + S_y^2 + S_z^2$.

Begin by determining the matrix forms of S_i . From these \vec{S}^2 can be determined. Use that S_i can be found via

$$S_i = S_i^{(1)} \otimes I_{S^{(2)}} + I_{S^{(1)}} \otimes S_i^{(2)}, \quad (7.27)$$

where $I_{S^{(2)}}$ is the identity matrix with dimensions corresponding to $S_i^{(2)}$. The tensor product is evaluated by using that in general

$$A \otimes B \doteq \begin{bmatrix} A_{11}B & A_{12}B & A_{13}B & \dots \\ A_{21}B & A_{22}B & A_{23}B & \dots \\ A_{31}B & A_{32}B & A_{33}B & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad (7.28)$$

where B in the matrix denotes an entire matrix multiplied to A_{ij} . In this case the two matrices are two-dimensional, so

$$\begin{aligned} S_x &= S_x^{(1)} \otimes I + I \otimes S_x^{(2)} \\ &= \begin{bmatrix} 0 \cdot I & \frac{\hbar}{2} \cdot I \\ \frac{\hbar}{2} \cdot I & 0 \cdot I \end{bmatrix} + \begin{bmatrix} 1 \cdot S_x^{(2)} & 0 \cdot S_x^{(2)} \\ 0 \cdot S_x^{(2)} & 1 \cdot S_x^{(2)} \end{bmatrix} \\ &= \frac{\hbar}{2} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} + \frac{\hbar}{2} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\ &= \frac{\hbar}{2} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \end{aligned} \quad (7.29)$$

$$\begin{aligned} S_y &= S_y^{(1)} \otimes I + I \otimes S_y^{(2)} \\ &= \begin{bmatrix} 0 \cdot I & -\frac{i\hbar}{2} \cdot I \\ \frac{i\hbar}{2} \cdot I & 0 \cdot I \end{bmatrix} + \begin{bmatrix} 1 \cdot S_y^{(2)} & 0 \cdot S_y^{(2)} \\ 0 \cdot S_y^{(2)} & 1 \cdot S_y^{(2)} \end{bmatrix} \\ &= \frac{\hbar}{2} \begin{bmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{bmatrix} + \frac{\hbar}{2} \begin{bmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{bmatrix} \\ &= \frac{\hbar}{2} \begin{bmatrix} 0 & -i & -i & 0 \\ i & 0 & 0 & -i \\ i & 0 & 0 & -i \\ 0 & i & i & 0 \end{bmatrix}, \end{aligned} \quad (7.30)$$

$$\begin{aligned} S_z &= S_z^{(1)} \otimes I + I \otimes S_z^{(2)} \\ &= \begin{bmatrix} \frac{\hbar}{2} \cdot I & 0 \cdot I \\ 0 \cdot I & -\frac{\hbar}{2} \cdot I \end{bmatrix} + \begin{bmatrix} 1 \cdot S_z^{(2)} & 0 \cdot S_z^{(2)} \\ 0 \cdot S_z^{(2)} & 1 \cdot S_z^{(2)} \end{bmatrix} \\ &= \frac{\hbar}{2} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} + \frac{\hbar}{2} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \\ &= \hbar \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \end{aligned} \quad (7.31)$$

At last \vec{S}^2 can be determined viz

$$\vec{S}^2 = S_x^2 + S_y^2 + S_z^2 = \hbar^2 \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}. \quad (7.32)$$

2. Show by explicit diagonalization of the \vec{S}^2 -matrix that the singlet and triplet states are the eigenvectors with the appropriate eigenvalues.

The eigenvalues of \vec{S}^2

$$\vec{S}^2 : \quad \lambda_i = j(j+1)\hbar^2 = 0, 2\hbar^2, 2\hbar^2, 2\hbar^2 \quad \Rightarrow \quad j = 0, 1, 1, 1, \quad (7.33)$$

where $j = 1$ is threefold degenerate (the triplet state) and $j = 0$ is non-degenerate (the singlet state). The corresponding eigenvectors

$$\begin{aligned} \vec{S}^2 : \quad \lambda_1^{(j=0,m=0)} = 0 : \quad |\lambda_1\rangle &= \begin{bmatrix} \langle +, +|\lambda_1 \rangle \\ \langle +, -|\lambda_1 \rangle \\ \langle -, +|\lambda_1 \rangle \\ \langle -, -|\lambda_1 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \\ \vec{S}^2 : \quad \lambda_2^{(j=1,m=1)} = \frac{\hbar^2}{2} : \quad |\lambda_2\rangle &= \begin{bmatrix} \langle +, +|\lambda_2 \rangle \\ \langle +, -|\lambda_2 \rangle \\ \langle -, +|\lambda_2 \rangle \\ \langle -, -|\lambda_2 \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \\ \vec{S}^2 : \quad \lambda_3^{(j=1,m=-1)} = \frac{\hbar^2}{2} : \quad |\lambda_3\rangle &= \begin{bmatrix} \langle +, +|\lambda_3 \rangle \\ \langle +, -|\lambda_3 \rangle \\ \langle -, +|\lambda_3 \rangle \\ \langle -, -|\lambda_3 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \\ \vec{S}^2 : \quad \lambda_4^{(j=1,m=0)} = \frac{\hbar^2}{2} : \quad |\lambda_4\rangle &= \begin{bmatrix} \langle +, +|\lambda_4 \rangle \\ \langle +, -|\lambda_4 \rangle \\ \langle -, +|\lambda_4 \rangle \\ \langle -, -|\lambda_4 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \end{aligned} \quad (7.34)$$

The eigenvectors must be simultaneous eigenstates of S_z . The eigenvalues of S_z

$$S_z : \quad \lambda'_i = m\hbar = -\hbar, 0, \hbar, 0 \quad \Rightarrow \quad m = -1, 0, 1, 0. \quad (7.35)$$

As is apparent from the eigenvalues of the two matrices, the allowed states are

$$|j = 1, m = 1\rangle, |j = 1, m = 0\rangle, |j = 1, m = -1\rangle, |j = 0, m = 0\rangle, \quad (7.36)$$

where the triplet state is $j = 1$ and the singlet state is $j = 0$. The eigenvectors of \vec{S}^2 listed above are also eigenvectors of S_z - they need not necessarily be so! There exists simultaneous eigenvectors since the two observables commute, but there also exists eigenvectors unique to the two matrices. For example $|+, -\rangle$ is an eigenvector of S_z but not of \vec{S}^2 . So one should really check that the eigenvectors are simultaneous!³

As a last thing the linear combinations of $|+, -\rangle = |m_1, m_2\rangle$ which results in $|j, m\rangle$ will be determined. Begin by determining the maximum state. Since $m = m_1 + m_2$ and $m = -j,..j$ the case of $m = 1$ and $m = -1$ are trivial

$$|\psi_2\rangle = |j = 1, m = 1\rangle = |+, +\rangle, \quad (7.37)$$

$$|\psi_3\rangle = |j = 1, m = -1\rangle = |-, -\rangle. \quad (7.38)$$

³This can be done by checking that $S_z \cdot \text{eigenvec} = \text{eigenval} \cdot \text{eigenvec}$, cf. equation (5.45).

From this $|j = 1, m = 0\rangle$ can be found by using the lowering or raising operator. Alternatively; the eigenvector corresponding to $\lambda_1 = 0$ corresponds, cf. the eigenvalue, to $|j = 0, m = 0\rangle$. $|\psi_2\rangle$ has both particles in spin up, and so must correspond to $|j = 1, m = 1\rangle$. $|\psi_3\rangle$ has both particles in spin down and must correspond to $|j = 1, m = -1\rangle$. $|\psi_3\rangle$ has $j = 1$ and $m = 0$, so it must correspond to $|j = 1, m = 0\rangle$. That is

$$|\psi_1\rangle = |j = 0, m = 0\rangle = \frac{1}{\sqrt{2}}(|+, -\rangle - |-, +\rangle), \quad (7.39)$$

$$|\psi_4\rangle = |j = 1, m = 0\rangle = \frac{1}{\sqrt{2}}(|+, -\rangle + |-, +\rangle). \quad (7.40)$$

Example 7.3.

In a system of three spin $\frac{1}{2}$ particles, a basis for the states of the system is given by the eight product vectors $|m_1\rangle \otimes |m_2\rangle \otimes |m_3\rangle$, where $m_i = \pm \frac{1}{2}$. In terms of these product vectors, find the eigenvectors of the total angular momentum operators \vec{J}^2 and J_z , where $\vec{J} = \vec{S}_1 + \vec{S}_2 + \vec{S}_3$.

Since three spin half particles are involved, the maximum value of j must be $j = \frac{3}{2}$. Now, to find the minimum value consider the scenarios; If all three particles point in the same direction, the value is $j = \frac{3}{2}$ - that goes for the one or the other direction. If two particles point in the same direction $j = \frac{1}{2}$ - these are the only options. Therefore

$$j = \frac{3}{2} \Rightarrow m = \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}, \quad (7.41)$$

$$j = \frac{1}{2} \Rightarrow m = \frac{1}{2}, -\frac{1}{2}. \quad (7.42)$$

Since the above only gives rise to six different kets on the form $|j, m\rangle$, two vectors must be degenerate! Hence, the procedure is to determine the matrix-representations of \vec{J}^2 and J_z and determine the simultaneous eigenvectors of the two matrices. Since there are three particles, the operators S_i are given by taking the tensor product of three two-dimensional operators. Alternatively use that in example 8.2 S_i for two spin $\frac{1}{2}$ particles have been determined. Hence, in "adding" one more spin half to these results will suffice, i.e.

$$S_x = S_x^{(1)} \otimes I_{S^{(2)}} + I_{S^{(1)}} \otimes S_x^{(2)}. \quad (7.43)$$

However now

$$S_x^{(1)} = \frac{\hbar}{2} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad S_x^{(2)} = \frac{\hbar}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (7.44)$$

where it is clear that $S_x^{(1)}$ is already the tensor product between two of the spin half particles. This results in

$$\begin{aligned} S_x &= \frac{\hbar}{2} \begin{bmatrix} 0 & I_{S^{(2)}} & I_{S^{(2)}} & 0 \\ I_{S^{(2)}} & 0 & 0 & I_{S^{(2)}} \\ I_{S^{(2)}} & 0 & 0 & I_{S^{(2)}} \\ 0 & I_{S^{(2)}} & I_{S^{(2)}} & 0 \end{bmatrix} + \begin{bmatrix} S_x^{(2)} & 0 & 0 & 0 \\ 0 & S_x^{(2)} & 0 & 0 \\ 0 & 0 & S_x^{(2)} & 0 \\ 0 & 0 & 0 & S_x^{(2)} \end{bmatrix} \\ &= \frac{\hbar}{2} \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} + \frac{\hbar}{2} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \end{aligned} \quad (7.45)$$

where every zero in the 4×4 notation is a 2×2 matrix of zeros.

$$S_y = \frac{i\hbar}{2} \begin{bmatrix} 0 & -I_{S^{(2)}} & -I_{S^{(2)}} & 0 \\ I_{S^{(2)}} & 0 & 0 & -I_{S^{(2)}} \\ I_{S^{(2)}} & 0 & 0 & -I_{S^{(2)}} \\ 0 & I_{S^{(2)}} & I_{S^{(2)}} & 0 \end{bmatrix} + \begin{bmatrix} S_y^{(2)} & 0 & 0 & 0 \\ 0 & S_y^{(2)} & 0 & 0 \\ 0 & 0 & S_y^{(2)} & 0 \\ 0 & 0 & 0 & S_y^{(2)} \end{bmatrix}$$

$$= \frac{i\hbar}{2} \begin{bmatrix} 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} + \frac{i\hbar}{2} \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (7.46)$$

$$S_z = \hbar \begin{bmatrix} I_{S^{(2)}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -I_{S^{(2)}} \end{bmatrix} + \begin{bmatrix} S_z^{(2)} & 0 & 0 & 0 \\ 0 & S_z^{(2)} & 0 & 0 \\ 0 & 0 & S_z^{(2)} & 0 \\ 0 & 0 & 0 & S_z^{(2)} \end{bmatrix}$$

$$= \hbar \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} + \frac{\hbar}{2} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad (7.47)$$

which results in

$$\vec{S}^2 = S_x S_x + S_y S_y + S_z S_z = \hbar^2 \begin{bmatrix} \frac{15}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{7}{4} & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & \frac{7}{4} & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{7}{4} & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & \frac{7}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \frac{7}{4} & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & \frac{7}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{15}{4} \end{bmatrix}. \quad (7.48)$$

Because of the dimensionality of the involved matrices it is convenient to use a computer program to compute the eigenvectors and values. In this instance the function "genvec(A,B)" of Mathcad has been used to determine the simultaneous eigenvectors of matrices A and B . the results

$$|\lambda_1\rangle = \sqrt{\frac{2}{3}} \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 0 \\ 0.5 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad |\lambda_2\rangle = \sqrt{\frac{1}{3}} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad |\lambda_3\rangle = \sqrt{\frac{2}{3}} \begin{bmatrix} 0 \\ -1 \\ 0.5 \\ 0 \\ 0.5 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad |\lambda_4\rangle = \sqrt{\frac{1}{3}} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (7.49)$$

$$|\lambda_5\rangle = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad |\lambda_6\rangle = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad |\lambda_7\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad |\lambda_8\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad (7.50)$$

where

$$|\lambda_i\rangle = \begin{bmatrix} \langle +, +, + | \lambda_i \rangle \\ \langle +, +, - | \lambda_i \rangle \\ \langle +, -, + | \lambda_i \rangle \\ \langle +, -, - | \lambda_i \rangle \\ \langle -, +, + | \lambda_i \rangle \\ \langle -, +, - | \lambda_i \rangle \\ \langle -, -, + | \lambda_i \rangle \\ \langle -, -, - | \lambda_i \rangle \end{bmatrix}. \quad (7.51)$$

Since the vectors are the tensor products between the vectors with two and 4 elements, lock the leftmost index and permute the two at the right cf. how it is done in example 8.2. Next repeat the permutation of the two at the right but switch the index at the left. What remains is to label the eigenvectors. To this end use that

$$S_z |\lambda_i\rangle = m\hbar |\lambda_i\rangle \quad \text{and} \quad \vec{S}^2 |\lambda_i\rangle = j(j+1)\hbar^2 |\lambda_i\rangle. \quad (7.52)$$

By inserting the different eigenvectors determine the corresponding eigenvalues (j, m) . The eigenvectors in the common notation; Ladder 1

$$\begin{aligned} |j = \frac{3}{2}, m = \frac{3}{2}\rangle &= |+, +, +\rangle \\ |j = \frac{3}{2}, m = \frac{1}{2}\rangle &= \frac{1}{\sqrt{3}}(|+, -, +\rangle + |-, +, +\rangle - |+, +, -\rangle) \\ |j = \frac{3}{2}, m = -\frac{1}{2}\rangle &= \frac{1}{\sqrt{3}}(|+, -, -\rangle + |-, +, -\rangle + |-, -, +\rangle) \\ |j = \frac{3}{2}, m = -\frac{3}{2}\rangle &= |-, -, -\rangle. \end{aligned} \quad (7.53)$$

Ladder 2

$$\begin{aligned} |j = \frac{1}{2}, m = \frac{1}{2}\rangle &= \frac{1}{\sqrt{2}}(|-, +, +\rangle - |+, -, +\rangle) \\ |j = \frac{1}{2}, m = \frac{-1}{2}\rangle &= \frac{2}{\sqrt{3}}(\frac{1}{2}|+, -, +\rangle + \frac{1}{2}|-, +, +\rangle - |+, +, -\rangle) \\ |j = \frac{1}{2}, m = -\frac{1}{2}\rangle &= \sqrt{\frac{2}{3}}(\frac{1}{2}|+, -, -\rangle + \frac{1}{2}|-, +, -\rangle - |-, -, +\rangle) \\ |j = \frac{1}{2}, m = -\frac{3}{2}\rangle &= \frac{1}{\sqrt{2}}(|-, +, -\rangle - |+, -, -\rangle). \end{aligned} \quad (7.54)$$

Example 7.4.

- Assume that an electron is in a state of orbital angular momentum $l = 2$. By use of the rules for angular momentum addition, construct the states with total angular momentum $j = \frac{5}{2}$ and z components $m = \frac{5}{2}, \frac{3}{2}$ as linear combinations of eigenstates of L_z and S_z .

The objective here is to determine states $|j = \frac{5}{2}, m = \frac{5}{2}\rangle$ and $|j = \frac{5}{2}, m = \frac{3}{2}\rangle$ in terms of $|m_1, m_2\rangle$. First, I need to determine the allowed values of j, j_1, j_2, m_1, m_2 . Since an electron is a spin half particle $j_2 = \frac{1}{2}$, and since $l = 2$

$$j_1 = l = 2 \quad \text{and} \quad j_2 = \frac{1}{2}. \quad (7.55)$$

Even though I only have to consider the case of $j = \frac{5}{2}$, I still need the allowed values of j in order to determine $m = -j, \dots, j$. The values of j are given by

$$|j_1 - j_2| \leq j \leq j_1 + j_2 \quad j = \frac{3}{2}, \frac{5}{2}, \quad (7.56)$$

which results in

$$\begin{aligned} m_{j=\frac{5}{2}} = -j, \dots, j &\Rightarrow m = -\frac{5}{2}, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \\ m_{j=\frac{3}{2}} = -j, \dots, j &\Rightarrow m = -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}. \end{aligned} \quad (7.57)$$

The allowed m_i are given from the relation

$$\begin{aligned} m_i = -j_i, \dots, j_i &\Rightarrow m_1 = -2, -1, 0, 1, 2, \\ &m_2 = -\frac{1}{2}, \frac{1}{2}. \end{aligned} \quad (7.58)$$

From which it is clear that the total number of states are 10.

Ladder 1: $j = \frac{5}{2}$ In the case of $j = \frac{5}{2}$ the highest state is $m = \frac{5}{2}$. This can only result from $m = m_1 + m_2 = 2 + \frac{1}{2}$. Thereby

$$|j = \frac{5}{2}, m = \frac{5}{2}\rangle = |m_1 = 2, m_2 = \frac{1}{2}\rangle. \quad (7.59)$$

This was the first state to be determined. The next, $|j = \frac{5}{2}, m = \frac{3}{2}\rangle$, is determined by lowering $|j = \frac{5}{2}, m = \frac{5}{2}\rangle$. The lowering operator is defined viz

$$J_- |j, m\rangle = \sqrt{(j+m)(j-m+1)} \hbar |j, m-1\rangle. \quad (7.60)$$

By using the lowering operator on the left hand side of equation 7.59 the state is lowered. However the right hand side also needs to be lowered. In order to do so, use that

$$J_- = J_-^{(1)} + J_-^{(2)}, \quad (7.61)$$

where $J_-^{(i)}$ are defined in analog to J_- . By using the lowering operators

$$\begin{aligned} J_- |j = \frac{5}{2}, m = \frac{5}{2}\rangle &= \sqrt{\left(\frac{5}{2} + \frac{5}{2}\right)\left(\frac{5}{2} - \frac{5}{2} + 1\right)} \hbar |j = \frac{5}{2}, m = \frac{3}{2}\rangle \\ &= \sqrt{5} \hbar |j = \frac{5}{2}, m = \frac{3}{2}\rangle \\ &= (J_{1-} + J_{2-}) |m_1 = 2, m_2 = \frac{1}{2}\rangle \\ &= J_{1-} |m_1 = 2, m_2 = \frac{1}{2}\rangle + J_{2-} |m_1 = 2, m_2 = \frac{1}{2}\rangle \\ &= \sqrt{(2+2)(2-2+1)} \hbar |m_1 = 1, m_2 = \frac{1}{2}\rangle \\ &\quad + \sqrt{\left(\frac{1}{2} + \frac{1}{2}\right)\left(\frac{1}{2} - \frac{1}{2} + 1\right)} \hbar |m_1 = 1, m_2 = -\frac{1}{2}\rangle \\ &= \sqrt{4} \hbar |m_1 = 1, m_2 = \frac{1}{2}\rangle + \hbar |m_1 = 2, m_2 = -\frac{1}{2}\rangle. \end{aligned} \quad (7.62)$$

From which it is clear that

$$|j = \frac{5}{2}; m = \frac{3}{2}\rangle = \sqrt{\frac{4}{5}} |m_1 = 1, m_2 = \frac{1}{2}\rangle + \frac{1}{\sqrt{5}} |m_1 = 2, m_2 = -\frac{1}{2}\rangle. \quad (7.63)$$

2. Find the normalized state with $j = \frac{3}{2}$ and $m = \frac{3}{2}$.

This corresponds to finding the highest state in the second ladder.

Ladder 2: $j = \frac{3}{2}, m = \frac{3}{2}$ can result from $|m_1 = 2, m_2 = -\frac{1}{2}\rangle$ or $|m_1 = 1, m_2 = \frac{1}{2}\rangle$. Thus, in general the state can be described as a linear combination between the two states

$$|j = \frac{3}{2}; m = \frac{3}{2}\rangle = c_1 |m_1 = 1, m_2 = \frac{1}{2}\rangle + c_2 |m_1 = 2, m_2 = -\frac{1}{2}\rangle. \quad (7.64)$$

To determine the coefficients I use that the coefficients are Clebsch-Gordan coefficients and can therefore be taken as real. Besides this, I require normalization and the different basis vectors, $|j, m\rangle$, should be orthogonal. From orthogonality I find

$$\langle j = \frac{5}{2}; m = \frac{3}{2} | j = \frac{3}{2}; m = \frac{3}{2} \rangle = c_1 \sqrt{\frac{4}{5}} + c_2 \sqrt{\frac{1}{5}} = 0 \Rightarrow c_1 = -\frac{c_2}{2}. \quad (7.65)$$

Combining this with normalization, $c_1^2 + c_2^2 = 1$, I find

$$c_1 = -\frac{1}{\sqrt{5}} \quad \text{and} \quad c_2 = \sqrt{\frac{4}{5}}. \quad (7.66)$$

Thereby

$$|j = \frac{3}{2}; m = \frac{3}{2}\rangle = -\frac{1}{\sqrt{5}} |m_1 = 1, m_2 = \frac{1}{2}\rangle + \sqrt{\frac{4}{5}} |m_1 = 2, m_2 = -\frac{1}{2}\rangle. \quad (7.67)$$

By applying the lowering operator to the states with $j = \frac{5}{2}, \frac{3}{2}$ all the 10 states can be found. The result is shown in figure 7.1.

	2,1/2	2,-1/2	1,1/2	1,-1/2	0,1/2	0,-1/2	-1,1/2	-1,-1/2	-2,1/2	-2,-1/2
j, m_j	5/2,5/2	1	0	0	0	0	0	0	0	0
	5/2,3/2	0	$1/\sqrt{5}$	$2/\sqrt{5}$	0	0	0	0	0	0
	5/2,1/2	0	0	0	$\sqrt{2}/\sqrt{5}$	$\sqrt{3}/\sqrt{5}$	0	0	0	0
	5/2,-1/2	0	0	0	0	0	$\sqrt{3}/\sqrt{5}$	$\sqrt{2}/\sqrt{5}$	0	0
	5/2,-3/2	0	0	0	0	0	0	$2/\sqrt{5}$	$1/\sqrt{5}$	0
	5/2,-5/2	0	0	0	0	0	0	0	0	1
	3/2,3/2	0	$2/\sqrt{5}$	$-1/\sqrt{5}$	0	0	0	0	0	0
	3/2,1/2	0	0	0	$\sqrt{3}/\sqrt{5}$	$-\sqrt{2}/\sqrt{5}$	0	0	0	0
	3/2,-1/2	0	0	0	0	$\sqrt{2}/\sqrt{5}$	$-\sqrt{3}/\sqrt{5}$	0	0	0
	3/2,-3/2	0	0	0	0	0	0	$1/\sqrt{5}$	$-2/\sqrt{5}$	0

Figure 7.1: The Clebsch-Gordan coefficients for an electron with $l = 2$.

3. Summarize your result in terms of Clebsch-Gordan coefficients.

The Clebsch-Gordan coefficients are the coefficients on the right hand side of $|j, m\rangle = \dots |m_1, m_2\rangle +$

... $|m_1, m_2\rangle$ These are given by

$$\begin{aligned}
 & \langle m_1 = 2, m_2 = \frac{1}{2} | j = \frac{5}{2}, m = \frac{5}{2} \rangle = 1, \\
 & \langle m_1 = 1, m_2 = \frac{1}{2} | j = \frac{5}{2}, m = \frac{3}{2} \rangle = \sqrt{\frac{4}{5}}, \\
 & \langle m_1 = 2, m_2 = -\frac{1}{2} | j = \frac{5}{2}, m = \frac{3}{2} \rangle = \sqrt{\frac{1}{5}}, \\
 & \langle m_1 = 1, m_2 = \frac{1}{2} | j = \frac{3}{2}, m = \frac{3}{2} \rangle = -\sqrt{\frac{1}{5}}, \\
 & \langle m_1 = 2, m_2 = -\frac{1}{2} | j = \frac{3}{2}, m = \frac{3}{2} \rangle = \sqrt{\frac{4}{5}}.
 \end{aligned} \tag{7.68}$$

8 | Approximation methods

8.1 Perturbation theory in QM

In the preceding chapters I have illustrated how to solve some simple quantum mechanical systems exactly. Most quantum mechanical systems, however, cannot be solved exactly, and so instead approximate methods are used. One such method is perturbation theory. In the words of Paul Dirac (Dirac, 1930, p.167)

“... Most quantum problems, however, cannot be solved exactly with the present resources of mathematics, as they lead to equations whose solutions cannot be expressed in finite terms with the help of the ordinary functions of analysis. For such problems one can often use a perturbation method. This consists in splitting up the Hamiltonian into two parts, one of which must be simple and the other small. The first part may then be considered as the Hamiltonian of a simplified or unperturbed system, which can be dealt with exactly, and the addition of the second will then require small corrections, of the nature of a perturbation, in the solution for the unperturbed system. The requirement that the first part shall be simple requires in practice that it shall not involve the time explicitly.”

That is in perturbation theory one assume that the Hamiltonian of the system under consideration can be described by a known Hamiltonian plus a relatively small perturbative Hamiltonian. The method of procedure for determining the energy levels and eigenstates of the Hamiltonian depends on whether or not the energy levels of the Hamiltonian are degenerate and whether or not the perturbative Hamiltonian depends on time. In this chapter I will review how to deal with the different cases.

Non-degenerate, time-independent perturbation theory

Non-degenerate, time-independent perturbation theory deals with the case where the perturbative Hamiltonian is time-independent. Since the non-perturbative Hamiltonian always is considered time independent, this means that the total Hamiltonian is considered time independent in this case. The total Hamiltonian is taken to be on the form

$$H = H_0 + \lambda H', \quad (8.1)$$

where $0 \leq \lambda \leq 1$ is a parameter, representing the strength of the perturbation, that is introduced for book-keeping purposes. H_0 is the unperturbed Hamiltonian. It is assumed that the eigenstates and eigenvalues of the unperturbed Hamiltonian is known

$$H_0 |n^0\rangle = E_n^{(0)} |n^0\rangle. \quad (8.2)$$

Since the total Hamiltonian is time independent, separation of variables can be used to separate the Schrödinger equation into a time-dependent and time-independent equation. Since I have already solved the time-dependent one, I will focus on the time-independent one here. This will be on the form

$$H |n\rangle = E_n |n\rangle. \quad (8.3)$$

The perturbation changes $|n^0\rangle$ only slightly, and therefore the new eigenstate can be expanded in terms of $|n^0\rangle$. In mathematical terms; the perturbation does not change the Hilbert space, so the eigenvectors of the unperturbed Hamiltonian can be used to expand the eigenvectors of the total Hamiltonian. Since

the perturbation is assumed only to change the state slightly I will describe a given eigenstate of the total Hamiltonian as the corresponding eigenstate of the unperturbed Hamiltonian plus some perturbative term

$$|n\rangle = |n^0\rangle + |\Delta n\rangle, \quad (8.4)$$

where the perturbative term can be expanded in terms of the eigenvectors of the unperturbed Hamiltonian (minus the eigenvector $|n^0\rangle$ since this is already used)

$$|\Delta n\rangle = \sum_{k \neq n} c_k |k^0\rangle. \quad (8.5)$$

Similarly the energy of a given eigenstate can be described as the energy of the corresponding eigenstate of the unperturbed Hamiltonian plus some perturbative term

$$E_n = E_n^{(0)} + \Delta E_n. \quad (8.6)$$

The aim of perturbation theory is to determine ΔE_n and $|\Delta n\rangle$ so the eigenstates and eigenvalues of the total Hamiltonian is determined. To this end I insert the expression for E_n and $|n\rangle$ into equation (8.3). Thereby

$$(H_0 + \lambda H')(|n^0\rangle + |\Delta n\rangle) = (E_n^{(0)} + \Delta E_n)(|n^0\rangle + |\Delta n\rangle). \quad (8.7)$$

I let the first terms on the LHS and RHS¹ cancel and act from the left with the bra $\langle n^0|$. Since $\langle n^0|n^0\rangle = \delta_{n^0 n'^0}$, and $|\Delta n\rangle$ is expanded in terms of $|k^0\rangle$ where $k \neq n$, $\langle n^0|\Delta n\rangle = 0$ and $\langle n^0|n^0\rangle = 1$. For the third term on the LHS of equation (8.7) I use that H_0 is Hermitian and I can therefore let it operate to the left onto its eigenstate and obtain the eigenvalue. Since the eigenvalue is just a number the bra is allowed to operator onto the ket, but since $\langle n^0|\Delta n\rangle = 0$ the term vanishes. On the RHS I can let the bra operate onto the kets immediately since the eigenvalues are just numbers and numbers commute with all operators. Therefore the two last terms cancel using again that $\langle n^0|\Delta n\rangle = 0$. Thus I am left with

$$\lambda \langle n^0|H'|n\rangle = \Delta E_n, \quad (8.8)$$

where I have used that $\lambda \langle n^0|H'|n^0\rangle + \lambda \langle n^0|H'|\Delta n\rangle = \lambda \langle n^0|H'|n\rangle$. The effect of the perturbation is to shift the energy level (originally E_n) by an amount ΔE_n which can be determined by equation (8.8). However, to calculate the shift in the energy the state $|n\rangle$ has to be determined. In reality it is expanded in powers of λ

$$\begin{aligned} |n\rangle &= \sum_i \lambda^i |n^i\rangle \\ &= |n^0\rangle + \lambda |n^1\rangle + \lambda^2 |n^2\rangle + \dots \\ &= |n^0\rangle + \lambda \sum_{k \neq n} c_k^{(1)} |k^0\rangle + \lambda^2 \sum_{k \neq n} c_k^{(2)} |k^0\rangle + \dots \end{aligned} \quad (8.9)$$

Each of the terms $\lambda^i |n^i\rangle$ are correction terms that add up to $|\Delta n\rangle$. Since the expansion of $|\Delta n\rangle$ on $|k^0\rangle$ was for a general vector that was not n^0 the same method applies for each of the vectors $|n^i\rangle$. By inserting this power series expansion into equation (8.8) the shift in energy can be written as

$$\begin{aligned} \Delta E_n &= \lambda \langle n^0|H'|n^0\rangle + \lambda^2 \langle n^0|H'|n^1\rangle + \lambda^3 \langle n^0|H'|n^2\rangle + \dots \\ &= \sum_i \lambda^i \langle n^0|H'|n^{(i-1)}\rangle \\ &= \sum_i \lambda^i E_n^i, \end{aligned} \quad (8.10)$$

¹LHS=Left Hand Side, RHS=Righ Hand Side.

where I have defined $\langle n^0 | H' | n^{(i-1)} \rangle = E_n^i$. Therefore also the energy change can be described in terms of a power series in the energy². By using the power series expansions of the energy and the state in equation (8.3) I find

$$\begin{aligned} & \left(H_0 + \lambda H' \right) \left(|n^0\rangle + \lambda \sum_{k \neq n} c_k^{(1)} |k^0\rangle + \lambda^2 \sum_{k \neq n} c_k^{(2)} |k^0\rangle + \dots \right) \\ &= \left(E_n^{(0)} + \lambda E_n^{(1)} + \dots \right) \left(|n^0\rangle + \lambda \sum_{k \neq n} c_k^{(1)} |k^0\rangle + \lambda^2 \sum_{k \neq n} c_k^{(2)} |k^0\rangle + \dots \right). \end{aligned} \quad (8.11)$$

I then compare the terms of equal order of λ

$$\begin{aligned} \lambda^0 : \quad H_0 |n^0\rangle &= E_n^{(0)} |n^0\rangle, \\ \lambda^1 : \quad \sum_{k \neq n} c_k^1 H_0 |k^0\rangle + H' |n^0\rangle &= E_n^{(0)} \sum_{k \neq n} c_k^{(1)} |k^0\rangle + E_n^{(1)} |n^0\rangle, \\ \lambda^2 : \quad \sum_{k \neq n} c_k^{(2)} H_0 |k^0\rangle + \sum_{k \neq n} c_k^{(1)} H' |k^0\rangle &= E_n^{(0)} \sum_{k \neq n} c_k^{(2)} |k^0\rangle + E_n^{(1)} \sum_{k \neq n} c_k^{(1)} |k^0\rangle + E_n^{(2)} |n^0\rangle. \end{aligned} \quad (8.12)$$

The 0'th order term just returns the unperturbed, TI Schroedinger equation - this is a good thing, as the system should to the 0'th order not be perturbed! For the first order term I act with $\langle n^0 |$ from the left to find the first order correction to the energy

$$\begin{aligned} \lambda^1 : \quad E_n^{(1)} &= \frac{\langle n^0 | H' | n^0 \rangle}{\langle n^0 | n^0 \rangle} \\ &= \int \int \langle n^0 | x' \rangle \langle x' | H' | x'' \rangle \langle x'' | n^0 \rangle d^3 x' d^3 x'' \\ &= \int |\langle n^0 | x' \rangle|^2 \langle x' | H' | x' \rangle d^3 x', \end{aligned} \quad (8.13)$$

where if the states are normalized $\langle n^0 | n^0 \rangle = 1$. For the second order term I do the same from which I find

$$\lambda^2 : \quad E_n^{(2)} = \sum_{k \neq n} c_k^{(1)} \frac{\langle n^0 | H' | k^0 \rangle}{\langle n^0 | n^0 \rangle} = \sum_{k \neq n} c_k^{(1)} \langle n^0 | H' | k^0 \rangle, \quad (8.14)$$

where I have assumed the states are normalized in the above equality (i.e. that $\langle n' | n \rangle = \delta_{n'n}$). c_k^1 is determined by acting with $\langle k^0 |$ from the left on equation (8.12). Hereby

$$\lambda^1 : \quad \langle k'^0 | H' | n^0 \rangle + c_{k'}^{(1)} E_{k'}^{(1)} = E_n^{(0)} c_{k'}^{(1)} \Rightarrow c_{k'}^{(1)} = \frac{\langle k'^0 | H' | n^0 \rangle}{E_n^{(0)} - E_{k'}^{(0)}}. \quad (8.15)$$

k' is just a dummy index, so I let $k' \Rightarrow k$ and use the above in the second order correction to the energy. Thereby

$$\lambda^2 : \quad E_n^{(2)} = \sum_{k \neq n} \frac{|\langle k^0 | H' | n^0 \rangle|^2}{E_n^{(0)} - E_k^{(0)}}. \quad (8.16)$$

Collecting the first and second order correction

$$\Delta E_n = \sum_i \lambda^i E_n^i = \lambda \langle n^0 | H' | n^0 \rangle + \lambda^2 \sum_{k \neq n} \frac{|\langle k^0 | H' | n^0 \rangle|^2}{E_n^{(0)} - E_k^{(0)}} + \dots \quad (8.17)$$

Likewise for the eigenstate

$$|n\rangle = \sum_i \lambda^i |n^i\rangle = |n^0\rangle + \lambda \sum_{k \neq n} \frac{\langle k^0 | H' | n^0 \rangle}{E_n^{(0)} - E_k^{(0)}} |k^0\rangle + \dots \quad (8.18)$$

²Note that the expansion of the energy can be written as $E_n = E_n^{(0)} + \lambda E_n^{(1)} + \dots$

In the end, which this is, the strength of the perturbation is set to unity (i.e. $\lambda = 1$). Thereby

$$\Delta E_n = \underbrace{\langle n^0 | H' | n^0 \rangle}_{\mathcal{O}(\lambda)} + \underbrace{\sum_{k \neq n} \frac{|\langle k^0 | H' | n^0 \rangle|^2}{E_n^{(0)} - E_k^{(0)}}}_{\mathcal{O}(\lambda^2)} + \dots, \quad (8.19)$$

$$|n\rangle = \underbrace{|n^0\rangle}_{\mathcal{O}(\lambda^0)} + \underbrace{\sum_{k \neq n} \frac{\langle k^0 | H' | n^0 \rangle}{E_n^{(0)} - E_k^{(0)}}}_{\mathcal{O}(\lambda)} |k^0\rangle + \dots \quad (8.20)$$

Even though the λ 's are gone, one must remember where they were because they describe the order of the perturbative term! By collecting higher orders of lambda and sandwiching from the left with a bra higher order corrections to the energy and state vector can be found. However, the first corrections provides a very good approximation in most cases and so higher order terms are not needed³⁴.

Degenerate, time-independent perturbation theory

In the case of degenerate energy levels the same equations cannot be used since the denominators (in the λ^2 and λ term respectively) in equation (8.19) and (8.20) blow up if $E_n^0 = E_k^0$. An infinity hints that the physics are wrong, and indeed they are in this case. The error comes from the eigenstate. From equation (8.20) it is clear that the correction to a given eigenstate is taken as a perturbation with respect to a state that refers to the energy level. However, in the case of degeneracy several eigenstates can be $|n^0\rangle$. To fix this I use that a linear combination of the degenerate states is orthogonal to the non-degenerate states. I denote the set of degenerate states as g and the individual degenerate eigenvectors as $|m^0\rangle$ ⁵. From these definitions I can write a degenerate version of equation (8.9) as follows

$$|n\rangle = \sum_{m \in g} x_m |m^0\rangle + \lambda \sum_{k \notin g} c_k^{(1)} |k^0\rangle + \lambda^2 \sum_{k \notin g} c_k^{(2)} |k^0\rangle + \dots \quad (8.21)$$

where x_m is just the set of expansion coefficients amongst the degenerate eigenstates. I use the same procedure as in the non-degenerate case and so I must also represent the energy by an expansion in powers of λ . Since I have fixed the issue (the degeneracy of $|n^0\rangle$) I can use the same expansion as in the non-degenerate case

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \dots \quad (8.22)$$

By using these in the TI Schrödinger equation (equation (8.3)) alongside the definition of the Hamiltonian from equation (8.1)

$$\begin{aligned} & \left(H_0 + \lambda H' \right) \left(\sum_{m \in g} x_m |m^0\rangle + \lambda \sum_{k \notin g} c_k^{(1)} |k^0\rangle + \lambda^2 \sum_{k \notin g} c_k^{(2)} |k^0\rangle + \dots \right) \\ &= \left(E_n^{(0)} + \lambda E_n^{(1)} + \lambda^2 E_n^{(2)} + \dots \right) \left(\sum_{m \in g} x_m |m^0\rangle + \lambda \sum_{k \notin g} c_k^{(1)} |k^0\rangle + \lambda^2 \sum_{k \notin g} c_k^{(2)} |k^0\rangle + \dots \right). \end{aligned} \quad (8.23)$$

I compare the terms of the same order in λ

$$\begin{aligned} \lambda^0 : \quad & \sum_{m \in g} x_m H_0 |m^0\rangle = E_n^{(0)} \sum_{m \in g} x_m |m^0\rangle, \\ \lambda^1 : \quad & \sum_{m \in g} x_m H' |m^0\rangle + \sum_{k \notin g} c_k^{(1)} E_k^{(0)} |k^0\rangle = E_n^{(0)} \sum_{k \notin g} c_k^{(1)} |k^0\rangle + E_n^{(1)} \sum_{m \in g} x_m |m^0\rangle, \\ \lambda^2 : \quad & \sum_{k \notin g} c_k^{(1)} H' |k^0\rangle + \sum_{k \notin g} c_k^{(2)} E_k^{(0)} |k^0\rangle = E_n^{(0)} \sum_{k \notin g} c_k^{(2)} |k^0\rangle + E_n^{(1)} \sum_{k \notin g} c_k^{(1)} |k^0\rangle + E_n^{(2)} \sum_{m \in g} x_m |m^0\rangle, \end{aligned} \quad (8.24)$$

³The expressions also gets increasingly more complicated. If the expressions were simpler perhaps they would have been incorporated. A funny side note: P. A. M. Dirac was almost religious in the belief that math describing physical phenomena should be beautiful!. He would dismiss a physical theory on the basis of the mathematics being "not beautiful". This has frustrated many physicists pitching ideas to him.

⁴Dirac is the inventor of TDPT - this is perhaps why it is well explained!

⁵So the different eigenvectors $|m^0\rangle$ are the elements in the set of degenerate eigenvectors; g .

where $H_0 |m^0\rangle = E_n^{(0)} |m^0\rangle$. By acting from the left with $\langle m'^0|$ I find⁶

$$\lambda^1 : \quad \sum_{m \in g} x_m \langle m'^0 | H' | m^0 \rangle = E_n^{(1)} x_{m'}. \quad (8.25)$$

This is a matrix equation on the same form as equation (5.45). Cf. Cramers rule there only exists non-trivial solutions if

$$\det(H' - E_n^{(1)} I) = 0 \Rightarrow \{E_n^{(1)}\}, \{|l\rangle\}. \quad (8.26)$$

Diagonalizing the matrix H' ⁷ results in $E_n^{(1)}$ and x_m .⁸ Since the number of eigenvalues match the dimensions of the matrix, there will be several, different, corrections to the degenerate energy level. This is to be understood as the degenerate energy level splitting into distinct energy levels; the perturbation splits the degenerate energy level into distinct energy levels. Since there can be degeneracy in the energy levels when

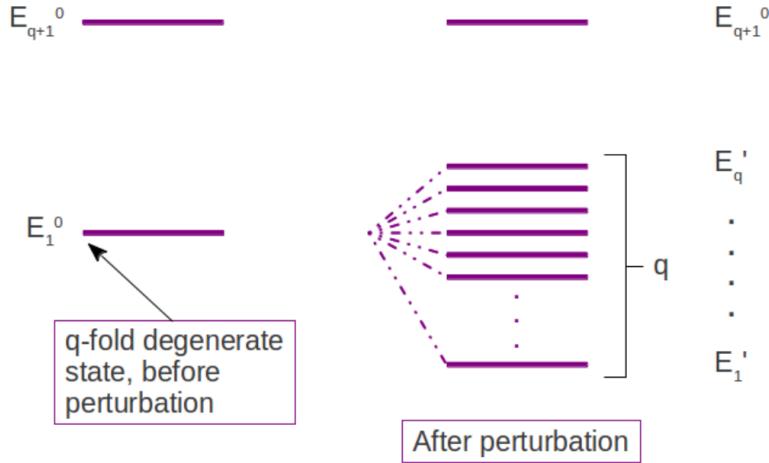


Figure 8.1: The splitting of the degenerate energy level.

diagonalizing a matrix, the diagonalization does not guarantee a complete lifting of the degeneracy. I will however limit myself to the case where this is so. If I lift the degeneracy I can use the same formulas, with the correction that the sum is only over non-degenerate energy eigenstates, as in the non-degenerate case to determine the higher order correction. Thereby I in the end find Collecting the first and second order correction

$$\Delta E_n = \underbrace{E_n^{(1)}}_{\mathcal{O}(\lambda)} + \underbrace{\sum_{k \notin g} \frac{|\langle k^0 | H' | n^0 \rangle|^2}{E_n^{(0)} - E_k^{(0)}}}_{\mathcal{O}(\lambda^2)} + \dots, \quad (8.27)$$

where $E_n^{(1)}$ are the eigenvalues of the degenerate perturbation matrix and can take different values. The second order correction is the same for each of these different values of $E_n^{(1)}$. For the eigenstate

$$|n\rangle = \underbrace{|l\rangle}_{\mathcal{O}(\lambda^0)} + \underbrace{\sum_{k \notin g} \frac{\langle k^0 | H' | l \rangle}{E_l^0 - E_k^0} |k^0\rangle}_{\mathcal{O}(\lambda)} + \dots, \quad (8.28)$$

where $|l\rangle$ are any of the eigenvectors obtained from diagonalizing the degenerate sub-matrix of the perturbative Hamiltonian. By comparison between equation (8.19) versus (8.27) and equation (8.20) versus (8.28) it is

⁶Nothing exciting happens for the zeroth order term.

⁷In practice the matrix to be diagonalized is the matrix containing the entire extra term, that is including λ .

⁸The different sets of x_m in turn results in the eigenvectors, $|l\rangle$, corresponding to the eigenvalues, $E_n^{(1)}$.

clear that the corrections in degenerate perturbation theory are reminiscent of the ones from non-degenerate perturbation theory. The procedure in degenerate perturbation theory is to find the basis in which the sub-matrix of the perturbative Hamiltonian that is degenerate is diagonal. If this basis lifts the degeneracy completely the non-degenerate perturbation theory can be applied to the eigenstates and eigenvalues of the degenerate perturbation matrix.

Time-dependent perturbation theory

Time-dependent perturbation theory is an approximate method for solving the Schrödinger equation in the case where $\frac{\partial H}{\partial t} \neq 0$ - that is in the case where separation of variables does not apply. The Hamiltonian is taken to consist of two additive parts; one simple part (H_0) for which the solution is known and one smaller part (H'). The simple part must contain any time dependency in order to be simple ($\frac{\partial H_0}{\partial t} = 0$), so the time dependency of the Hamiltonian must lie in the small part, i.e. $H' = H'(t)$. The Hamiltonian is therefore given by

$$H = H_0 + H'(t). \quad (8.29)$$

The system is taken to be in an eigenstate of H_0 at $t = 0s$. In some finite interval of time the system is then perturbed, in this case the system is a nucleus, and the perturbation is scattering with the WIMP. After the perturbation, the system is again characterized by the simple Hamiltonian, and the system is possibly in the same or a different eigenstate of the simple Hamiltonian. The scenario is shown in figure 8.2. While

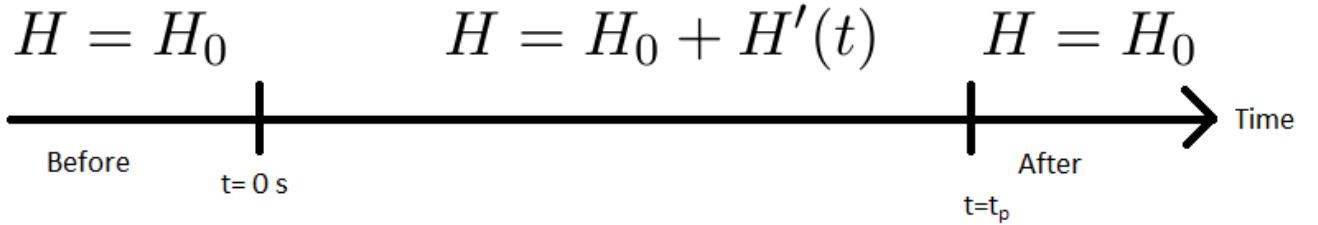


Figure 8.2: The development of the Hamiltonian as a function of time.

the perturbation is turned on ($0 \leq t \leq t_p$) the system is described by the Schrödinger equation on the form

$$i\hbar \frac{\partial |\alpha\rangle}{\partial t} = (H_0 + H'(t)) |\alpha\rangle, \quad (8.30)$$

where $|\alpha\rangle$ is the state-vector. Equation (8.30) is solved by assuming that the eigenvectors and eigenvalues of the unperturbed system is known. Since the unperturbed Hamiltonian is time independent the Schrödinger equation can in this case be solved via separation of variables. The time-dependent Schrödinger equation is trivially solved while the time-independent Schrödinger equation is given by

$$H_0 |n_0\rangle = E_{n_0} |n_0\rangle, \quad (8.31)$$

where E_{n_0} and $\{|n_0\rangle\}$ is taken to be known. Since the eigenvectors of H_0 make up a complete set, spanning the desired vector space, they are used to develop the state-vector of the system

$$|\alpha\rangle = \sum_{n_0} c_{n_0}(t) e^{-\frac{iE_{n_0}}{\hbar} t} |n_0\rangle, \quad (8.32)$$

where, compared to the case where the Hamiltonian is time-independent, the coefficients of the expansion are now time-dependent. The probability of the system ending up in a given state is given by the squared magnitude of the coefficients $c_{n_0}(t)$. To solve equation (8.30) equation (8.32) is inserted

$$i\hbar \sum_{n_0} \frac{\partial c_{n_0}(t)}{\partial t} e^{-\frac{iE_{n_0}}{\hbar} t} |n_0\rangle - i\hbar \sum_{n_0} c_{n_0}(t) \frac{iE_{n_0}}{\hbar} e^{-\frac{iE_{n_0}}{\hbar} t} |n_0\rangle = \sum_{n_0} (H_0 + H'(t)) |n_0\rangle c_{n_0}(t) e^{-\frac{iE_{n_0}}{\hbar} t}. \quad (8.33)$$

Equation (8.33) is multiplied from the left with a eigenbra of H_0 ; $\langle m_0 |$

$$\begin{aligned} i\hbar \sum_{n_0} \frac{\partial c_{n_0}(t)}{\partial t} e^{-\frac{iE_{n_0}}{\hbar}t} \langle m_0 | n_0 \rangle + \sum_{n_0} c_{n_0}(t) E_{n_0} e^{-\frac{iE_{n_0}}{\hbar}t} \langle m_0 | n_0 \rangle \\ = \sum_{n_0} \langle m_0 | (H_0 + H'(t)) | n_0 \rangle c_{n_0}(t) e^{-\frac{iE_{n_0}}{\hbar}t}. \end{aligned} \quad (8.34)$$

By using $\langle m_0 | n_0 \rangle = \delta_{m_0 n_0}$ the sums on the left hand side vanishes. Furthermore H_0 operating to the right on $|n_0\rangle$ results in $E_{n_0}|n_0\rangle$ since $|n_0\rangle$ is an eigenstate of H_0 . Since E_{n_0} is just a number, it can be taken outside the inner product, and again the orthogonality property can be used to kill the sum. In the end

$$i\hbar \frac{\partial c_{m_0}(t)}{\partial t} e^{-\frac{iE_{m_0}}{\hbar}t} + c_{m_0}(t) E_{m_0} e^{-\frac{iE_{m_0}}{\hbar}t} = c_{m_0}(t) E_{m_0} e^{-\frac{iE_{m_0}}{\hbar}t} + \sum_{n_0} \langle m_0 | H'(t) | n_0 \rangle c_{n_0}(t) e^{-\frac{iE_{n_0}}{\hbar}t}. \quad (8.35)$$

Two of the terms cancel out. By moving the exponential from the left hand side (LHS) to the right hand side (RHS)

$$i\hbar \frac{\partial c_{m_0}(t)}{\partial t} = \sum_{n_0} \langle m_0 | H'(t) | n_0 \rangle c_{n_0}(t) e^{i\omega_{m_0 n_0} t}, \quad (8.36)$$

where $\omega_{m_0 n_0} = \frac{E_{m_0} - E_{n_0}}{\hbar}$. A general solution to equation (8.36) is difficult, but in the case where $H' \ll H_0$ perturbation theory is applicable, and an approximative solution can be found. A dimensionless, continuous, real parameter, $\lambda \in [0, 1]$, representing the strength of the perturbation, is defined by

$$H'(t) \Rightarrow \lambda H'(t). \quad (8.37)$$

λ is "dialed" from 0 to 1 in a smooth, continuous way as the state of the system goes from $|n_0\rangle$ to $|\alpha\rangle$. The parameter makes sure that there is a continuous transition between the two states of the system. Next up the expansion coefficients, $c_{m_0}(t)$ and $c_{n_0}(t)$, are expanded in powers of λ

$$c_{m_0}(t) = c_{m_0}^{(0)}(t) + \lambda c_{m_0}^{(1)}(t) + \lambda^2 c_{m_0}^{(2)}(t) + \dots, \quad (8.38)$$

$$c_{n_0}(t) = c_{n_0}^{(0)}(t) + \lambda c_{n_0}^{(1)}(t) + \lambda^2 c_{n_0}^{(2)}(t) + \dots, \quad (8.39)$$

where the superscript "(0)" denotes unperturbed quantities, and higher numbers identify corrections of higher orders. If the method is to be of practical interest, a good approximation must be obtained by taking only one or two terms into account. Equation (8.38) and (8.39) is inserted in equation (8.36)

$$\begin{aligned} i\hbar \left(\frac{\partial c_{m_0}^0(t)}{\partial t} + \lambda \frac{\partial c_{m_0}^1(t)}{\partial t} + \lambda^2 \frac{\partial c_{m_0}^2(t)}{\partial t} + \dots \right) \\ = \sum_{n_0} (c_{n_0}^0(t) + \lambda c_{n_0}^1(t) + \lambda^2 c_{n_0}^2(t) + \dots) \lambda e^{i\omega_{m_0 n_0} t} \langle m_0 | H'(t) | n_0 \rangle, \end{aligned} \quad (8.40)$$

where the λ outside the parenthesis on the right hand side stems from $H'(t)$. Comparing powers of λ

$$\begin{aligned} \lambda^0 : & \quad i\hbar \frac{\partial c_{m_0}^{(0)}(t)}{\partial t} = 0 \Rightarrow c_{m_0}^{(0)}(t) = \text{const}, \\ \lambda^1 : & \quad i\hbar \frac{\partial c_{m_0}^{(1)}(t)}{\partial t} = \sum_{n_0} c_{n_0}^{(0)}(t) e^{i\omega_{m_0 n_0} t} \langle m_0 | H'(t) | n_0 \rangle, \\ \lambda^2 : & \quad i\hbar \frac{\partial c_{m_0}^{(2)}(t)}{\partial t} = \sum_{n_0} c_{n_0}^{(1)}(t) e^{i\omega_{m_0 n_0} t} \langle m_0 | H'(t) | n_0 \rangle. \end{aligned} \quad (8.41)$$

It is clear that the coefficients on the LHS and RHS not are of the same power - there is a recursion relation. The coefficient $c_{m_0}^{(1)}(t)$ describes the process in which a single photon is absorbed or emitted by an atom.

Higher order coefficients correspond to processes incorporating more photons (eg. $c_{m_0}^{(2)}(t)$ corresponds to a two photon process). Because the relation between the coefficients is recursive the lower order coefficients must be determined in order to determine a higher order coefficient. Therefore, to get started, $c_{m_0}^0(t)$ is determined from an analysis of the initial conditions. The system is taken to be in an eigenstate of H_0 denoted by $|a_0\rangle$ at $t = 0$ - where i is just a random label. Since the system at $t = 0$ is in an eigenstate of H_0 , the only non-zero coefficient must be the one corresponding to $|a_0\rangle$, i.e.

$$t = 0 : \quad c_{a_0} = 1 \quad \wedge \quad c_{i \neq a_0} = 0. \quad (8.42)$$

The coefficients of 0'th order, $c_{n_0}^{(0)}(t)$, refers to the initial, unperturbed state. The only coefficient differing from 0 is $c_{a_0}^{(0)}(t) = 1$. Thereby the sum in the expression for the first order coefficient can be removed and, by denoting the final state $\langle b_0 |$

$$i\hbar \frac{\partial c_{b_0}^{(1)}(t)}{\partial t} = e^{i\omega_{b_0 a_0} t} \langle b_0 | H'(t) | a_0 \rangle. \quad (8.43)$$

Moving $i\hbar$ to the RHS, and integrating

$$c_{b_0}^{(1)}(t) = \frac{1}{i\hbar} \int_0^t \langle b_0 | H'(t') | a_0 \rangle e^{i\omega_{b_0 a_0} t'} dt', \quad (8.44)$$

where often t_p is denoted by t . Using equation (8.44) in equation (8.41)

$$\begin{aligned} c_{b_0}^{(2)}(t) &= \frac{1}{i\hbar} \sum_{n_0} \int_0^t c_{n_0}^{(1)}(t'') e^{i\omega_{b_0 n_0} t''} \langle b_0 | H'(t'') | n_0 \rangle dt' \\ &= -\frac{1}{\hbar^2} \sum_{n_0} \int_0^t \langle b_0 | H'(t'') | n_0 \rangle e^{i\omega_{b_0 n_0} t''} \int_0^{t''} \langle n_0 | H'(t') | a_0 \rangle e^{i\omega_{n_0 a_0} t'} dt' dt''. \end{aligned} \quad (8.45)$$

The transition probability, via a general process, is given by

$$\text{Transition probability from } |a_0\rangle \text{ to } |b_0\rangle = P_{a_0 \Rightarrow b_0} = |c_{b_0}^{(1)}(t) + c_{b_0}^{(2)}(t) + \dots|^2. \quad (8.46)$$

The transition probability for an i -photon process, eg. single photon or tw-photon, is given by

$$P_{a_0 \Rightarrow b_0}^{(i)} = |c_{b_0}^{(i)}(t)|^2. \quad (8.47)$$

The transition probability depends on the nature of $H'(t)$. Calculating the transition probability concludes time-dependent perturbation theory(TDPT). Since the system state before perturbation is known, what is of interest is what happens to the system during the perturbation - whether or not the system jumps into another state. This probability depends on the size of the perturbation, and if the perturbation is small, the probability can be calculated from equation (8.46).

8.2 The WKB approximation

The WKB method is a technique for obtaining approximate solutions to the time-independent Schrödinger equation ($\frac{\partial H}{\partial t} = 0$) in one dimension⁹. It is particularly useful in calculating bound state energies and tunneling rates through a potential barrier, and is for this reason used in astrophysics when dealing with the tunneling probability relevant for fusion processes.

I begin with the one-dimensional Schrödinger equation in the position basis

$$-\frac{\hbar^2}{2M} \frac{\partial^2 \psi(x)}{\partial x^2} + V(x)\psi(x) = E\psi(x). \quad (8.48)$$

⁹The same technique can be applied to many other differential equations.

This equation can be rewritten as

$$\frac{\partial^2 \psi(x)}{\partial x^2} = -\frac{2M(E - V(x))}{\hbar^2} \psi(x) = -\frac{p^2}{\hbar^2} \psi(x), \quad (8.49)$$

where $p = \sqrt{2M(E - V(x))}$ is the classical momentum of a particle moving in a one-dimensional potential (Griffiths, 2004, p.346), $V(x)$, with energy E , provided that $E \geq V(x)$ for all x . where the wave function in general is a complex number, and therefore can be written as

$$\psi(x) = A(x)e^{i\phi(x)}, \quad (8.50)$$

where $\phi(x)$ is the phase of the complex number. By this notation equation (8.49) becomes

$$\begin{aligned} & \left(\frac{\partial^2 A(x)}{\partial x^2} + 2i \frac{\partial A(x)}{\partial x} \frac{\partial \phi(x)}{\partial x} + iA(x) \frac{\partial^2 \psi(x)}{\partial x^2} - A(x) \left(\frac{\partial \phi(x)}{\partial x} \right)^2 \right) e^{i\phi(x)} \\ &= -\frac{p^2}{\hbar^2} A(x) e^{i\phi(x)}. \end{aligned} \quad (8.51)$$

This is equivalent to two real equations, one for the real part and one for the imaginary part. For the real part

$$\frac{\partial^2 A(x)}{\partial x^2} = A(x) \left(\left(\frac{\partial \phi(x)}{\partial x} \right)^2 - \frac{p^2}{\hbar^2} \right). \quad (8.52)$$

For the imaginary part

$$\frac{\partial}{\partial x} \left(A(x)^2 \frac{\partial \phi(x)}{\partial x} \right) = 0. \quad (8.53)$$

Equation (8.52) and (8.53) are equivalent to the time-independent Schroedinger equation. Equation (8.53) is solved by integration. Denoting the integration constant C^2 , I find

$$A(x) = \frac{C}{\sqrt{\frac{\partial \phi(x)}{\partial x}}}, \quad (8.54)$$

where C is a real constant. Equation (8.52) cannot be solved in general, so an approximation is introduced; *it is assumed that the amplitude varies slowly, such that $\frac{\partial^2 A(x)}{\partial x^2} \approx 0$, and $\frac{\partial^2 A(x)}{A(x)} \approx 0$.* Thus, from equation (8.52) I find

$$\frac{\partial \phi(x)}{\partial x} \approx \pm \frac{p}{\hbar}, \quad (8.55)$$

which is solved by integration such that

$$\phi(x) \approx \pm \frac{1}{\hbar} \int p(x) dx. \quad (8.56)$$

Any constants from this integral may be absorbed into C , and C may therefore be complex. Thus, I find the approximate solution

$$\psi(x) \simeq \frac{C}{\sqrt{p(x)}} e^{\pm \frac{i}{\hbar} \int p(x) dx}. \quad (8.57)$$

The general, approximate solution will be a linear combination of the plus and minus solution. From equation (8.57) it is clear that the probability of finding the particle at position x is given by

$$|\psi(x)|^2 \simeq \frac{|C|^2}{p(x)}. \quad (8.58)$$

Equation (8.58) reveals that the probability of finding the particle at position x is inversely proportional to the classical momentum, and thereby the particle's velocity at that point. This makes sense since if the particle has a large velocity at one point, it will not spend as much time there, and the probability of finding the particle there will be smaller.

Example 8.1.

A simple harmonic oscillator (in one dimension) is subjected to a perturbation

$$H' = bx, \quad (8.59)$$

where b is a real constant.

1. Calculate the energy shift of the ground state to the lowest non-vanishing order.

Since the energy levels of H_0 (the SHO) is *not* degenerate, and the perturbation does not depend on time, I use TIPT in the non-degenerate version. I use equation (8.19) and evaluate from the "bottom". The first order correction to the energy

$$\Delta E_0^{(1)} = b \langle 0^0 | x | 0^0 \rangle = b \int_{-\infty}^{\infty} |\psi_0(x)|^2 x dx = 0, \quad (8.60)$$

where I used that $\psi_n(x) = N_n H_n(\sqrt{\frac{\omega M}{\hbar}}x) e^{-\frac{\omega M x^2}{2\hbar}}$. The second order correction to the energy

$$\Delta E_0^{(2)} = b^2 \sum_{k \neq 0} \frac{|\langle k^0 | x^2 | 0^0 \rangle|^2}{E_0^0 - E_k^0}. \quad (8.61)$$

In this case it is not so straight forward since there is an infinite number of energy levels. However, by using that $\langle n' | x | n \rangle = \sqrt{\frac{\hbar}{2M\omega}} (\sqrt{n} \delta_{n',n-1} + \sqrt{n+1} \delta_{n',n+1})$ I take

$$\langle k^0 | x | 0 \rangle = \sqrt{\frac{\hbar}{2M\omega}} (\sqrt{n} \delta_{k,0-1} + \sqrt{0+1} \delta_{k,0+1}) = \sqrt{\frac{\hbar}{2M\omega}} \delta_{k,1}. \quad (8.62)$$

By using this the second order correction can be found to be

$$\Delta E_0^{(2)} = b^2 \sum_{k \neq 0} \frac{|\langle 1^0 | x^2 | 0^0 \rangle|^2}{E_0^0 - E_1^0} = -\frac{b^2}{2M\omega^2}, \quad (8.63)$$

where I used that $\Delta E^0 = \hbar\omega = -(E_0^0 - E_1^0)$. So, the energy shift to the lowest non-vanishing order (second) is given by

$$\begin{aligned} \Delta E_0 &= \langle 0^0 | H' | 0^0 \rangle + \sum_{k \neq 0} \frac{|\langle k^0 | H' | 0^0 \rangle|^2}{E_0^0 - E_k^0} + \dots \\ &= -\frac{b^2}{2M\omega^2} + \dots \end{aligned} \quad (8.64)$$

2. Solve this problem exactly and compare with your previous result.

The TISE¹⁰ in this case is given by

$$-\frac{\hbar^2}{2M} \frac{\partial^2 \langle x | \alpha \rangle}{\partial x^2} + \frac{1}{2} M \omega^2 x^2 \langle x | \alpha \rangle + bx \langle x | \alpha \rangle = E \langle x | \alpha \rangle. \quad (8.65)$$

This is a second order, non-homogeneous, non-linear, partial differential equation, and as such not an easy equation to solve¹¹. Therefore I search for a change of variables so that the equation takes a

¹⁰TI= Time Independent, SE=Schroedinger Equation.

¹¹If I were to solve it I would use a generalized power series approach.

familiar form - preferably the form of the simple harmonic oscillator! Since we have a terms with x^2 and one with x it looks factorisable. Indeed I find

$$\frac{1}{2}M\omega^2x^2 + bx = \frac{1}{2}M\omega^2\left(x + \frac{b}{M\omega^2}\right)^2 - \left(\frac{b}{M\omega^2}\right)^2. \quad (8.66)$$

Inspired by this I defined $x' = x + \frac{b}{M\omega^2}$ as a new variable. Since the derivative of x' with respect to x is 1 I can write the TISE in the new variable as

$$-\frac{\hbar^2}{2M}\frac{\partial^2 \langle x|\alpha \rangle}{\partial x'^2} + \frac{1}{2}M\omega^2\left(x'^2 - \left(\frac{b}{M\omega^2}\right)^2\right)\langle x|\alpha \rangle = E\langle x|\alpha \rangle, \quad (8.67)$$

which can be rewritten as

$$-\frac{\hbar^2}{2M}\frac{\partial^2 \langle x|\alpha \rangle}{\partial x'^2} + \frac{1}{2}M\omega^2x'^2\langle x|\alpha \rangle = \left(E + \frac{b^2}{2M\omega^2}\right)\langle x|\alpha \rangle. \quad (8.68)$$

This is the same form as the simple harmonic oscillator with the new eigenvalue

$$E' = E + \frac{b^2}{2M\omega^2}. \quad (8.69)$$

In comparison with before

$$\Delta E_0 = E - E' = -\frac{b^2}{2M\omega^2}, \quad (8.70)$$

which is exactly the same result as obtained from TI perturbation theory.

Example 8.2.

Consider the Hamiltonian

$$H_0 = \frac{p_x^2}{2M} + \frac{1}{2}M\omega^2x^2. \quad (8.71)$$

This Hamiltonian is subjected to a perturbation on the form

$$H' = \frac{1}{2}\varepsilon M\omega^2x^2. \quad (8.72)$$

What is the correction to the GS energy to second order?

Since the Harmonic oscillator (in 1D) is not degenerate I will use equation (8.19)

$$\Delta E_0 = \langle 0^0 | H' | 0^0 \rangle + \sum_{k \neq n} \frac{|\langle k^0 | H' | 0^0 \rangle|^2}{E_0^{(0)} - E_k^{(0)}} + \dots \quad (8.73)$$

The first order correction

$$\begin{aligned} \Delta E_0^{(1)} &= \langle 0^0 | H' | 0^0 \rangle \\ &= \frac{1}{2}\varepsilon M\omega^2 \langle 0^0 | x^2 | 0^0 \rangle \\ &= \frac{1}{2}\varepsilon M\omega^2 \frac{\hbar}{2M\omega} \langle 0^0 | (a^\dagger + a)^2 | 0^0 \rangle \\ &= \frac{\varepsilon\hbar\omega}{4} \left(\langle 0^0 | a^\dagger a | 0^0 \rangle + \langle 0^0 | a^\dagger a^\dagger | 0^0 \rangle + \langle 0^0 | aa^\dagger | 0^0 \rangle + \langle 0^0 | aa | 0^0 \rangle \right) \\ &= \frac{\varepsilon\hbar\omega}{4} \langle 0^0 | aa^\dagger | 0^0 \rangle = \frac{\varepsilon\hbar\omega}{4}. \end{aligned} \quad (8.74)$$

The second order correction

$$\Delta E_0^{(2)} = \sum_{k \neq n} \frac{|\langle k^0 | H' | 0^0 \rangle|^2}{E_0^{(0)} - E_k^{(0)}}. \quad (8.75)$$

To evaluate this the matrix elements need to be evaluated. In analog with the first order correction

$$\begin{aligned} \langle k^0 | H' | 0^0 \rangle &= \frac{\varepsilon \hbar \omega}{4} \left(\langle k^0 | a^\dagger a | 0^0 \rangle + \langle k^0 | a^\dagger a^\dagger | 0^0 \rangle + \langle k^0 | a a^\dagger | 0^0 \rangle + \langle k^0 | a a | 0^0 \rangle \right) \\ &= \frac{\varepsilon \hbar \omega}{4} \left(\langle k^0 | a^\dagger a^\dagger | 0^0 \rangle \delta_{k,2} + \langle k^0 | a a^\dagger | 0^0 \rangle \delta_{k,0} \right) \\ &= \frac{\varepsilon \hbar \omega}{4} \left(\sqrt{2} \delta_{k,2} + \delta_{k,0} \right). \end{aligned} \quad (8.76)$$

Since the sum is only over $k \neq n$ the $k = 0$ is not included in the second order correction

$$\Delta E_0^{(2)} = \left(\frac{\sqrt{2} \varepsilon \hbar \omega}{4} \right)^2 \frac{1}{\frac{1}{2} \hbar \omega - \frac{5}{2} \hbar \omega} = -\frac{\varepsilon^2 \hbar \omega}{16}. \quad (8.77)$$

Example 8.3.

Consider a particle in a two-dimensional potential

$$V_0 = \begin{cases} 0 & \text{for } 0 \leq x \leq L, 0 \leq y \leq L \\ \infty & \text{for Else} \end{cases}. \quad (8.78)$$

1. Find the eigenfunctions for the ground state and the first excited state.

This is identical to the free particle in a box in two dimensions. Taking only the TI part of $\langle x | \alpha \rangle = \psi(x, y, z, t)_{n_x n_y n_z} = B \sin\left(\frac{n_x \pi}{L} x\right) \sin\left(\frac{n_y \pi}{L} y\right) \sin\left(\frac{n_z \pi}{L} z\right) e^{-\frac{i E n_x n_y n_z}{\hbar} t}$ I find

$$\langle \vec{x} | n \rangle = \psi(x, y)_{n_x, n_y} = \frac{2}{L} \sin\left(\frac{n_x \pi}{L} x\right) \sin\left(\frac{n_y \pi}{L} y\right), \quad (8.79)$$

from which

$$\begin{aligned} \psi(x, y)_{1,1} &= \frac{2}{L} \sin\left(\frac{\pi}{L} x\right) \sin\left(\frac{\pi}{L} y\right), \\ \psi(x, y)_{2,1} &= \frac{2}{L} \sin\left(\frac{2\pi}{L} x\right) \sin\left(\frac{\pi}{L} y\right), \\ \psi(x, y)_{1,2} &= \frac{2}{L} \sin\left(\frac{\pi}{L} x\right) \sin\left(\frac{2\pi}{L} y\right). \end{aligned} \quad (8.80)$$

2. Now a perturbative term is added. The perturbative terms is given by

$$H' = \begin{cases} \lambda x y & \text{for } 0 \leq x \leq L, 0 \leq y \leq L \\ 0 & \text{for Else} \end{cases}. \quad (8.81)$$

Obtain the zeroth order energy eigenfunctions and the first order energy shifts for the ground state and the first excited state.

The GS is non-degenerate so non-degenerate perturbation theory can be used here. The first excited states are degenerate so here degenerate perturbation theory must be used (the perturbative Hamiltonian does not depend on time so TDPT is not used here).

For the GS the zeroth order energy eigenfunction is given by equation (8.80). The first order shift to the energy is given by

$$\begin{aligned}\Delta E_{1,1}^{(1)} &= \langle 1^0 | H' | 1^0 \rangle \\ &= \frac{4\lambda}{L^2} \int_0^L \int_0^L \sin\left(\frac{\pi}{L}x\right)^2 \sin\left(\frac{\pi}{L}y\right)^2 xy dxdy \\ &= \frac{L^2\lambda}{4}.\end{aligned}\quad (8.82)$$

For the first excited state there is degeneracy, and so degenerate perturbation theory will be used instead of non-degenerate perturbation theory. I construct the matrix $\langle n' | H' | n \rangle$ where $n, n' = 2, 1; 1, 2^{12}$. That is

$$\begin{aligned}H' &\doteq \begin{bmatrix} \langle 1, 2 | H' | 1, 2 \rangle & \langle 1, 2 | H' | 2, 1 \rangle \\ \langle 2, 1 | H' | 1, 2 \rangle & \langle 2, 1 | H' | 2, 1 \rangle \end{bmatrix} \\ &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \\ &= \frac{\lambda L^2}{4} \begin{bmatrix} 1 & \frac{1024}{81\pi^4} \\ \frac{1024}{81\pi^4} & 1 \end{bmatrix},\end{aligned}\quad (8.83)$$

where

$$\begin{aligned}A &\equiv \frac{4\lambda}{L^2} \int_0^L \int_0^L \sin\left(\frac{\pi}{L}x\right)^2 \sin\left(\frac{2\pi}{L}y\right)^2 dxdy, \\ B &\equiv \frac{4\lambda}{L^2} \int_0^L \int_0^L \sin\left(\frac{\pi}{L}x\right) \sin\left(\frac{2\pi}{L}y\right) \sin\left(\frac{2\pi}{L}x\right) \sin\left(\frac{\pi}{L}y\right) dxdy, \\ C &\equiv \frac{4\lambda}{L^2} \int_0^L \int_0^L \sin\left(\frac{\pi}{L}x\right) \sin\left(\frac{2\pi}{L}y\right) \sin\left(\frac{2\pi}{L}x\right) \sin\left(\frac{\pi}{L}y\right) dxdy, \\ D &\equiv \frac{4\lambda}{L^2} \int_0^L \int_0^L \sin\left(\frac{2\pi}{L}x\right)^2 \sin\left(\frac{\pi}{L}y\right)^2 dxdy.\end{aligned}\quad (8.84)$$

The eigenvalues of H' are the first order corrections to the energy. That is

$$\Delta E_{2,1}^{(1)} = \frac{\lambda^2 L^2}{324\pi^4} (81\pi^4 \pm 1024).\quad (8.85)$$

As is apparent from the above; the energy level of the first excited state is, to first order, split into two distinct energy states. Hence the degeneracy is lifted!

The eigenvectors of H' corresponds to the zeroth order energy eigenvectors. Denoting the two energy levels the first excited energy level has split into by $E_{2,i}$ where $i = 1, 2$, the corresponding eigenvectors can be found to be

$$|E_{2,1}\rangle = \frac{1}{\sqrt{2}}(|1, 2\rangle - |2, 1\rangle),\quad (8.86)$$

$$|E_{2,2}\rangle = \frac{1}{\sqrt{2}}(|1, 2\rangle + |2, 1\rangle).\quad (8.87)$$

Example 8.4.

Consider a spin-less particle in a two-dimensional infinite square well

$$V_0 = \begin{cases} 0 & \text{for } 0 \leq x \leq a, 0 \leq y \leq a \\ \infty & \text{for Else} \end{cases}.\quad (8.88)$$

¹²That is a matrix of the degenerate eigenfunctions!

1. What are the energy eigenvalues for the three lowest states? Is there any degeneracy?

Inside the box the particle is treated as a free particle in two dimensions. The energy is just the sum of two free particles in a 1D box

$$E_{n_x, n_y}^{(0)} = \frac{\hbar^2 \pi^2}{2ML^2} (n_x^2 + n_y^2), \quad (8.89)$$

where $n_i = 1, 2, 3, \dots$. Note that $n_i = 0$ is not a possibility since this value would make the wave-function, which is a product of the two wave-functions of a 1D free particle in a box and $\psi \propto \sin(n_x)\sin(n_y)$, zero. Therefore the GS is given by

$$E_{1,1}^{(0)} = \frac{\hbar^2 \pi^2}{ML^2}. \quad (8.90)$$

There is no degeneracy here. The following energy states

$$E_{2,1}^{(0)} = E_{1,2}^{(0)} = \frac{5\hbar^2 \pi^2}{2ML^2}, \quad (8.91)$$

where there is two-fold degeneracy.

2. Now a perturbative potential is added

$$H' = \lambda xy. \quad (8.92)$$

- (a) Is the energy shift due to the perturbation linear or quadratic in λ for each of the three states?

Cf. equation (8.82) and (8.85) $\Delta E_{1,1}^{(1)} \propto \lambda$ and $\Delta E_{2,1}^{(1)} \propto \lambda^2$.

- (b) Draw an energy diagram representing before and after the perturbation.

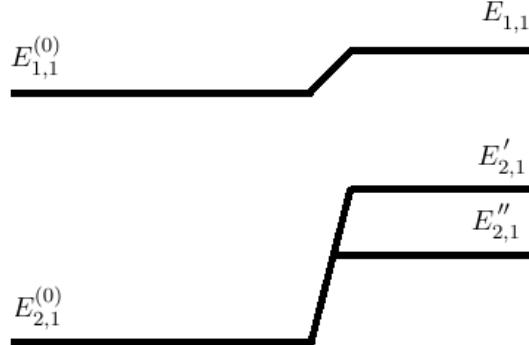


Figure 8.3

Example 8.5.

Consider an isotropic SHO in two dimensions. The Hamiltonian is

$$H = \frac{1}{2M}(p_x^2 + p_y^2) + \frac{M\omega^2}{2}(x^2 + y^2). \quad (8.93)$$

1. What are the energies of the three lowest-lying states? Is there any degeneracy?

The energy is given by the sum of the two harmonic oscillators

$$E_{n_x, n_y} = \hbar\omega\left(\frac{1}{2} + n_x\right) + \hbar\omega\left(\frac{1}{2} + n_y\right) = \hbar\omega(1 + n_x + n_y). \quad (8.94)$$

So, the lowest state

$$E_{0,0} = \hbar\omega. \quad (8.95)$$

There is no degeneracy. The next lowest level(s)

$$E_{1,0} = E_{0,1} = 2\hbar\omega, \quad (8.96)$$

where there is twofold degeneracy.

2. Now apply the perturbation

$$H' = \delta M\omega^2 xy, \quad (8.97)$$

where $\delta \ll 1$ is a real number. Find the zeroth order energy eigenket and first order corrections to the energy for each state.

Since the GS is non-degenerate, the zeroth order energy eigenstate is given by

$$|0\rangle \otimes |0\rangle = |0^{(0)}\rangle \otimes |0^{(0)}\rangle, \quad (8.98)$$

where " $|x\rangle \otimes |y\rangle$ " is chosen, i.e. the first state corresponds to x and the second state to y . For the first order correction to the energy

$$\begin{aligned} \Delta E_{0,0} &\simeq \langle 0^{(0)} | \otimes \langle 0^{(0)} | H' | 0^{(0)} \rangle \otimes | 0^{(0)} \rangle \\ &= \delta m\omega^2 \langle 0^{(0)} | x | 0^{(0)} \rangle \langle 0^{(0)} | y | 0^{(0)} \rangle, \end{aligned} \quad (8.99)$$

since the operator only operates onto its corresponding state. Using $\langle n' | x | n \rangle = \sqrt{\frac{\hbar}{2M\omega}}(\sqrt{n}\delta_{n',n-1} + \sqrt{n+1}\delta_{n',n+1})$ the above can be evaluated to be zero. I.e. there is no correction to the energy after the perturbation is applied (to first order).

In the case of the degenerate states I need to set up the degenerate matrix, H' and diagonalize it to find the eigenvalues and eigenvectors which are the first order correction(s) to the energy and the zeroth order correction to the eigenvector. In this case the basis is given by

$$\text{Basis } |0^{(0)}\rangle \otimes |1^{(0)}\rangle \text{ and } |1^{(0)}\rangle \otimes |0^{(0)}\rangle. \quad (8.100)$$

The basis must be the degenerate states and these are the states of the unperturbed Hamiltonian! From this I set up the degenerate matrix

$$\begin{aligned} H' &= \begin{bmatrix} \langle 1^{(0)} | \otimes \langle 0^{(0)} | H' | 0^{(0)} \rangle \otimes | 1^{(0)} \rangle & \langle 0^{(0)} | \otimes \langle 1^{(0)} | H' | 0^{(0)} \rangle \otimes | 1^{(0)} \rangle \\ \langle 1^{(0)} | \otimes \langle 0^{(0)} | H' | 1^{(0)} \rangle \otimes | 0^{(0)} \rangle & \langle 0^{(0)} | \otimes \langle 1^{(0)} | H' | 1^{(0)} \rangle \otimes | 0^{(0)} \rangle \end{bmatrix} \\ &= \begin{bmatrix} 0 & \delta m\omega^2 \langle 1^{(0)} | x | 0^{(0)} \rangle \langle 0^{(0)} | y | 1^{(0)} \rangle \\ \delta m\omega^2 \langle 0^{(0)} | x | 1^{(0)} \rangle \langle 1^{(0)} | y | 0^{(0)} \rangle & 0 \end{bmatrix} \\ &= \frac{\delta\hbar\omega}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \end{aligned} \quad (8.101)$$

The eigenvalues and corresponding eigenvectors

$$\lambda_1 = \frac{\delta\hbar\omega}{2} \Rightarrow |\lambda_1\rangle = \frac{1}{\sqrt{2}}(|0^{(0)}\rangle \otimes |1^{(0)}\rangle + |1^{(0)}\rangle \otimes |0^{(0)}\rangle), \quad (8.102)$$

$$\lambda_2 = -\frac{\delta\hbar\omega}{2} \Rightarrow |\lambda_2\rangle = \frac{1}{\sqrt{2}}(|0^{(0)}\rangle \otimes |1^{(0)}\rangle - |1^{(0)}\rangle \otimes |0^{(0)}\rangle), \quad (8.103)$$

where, as mentioned before, $\lambda_{1,2} = \Delta E_{1,0}^{(1)} = \Delta E_{0,1}^{(1)}$. So, to recap; after the perturbation the GS will remain the same while the energy levels $E_{1,0}$ and $E_{0,1}$ will split cf.

$$E_{0,0} = \hbar\omega \quad \text{and} \quad |0,0\rangle, \quad (8.104)$$

$$E_{0,1} = 2\hbar\omega + \frac{\delta\hbar\omega}{2} \quad \text{and} \quad |0, 1\rangle = \frac{1}{\sqrt{2}}(|0^{(0)}\rangle \otimes |1^{(0)}\rangle + |1^{(0)}\rangle \otimes |0^{(0)}\rangle), \quad (8.105)$$

$$E_{1,0} = 2\hbar\omega - \frac{\delta\hbar\omega}{2} \quad \text{and} \quad |1, 0\rangle = \frac{1}{\sqrt{2}}(|0^{(0)}\rangle \otimes |1^{(0)}\rangle - |1^{(0)}\rangle \otimes |0^{(0)}\rangle). \quad (8.106)$$

To first order in the energy and zeroth order in the eigenvectors.

Example 8.6.

Evaluate the matrix elements (or expectation values) given below. If any vanishes, explain why it vanishes using simple symmetry (or other) arguments.

1. $\langle n=2, l=1, m=0 | x | n=2, l=0, m=0 \rangle$.

This is zero. To show that I use that x is a component of \vec{r} which is a vector operator. A vector operator is again a rank 1 spherical tensor operator, and therefore x must be a linear combination rank 1 spherical tensors of the form $T_q^{(k)}$ - where $k = 1$ is the rank and $q \in [-1, 0, 1]$ is the magnetic number. In fact, x can be written as¹³

$$x = \frac{T_{-1}^{(1)} - T_1^{(1)}}{\sqrt{2}}, \quad (8.107)$$

where

$$T_q^{(k)} = \sqrt{\frac{4\pi}{3}} \vec{r} \cdot Y_k^q. \quad (8.108)$$

From this

$$\begin{aligned} \langle n, l, m | x | n', l', m' \rangle &= \langle n, l, m | \frac{T_{-1}^{(1)} - T_1^{(1)}}{\sqrt{2}} | n', l', m' \rangle \\ &= \frac{1}{\sqrt{2}} \langle n, l, m | T_{-1}^{(1)} | n', l', m' \rangle \\ &\quad - \frac{1}{\sqrt{2}} \langle n, l, m | T_1^{(1)} | n', l', m' \rangle. \end{aligned} \quad (8.109)$$

Hereafter I use that $\langle n, l, m | T_1^{(1)} | n', l', m' \rangle = 0$ unless $m = q + m'$. Since in this case $q = \pm 1$ and $m = m' = 0$ the matrix element is zero!

2. $\langle n=2, l=1, m=0 | p_z | n=2, l=0, m=0 \rangle$.

In this case I use that the momentum can be described via the commutator of the position operator with the Hamiltonian operator via

$$[z, H] = \frac{i\hbar}{M} p_z, \quad (8.110)$$

where I have used that $[x_i, f(\vec{x}, \vec{p})] = i\hbar \frac{\partial f(\vec{x}, \vec{p})}{\partial p_i}$.¹⁴ By using this

$$\begin{aligned} \langle n, l, m | p_z | n', l', m' \rangle &= \frac{M}{i\hbar} \langle n, l, m | [z, H] | n', l', m' \rangle \\ &= \frac{M}{i\hbar} \langle n, l, m | zH | n', l', m' \rangle - \frac{M}{i\hbar} \langle n, l, m | Hz | n', l', m' \rangle \\ &= \frac{M}{i\hbar} \langle n, l, m | z | n', l', m' \rangle (E_{n', l', m'} - E_{n, l, m}), \end{aligned} \quad (8.111)$$

¹³ $T_0^{(1)} = z$

¹⁴ Likewise; $[p_i, f(\vec{x}, \vec{p})] = -i\hbar \frac{\partial f(\vec{x}, \vec{p})}{\partial x_i}$.

where $E_{n',l',m'} - E_{n,l,m} = E_{2,0,0} - E_{2,1,0} = 0$ in this case because of degeneracy (see figure 8.4).

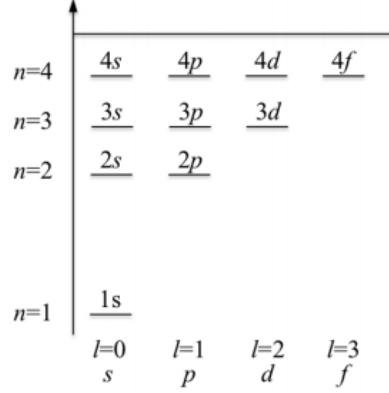


Figure 8.4: The lower energy levels of hydrogen. The degeneracy of $2s$ and $2p$ is evident.

3. $\langle j = \frac{9}{2}, m = \frac{7}{2}, l = 4 | L_z | j = \frac{9}{2}, m = \frac{7}{2}, l = 4 \rangle$ for an electron in a central force field.

The above is not directly evaluate-able since $|j, l, m\rangle$ is not an eigenstate of L_z . Instead I use that

$$J_z = J_{1z} + J_{2z} = L_z + S_z. \quad (8.112)$$

In combination with

$$J_z |j, l, m\rangle = m\hbar |j, l, m\rangle, \quad (8.113)$$

$$\langle j, m, l | S_z | j, m, l \rangle |_{j=l \pm \frac{1}{2}, m} = \pm \frac{m\hbar}{2l+1}, \quad (8.114)$$

where I took the last identity from Sakurai (1994)[p.338]. The result

$$\langle j, m, l | L_z | j, m, l \rangle = \frac{7}{2}\hbar - \frac{7\hbar}{2(24+1)} = \frac{28}{9}\hbar. \quad (8.115)$$

4. $\langle \text{singlet}, m_s = 0 | S_z^{(e^-)} - S_z^{(e^+)} | \text{triplet}, m_s = 0 \rangle$ for an s-state positronium.

I use the definitions of the singlet and triplet states. The triplet state

$$\vec{S}^2 : \quad \lambda_4^{(j=1, m=0)} = \frac{\hbar^2}{2}, \quad |\lambda_4\rangle = \begin{bmatrix} \langle +, + | \lambda_4 \rangle \\ \langle +, - | \lambda_4 \rangle \\ \langle -, + | \lambda_4 \rangle \\ \langle -, - | \lambda_4 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \quad (8.116)$$

The above is also an eigenvector of S_z since $[S_z, \vec{S}^2] = 0$. Reformulated in a normalized way

$$|j = 1, m_s = 0\rangle = \frac{1}{\sqrt{2}} \left(|+, -\rangle + |-, +\rangle \right). \quad (8.117)$$

The singlet state

$$\vec{S}^2 : \quad \lambda_1^{(j=0, m=0)} = 0, \quad |\lambda_1\rangle = \begin{bmatrix} \langle +, + | \lambda_1 \rangle \\ \langle +, - | \lambda_1 \rangle \\ \langle -, + | \lambda_1 \rangle \\ \langle -, - | \lambda_1 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}. \quad (8.118)$$

Reformulated in a normalized way

$$|j = 0, m_s = 0\rangle = \frac{1}{\sqrt{2}} \left(|+, -\rangle - |-, +\rangle \right), \quad (8.119)$$

with the singlet and triplet states expressed in terms of the electron states I can now evaluate the matrix element. I let the operators act on the triplet state

$$\begin{aligned} \left(S_z^{(e^-)} - S_z^{(e^+)} \right) |j = 1, m_s = 0\rangle &= \left(S_z^{(e^-)} - S_z^{(e^+)} \right) \frac{1}{\sqrt{2}} \left(|+, -\rangle + |-, +\rangle \right) \\ &= \frac{1}{\sqrt{2}} \left(\frac{\hbar}{2} + \frac{\hbar}{2} \right) |+, -\rangle + \frac{1}{\sqrt{2}} \left(-\frac{\hbar}{2} - \frac{\hbar}{2} \right) |-, +\rangle \\ &= \frac{\hbar}{\sqrt{2}} \left(|+, -\rangle - |-, +\rangle \right) \\ &= \hbar |j = 0, m_s = 0\rangle. \end{aligned} \quad (8.120)$$

So, the matrix element is found to be

$$\begin{aligned} \hbar \langle j = 0, m_s = 0 | S_z^{(e^-)} - S_z^{(e^+)} | j = 1, m_s = 0 \rangle, \\ \hbar \langle j = 0, m_s = 0 | j = 0, m_s = 0 \rangle. \end{aligned} \quad (8.121)$$

5. $\langle \vec{S}_1 \cdot \vec{S}_2 \rangle$ for the ground state of a hydrogen molecule (H_2).

I use that

$$\vec{S}_1 \cdot \vec{S}_2 = \frac{1}{2} (\vec{S}^2 - \vec{S}_1^2 - \vec{S}_2^2). \quad (8.122)$$

I take the lowest state to be a spin singlet with $m_s = 0$. Thereby $\langle \vec{S} \rangle = 0$ and $\langle \vec{S}_1^2 \rangle = \langle \vec{S}_2^2 \rangle = \frac{1}{2}(\frac{1}{2} + 1)\hbar^2 = \frac{3}{4}\hbar^2$. Thereby

$$\langle \vec{S}_1 \cdot \vec{S}_2 \rangle = \frac{1}{2} (\vec{S}^2 - \vec{S}_1^2 - \vec{S}_2^2) = \frac{1}{2} (0 - \frac{3}{4}\hbar^2 - \frac{3}{4}\hbar^2) = -\frac{3}{4}\hbar^2. \quad (8.123)$$

Example 8.7.

A one-electron atom whose GS is non-degenerate is placed in a uniform electric field in the z -direction. Obtain an approximate expression for the induced electric dipole moment of the GS by considering the expectation value of ez with respect to the perturbed-state vector computed to first order. Show that the same expression can also be found from the energy shift of the GS computed to second order. Ignore spin.

In this exercise the GS of a hydrogen-like atom is considered. That is the state $|n, l, m\rangle = |1, 0, 0\rangle$. Because the electric field is uniform, and in the z -direction, the perturbative Hamiltonian is given by

$$H' = -ez|\vec{E}|. \quad (8.124)$$

Since the GS is non-degenerate the state vector to first order is given by equation (8.20)

$$|1, 0, 0\rangle \simeq |(1, 0, 0)^0\rangle + \sum_{k \neq (1, 0, 0)^0} \frac{\langle k^0 | H' | (1, 0, 0)^0 \rangle}{E_{(1, 0, 0)^0}^{(0)} - E_k^{(0)}} |k^0\rangle. \quad (8.125)$$

So, the expectation value

$$\begin{aligned} \langle H' \rangle_{1,0,0} &= -e|\vec{E}| \langle 1, 0, 0 | z | 1, 0, 0 \rangle \\ &\simeq -e|\vec{E}| \left(-e|\vec{E}| \sum_{k \neq (1, 0, 0)^0} \frac{\langle (1, 0, 0)^0 | z | k^0 \rangle}{E_{(1, 0, 0)^0}^{(0)} - E_k^{(0)}} \langle k^0 | + \langle (1, 0, 0)^0 | \right) \\ &\quad \cdot z \left(|(1, 0, 0)^0\rangle - e|\vec{E}| \sum_{k \neq (1, 0, 0)^0} \frac{\langle k^0 | z | (1, 0, 0)^0 \rangle}{E_{(1, 0, 0)^0}^{(0)} - E_k^{(0)}} |k^0\rangle \right). \end{aligned} \quad (8.126)$$

To evaluate this I use that $\langle n', l', m' | z | n, l, m \rangle = 0$ unless $l' = l \pm 1$ and $m = m'$. So, the element $\langle (1, 0, 0)^0 | z | (1, 0, 0)^0 \rangle = 0$. So, the remainder

$$\begin{aligned} \langle H' \rangle_{1,0,0} &\simeq -e|\vec{E}| \left(-2e|\vec{E}| \sum_{k \neq (1,0,0)^0} \frac{|\langle (1,0,0)^0 | z | k^0 \rangle|^2}{E_{(1,0,0)^0}^{(0)} - E_k^{(0)}} \right. \\ &\quad \left. + e^2 |\vec{E}|^2 \sum_{k' \neq (1,0,0)^0} \sum_{k \neq (1,0,0)^0} \frac{\langle k^0 | z | (1,0,0)^0 \rangle \langle (1,0,0)^0 | z | k'^0 \rangle}{E_{(1,0,0)^0}^{(0)} - E_k^{(0)} E_{(1,0,0)^0}^{(0)} - E_{k'}^{(0)}} \langle k'^0 | z | k^0 \rangle \right). \end{aligned} \quad (8.127)$$

The last term vanishes since $0 \neq \langle k^0 | z | (1,0,0)^0 \rangle$ if $k = n, 1, 0$ and likewise for $\langle (1,0,0)^0 | z | k'^0 \rangle$. However, $\langle k'^0 | z | k^0 \rangle = 0$ in these cases. Therefore

$$\langle H' \rangle_{1,0,0} \simeq 2e^2 |\vec{E}|^2 \sum_{k \neq (1,0,0)^0} \frac{|\langle (1,0,0)^0 | z | k^0 \rangle|^2}{E_{(1,0,0)^0}^{(0)} - E_k^{(0)}}. \quad (8.128)$$

Example 8.8.

The Hamiltonian for a two-state system can be written as

$$H = H_0 + H' \doteq \begin{bmatrix} E_1^{(0)} & 0 \\ 0 & E_2^{(0)} \end{bmatrix} + \lambda \begin{bmatrix} 0 & \Delta \\ \Delta & 0 \end{bmatrix} = \begin{bmatrix} E_1^{(0)} & \lambda\Delta \\ \lambda\Delta & E_2^{(0)} \end{bmatrix}. \quad (8.129)$$

The energy eigenfunctions for the unperturbed problems ($\lambda = 0$) are given by

$$|\phi_1^{(0)}\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |\phi_2^{(0)}\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (8.130)$$

1. Solve this problem exactly to find the energy eigenvectors $|E_{1,2}\rangle$ and the energy eigenvalues $E_{1,2}$.

The problem can be set up as a matrix equation on the form

$$H\vec{x} = E\vec{x}. \quad (8.131)$$

Cf. Cramers rule (equation (5.46)) this only has non-trivial solutions when

$$\det(H - EI) = 0. \quad (8.132)$$

From this I find the eigenvalues to be

$$E_{1,2} = \frac{E_1^{(0)} + E_2^{(0)} \mp \sqrt{(E_2^{(0)} - E_1^{(0)})^2 + 4\lambda^2\Delta^2}}{2}. \quad (8.133)$$

By insertion of the eigenvalues into equation (8.131) the components of the eigenvectors can be found. This will result in two non-independent equations, and so one of the coefficients can be chosen to be 1 in order to determine the other. The unnormalized eigenvectors are therefore

$$|E_1\rangle = \begin{bmatrix} 1 \\ \frac{E_1 - E_2^{(0)}}{\lambda\Delta} \end{bmatrix}; \quad |E_2\rangle = \begin{bmatrix} \frac{E_2 - E_1^{(0)}}{\lambda\Delta} \\ 1 \end{bmatrix}. \quad (8.134)$$

The reason for the odd way to list the eigenvectors is to compare with the results from perturbation theory easily.

2. Assuming that $\lambda|\Delta| \ll |E_1^{(0)} - E_2^{(2)}|$, solve the same problem using time-independent perturbation theory up to first order in the energy eigenfunctions and up to second order in the energy eigenvalues. Compare with the exact results.

From equation (8.19)

$$\Delta E_n = \langle n^0 | H' | n^0 \rangle + \sum_{k \neq n} \frac{|\langle k^0 | H' | n^0 \rangle|^2}{E_n^0 - E_k^0} + \dots, \quad (8.135)$$

where $\langle n^0 | H' | n^0 \rangle$ is the diagonal elements of the *potential* matrix given by

$$H' \doteq \begin{bmatrix} 0 & \lambda\Delta \\ \lambda\Delta & 0 \end{bmatrix}. \quad (8.136)$$

Hence the diagonal elements are zero, and so the first order correction to the energy (for both energy eigenvalues) is zero. For the second order correction

$$\begin{aligned} \Delta E_1 &= \lambda \langle 1^0 | H' | 1^0 \rangle + \lambda^2 \sum_{k \neq 1} \frac{|\langle k^0 | H' | 1^0 \rangle|^2}{E_1^0 - E_k^0} + \dots \\ &= \lambda^2 \frac{|\Delta|^2}{E_1^{(0)} - E_2^{(0)}}, \\ \Delta E_2 &= \langle 2^0 | H' | 2^0 \rangle + \sum_{k \neq 1} \frac{|\langle k^0 | H' | 2^0 \rangle|^2}{E_2^0 - E_k^0} + \dots \\ &= \lambda^2 \frac{|\Delta|^2}{E_2^{(0)} - E_1^{(0)}} + \dots \end{aligned} \quad (8.137)$$

To compare with the exact result I consider E_1 as an example. I can rewrite it as¹⁵

$$\begin{aligned} E_1 &= \frac{E_1^{(0)} + E_2^{(0)}}{2} + \frac{E_2^{(0)} - E_1^{(0)}}{2} \sqrt{1 + \frac{4\lambda^2\Delta^2}{(E_2^{(0)} - E_1^{(0)})^2}} \\ &\simeq \frac{E_1^{(0)} + E_2^{(0)}}{2} + \frac{E_2^{(0)} - E_1^{(0)}}{2} \left(1 + \frac{2\lambda^2\Delta^2}{(E_2^{(0)} - E_1^{(0)})^2} \right) \\ &= E_1^{(0)} + \frac{\lambda^2\Delta^2}{E_1^{(0)} - E_2^{(0)}}, \end{aligned} \quad (8.138)$$

which is the same as the exact case¹⁶. For the eigenstates

$$\begin{aligned} |E_1\rangle &= |E_1^{(0)}\rangle + \lambda \sum_{k \neq 1} \frac{\langle k'^0 | H' | E_1^{(0)} \rangle}{E_1^{(0)} - E_{k'}^{(0)}} |k^0\rangle + \dots \\ &= |E_1^{(0)}\rangle + \frac{\lambda\Delta}{E_1^{(0)} - E_2^{(0)}} |E_2^{(0)}\rangle + \dots, \\ |E_2\rangle &= |E_2^{(0)}\rangle + \lambda \sum_{k \neq 1} \frac{\langle k'^0 | H' | E_2^{(0)} \rangle}{E_2^{(0)} - E_{k'}^{(0)}} |k^0\rangle + \dots \\ &= |E_2^{(0)}\rangle + \frac{\lambda\Delta}{E_2^{(0)} - E_1^{(0)}} |E_1^{(0)}\rangle + \dots. \end{aligned} \quad (8.139)$$

¹⁵where I use $\lambda|\Delta| \ll |E_1^{(0)} - E_2^{(2)}|$ to Taylor expand the square root to second order; $\sqrt{1+x} = 1 + \frac{x}{2} + \dots$

¹⁶A completely analogous procedure can be performed for E_2 .

Inserting the eigenvalues in the exact results and Taylor expanding the square roots to second order will reveal the eigenvectors above found from perturbation theory

$$\begin{aligned}
|E_1\rangle &\doteq \begin{bmatrix} 1 \\ \frac{E_1 - E_1^{(0)}}{\lambda\Delta} \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ \frac{1}{\lambda\Delta} \left(\frac{E_1^{(0)} + E_2^{(0)}}{2} - \frac{(E_2^{(0)} - E_1^{(0)})}{2} \right) \sqrt{1 + \frac{4\lambda^2\Delta^2}{(E_2^{(0)} - E_1^{(0)})^2}} - E_1^{(0)} \end{bmatrix} \\
&\simeq \begin{bmatrix} 1 \\ \frac{1}{\lambda\Delta} \left(\frac{E_1^{(0)} + E_2^{(0)}}{2} - \frac{E_2^{(0)} - E_1^{(0)}}{2} \left(1 + \frac{2\lambda^2\Delta^2}{(E_2^{(0)} - E_1^{(0)})^2} \right) - E_1^{(0)} \right) \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ -\frac{1}{\lambda\Delta} \frac{E_2^{(0)} - E_1^{(0)}}{2} \frac{2\lambda^2\Delta^2}{(E_2^{(0)} - E_1^{(0)})^2} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{\lambda\Delta}{E_1^{(0)} - E_2^{(0)}} \end{bmatrix},
\end{aligned} \tag{8.140}$$

which is identical to equation (8.139).

3. Suppose the two unperturbed energy levels are "almost degenerate", i.e. $\lambda|\Delta| \gg |E_1^{(0)} - E_2^{(2)}|$. Show that the exact results obtained earlier closely resemble what you would expect by applying degenerate perturbation theory to this problem with $E_1^{(0)} = E_2^{(0)} = E^{(0)}$.

In this case the Hamiltonian will be on the form

$$\begin{aligned}
H &= H_0 + H' \\
&\doteq E^{(0)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \lambda\Delta \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.
\end{aligned} \tag{8.141}$$

I diagonalize $\lambda H'$. The eigenvalues are the first order correction to the energy whereas the eigenvectors are the eigenvectors of the energy levels. I find

$$E_1 = E^{(0)} + \lambda\Delta \quad \text{and} \quad E_2 = E^{(0)} - \lambda\Delta, \tag{8.142}$$

$$|E_1\rangle = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad |E_2\rangle = \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \tag{8.143}$$

Letting $E_1^{(0)} = E_2^{(0)} = E^{(0)}$ in equation (8.133) results in $E_i = E^{(0)} \pm \lambda\Delta$ just as above. Using $E_i = E^{(0)} \pm \lambda\Delta$ in equation (8.134) results in $|E_i\rangle = \begin{bmatrix} \pm 1 \\ 1 \end{bmatrix}$ - just as above. Therefore, in the limit of almost degenerate energy levels, the exact method and the degenerate perturbation theory reveals the same result.

Example 8.9.

To illustrate the calculations required, I analyze the semi classical case of an electric dipole. I take $H'(t)$ to be on the form

$$H'(t) = -\vec{\mu} \cdot \vec{E}_0 \cos(\omega t), \tag{8.144}$$

where \vec{E}_0 is the magnitude of the electrical field, and unrelated to the energy, E_n . By using equation (8.144) in (8.44) I find the squared amplitude to be

$$|c_{b_0}^1(t_p)|^2 = \left(\frac{E_0}{\hbar} \right)^2 |\langle b_0 | \mu | a_0 \rangle|^2 \left| \int_0^{t_p} (-) e^{i\omega_{b_0} a_0 t} \cos(\omega t) dt \right|^2. \tag{8.145}$$

I define the integral as I and evaluate it

$$\begin{aligned}
I &= \int_0^{t_p} (-) e^{i\omega_{b_0 a_0} t} \cos(\omega t) dt \\
&= \int_0^{t_p} (-) e^{i\omega_{b_0 a_0} t} \left(\frac{e^{i\omega t} + e^{-i\omega t}}{2} \right) dt \\
&= -\frac{1}{2} \int_0^{t_p} e^{i\omega_{b_0 a_0} t} (e^{i\omega t} + e^{-i\omega t}) dt \\
&= -\frac{1}{2} \left(\int_0^{t_p} e^{i(\omega_{b_0 a_0} + \omega)t} dt + \int_0^{t_p} e^{i(\omega_{b_0 a_0} - \omega)t} dt \right) \\
&= -\frac{1}{2} \left(\frac{e^{i(\omega_{b_0 a_0} + \omega)t_p} - 1}{i(\omega_{b_0 a_0} + \omega)} + \frac{e^{i(\omega_{b_0 a_0} - \omega)t_p} - 1}{i(\omega_{b_0 a_0} - \omega)} \right) \\
&= \frac{1}{2i} \left(\frac{1 - e^{i(\omega_{b_0 a_0} + \omega)t_p}}{\omega_{b_0 a_0} + \omega} + \frac{1 - e^{i(\omega_{b_0 a_0} - \omega)t_p}}{\omega_{b_0 a_0} - \omega} \right).
\end{aligned} \tag{8.146}$$

Using equation (8.146) in 8.145 and letting the magnitude-operator absorb the imaginary unit, I find

$$|c_{b_0}^1(t_p)|^2 = \left(\frac{E_0}{2\hbar} \right)^2 |\langle b_0 | \mu | a_0 \rangle|^2 \left| \frac{1 - e^{i(\omega_{b_0 a_0} + \omega)t_p}}{\omega_{b_0 a_0} + \omega} + \frac{1 - e^{i(\omega_{b_0 a_0} - \omega)t_p}}{\omega_{b_0 a_0} - \omega} \right|^2. \tag{8.147}$$

To get considerable probability, the denominator needs to explode. This gives rise to the criterion for a large transition probability

$$\omega \simeq |\omega_{b_0 a_0}|. \tag{8.148}$$

Equation (8.148) causes the denominator in one of the fractions to "explode" -the other one does not explode, and can therefore be neglected. I find

$$\begin{aligned}
\omega \simeq \omega_{b_0 a_0} \quad &|c_{b_0}^1(t_p)|^2 \simeq \left(\frac{E_0}{2\hbar} \right)^2 |\langle b_0 | \mu | a_0 \rangle|^2 \left| \frac{\sin(\frac{\omega_{b_0 a_0} - \omega}{2} t_p)}{\frac{\omega_{b_0 a_0} - \omega}{2}} \right|^2, \\
\omega \simeq -\omega_{b_0 a_0} \quad &|c_{b_0}^1(t_p)|^2 \simeq \left(\frac{E_0}{2\hbar} \right)^2 |\langle b_0 | \mu | a_0 \rangle|^2 \left| \frac{\sin(\frac{\omega_{b_0 a_0} + \omega}{2} t_p)}{\frac{\omega_{b_0 a_0} + \omega}{2}} \right|^2.
\end{aligned} \tag{8.149}$$

Using the definition of $\omega_{m_0 n_0} = \frac{E_{m_0} - E_{n_0}}{\hbar}$

$$\begin{aligned}
\omega_{m_0 n_0} > 0 : \quad &\omega_{m_0 n_0} = \frac{E_{b_0} - E_{a_0}}{\hbar} > 0 \\
&\Rightarrow E_{b_0} > E_{a_0} \Rightarrow \text{Absorption of 1 photon}, \\
\omega_{m_0 n_0} < 0 : \quad &\omega_{m_0 n_0} = \frac{E_{b_0} - E_{a_0}}{\hbar} < 0 \\
&\Rightarrow E_{b_0} < E_{a_0} \Rightarrow \text{Emission of 1 photon}.
\end{aligned} \tag{8.150}$$

In equation (8.149) $\omega_{b_0 a_0} \pm \omega$ is raised to the second power and sine is an odd function to the second power¹⁷. This means that the probability for absorption and emission of 1 photon is the same. The probability depends on the perturbation time and the frequency of the electromagnetic wave. The dependency of the perturbation time is shown in figure 8.5. As is evident from figure 8.5; the transition probability varies periodically with the perturbation time. To have greatest possibility of a transition, the perturbation time must be given by

$$t_p = \frac{(2k+1)\pi}{\omega_{b_0 a_0} \pm \omega}, \tag{8.151}$$

¹⁷Meaning that $\sin(-x) = -\sin(x)$.

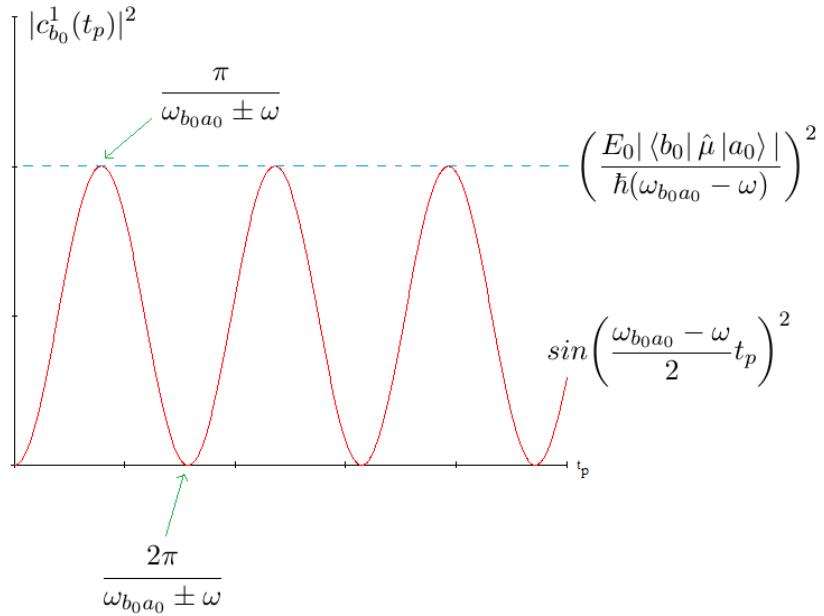


Figure 8.5: The transition probability as a function of perturbation time.

where $k = 1, 2, 3, \dots$. The dependency of the transition probability on the angular frequency of the electromagnetic wave is shown in figure 8.6. Here equation (8.149) is rewritten to

$$\omega \simeq \omega_{b_0 a_0} \quad |c_{b_0}^1(t_p)|^2 \simeq \left(\frac{E_0 |\langle b_0 | \hat{\mu} | a_0 \rangle|}{2\hbar} \right)^2 t_p^2 \left| \text{sinc}\left(\frac{\omega_{b_0 a_0} - \omega}{2} t_p \right) \right|^2. \quad (8.152)$$

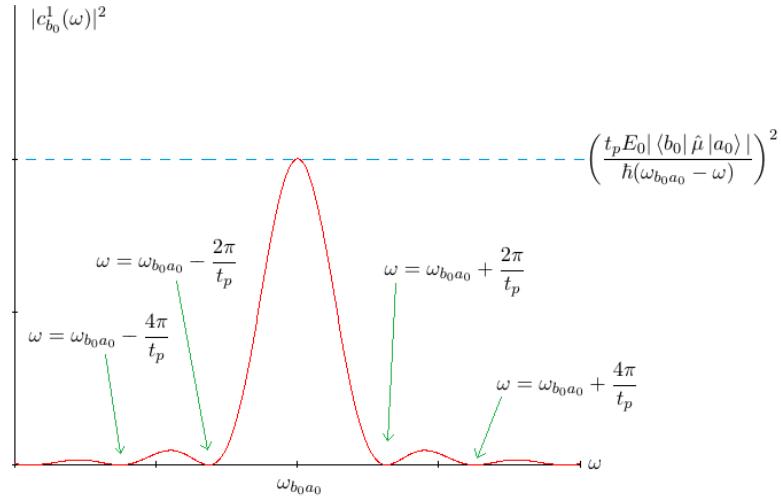


Figure 8.6: The transition probability as a function of the angular frequency of the electromagnetic wave.

As is evident from figure 8.6; the transition probability peaks around $\omega_{b_0 a_0}$. Thus, in order to maximize the transition probability, the time, t_p , and angular frequency, ω , must have certain values.

Example 8.10.

Consider a one-dimensional simple harmonic oscillator whose classical angular frequency is ω_0 . For $t < 0$ it is known to be in the ground state. For $t > 0$ there is also a time-dependent potential given by

$$H'(t) = F_0 x \cos(\omega t), \quad (8.153)$$

where F_0 is constant in both space and time. Obtain an expression for the expectation value $\langle x \rangle$ as a function of time using time-dependent perturbation theory to lowest non-vanishing order. Is this procedure valid for $\omega \simeq \omega_0$?

What is to be evaluated is

$$\langle x \rangle = \langle 0(t) | x | 0(t) \rangle, \quad (8.154)$$

where

$$|0(t)\rangle = \sum_n c_n(t) e^{-i\frac{H_0 t}{\hbar}} |n\rangle, \quad c_n(t) = c_n^{(0)} + c_n^{(1)}(t) + c_n^{(2)}(t) + \dots \quad (8.155)$$

To first order; $c_0(t) = c_0^{(0)}$ since $\langle 0 | x | 0 \rangle$ from the first order term ($c_0^{(1)}$) is zero. For the first energy level; $c_1(t) = c_1^{(0)} + c_1^{(1)}(t)$ where $c_1^0 = 0$ cf. the initial condition that the state at $t = 0$ is in the ground state. The rest of the coefficients, $c_n(t)$ ($n \geq 2$) vanish since

$$\langle n' | x | n \rangle = \sqrt{\frac{\hbar}{2M\omega_0}} \left(\sqrt{n+1} \delta_{n',+1} + \sqrt{n} \delta_{n',n-1} \right). \quad (8.156)$$

Therefore

$$|0(t)\rangle = c_0^{(0)} e^{-\frac{1}{2}i\omega t} |0\rangle + c_1^{(1)} e^{-\frac{3}{2}i\omega t} |1\rangle. \quad (8.157)$$

Since at $t = 0$ the state is in the ground state; $c_0^0 \simeq 1$. For the first excited state (from equation (8.44))

$$\begin{aligned} c_1^1(t) &= \frac{1}{i\hbar} \int_0^t \langle n | H'(t') | 0 \rangle e^{i\omega_0 t'} dt' \\ &= \frac{F_0 \langle n | x | 0 \rangle}{i\hbar} \int_0^t \cos(\omega t') e^{i\omega_0 t'} dt' \\ &= \frac{F_0 \langle n | x | 0 \rangle}{i\hbar} \left(\int_0^t \cos(\omega t') \cos(\omega_0 t') dt' + i \int_0^t \cos(\omega t') \sin(\omega_0 t') dt' \right) \\ &= \frac{F_0 \langle n | x | 0 \rangle}{2\hbar} \left(\frac{e^{-i(\omega-\omega_0)t} - 1}{\omega - \omega_0} - \frac{e^{i(\omega+\omega_0)t} - 1}{\omega + \omega_0} \right). \end{aligned} \quad (8.158)$$

So, to first order the expectation value can be written as

$$\begin{aligned} \langle 0(t) | x | 0(t) \rangle &= (\langle 1 | e^{\frac{i3\hbar\omega_0 t}{2}} c_1^{(1)*} + \langle 0 | e^{\frac{i\hbar\omega_0 t}{2}} c_0^{(0)} \\ &\quad \cdot x(c_0^{(0)*} e^{\frac{-i\hbar\omega_0 t}{2}} |0\rangle + c_1^{(1)} e^{\frac{-i3\hbar\omega_0 t}{2}} |1\rangle) \\ &= \langle 1 | x | 0 \rangle e^{i\omega_0 t} c_1^{(1)*} + \langle 0 | x | 1 \rangle e^{-i\omega_0 t} c_1^{(1)}, \end{aligned} \quad (8.159)$$

where equation (8.156) has been used again to make the terms with $\langle n | x | n \rangle$ vanish. Thereby

$$\langle x \rangle = \frac{F_0}{M} \frac{\cos(\omega t) - \cos(\omega_0 t)}{\omega^2 - \omega_0^2}, \quad (8.160)$$

where I have used that $\langle 1 | x | 0 \rangle = \sqrt{\frac{\hbar}{2M\omega_0}}$.

Example 8.11.

A one-dimensional SHO is in its ground state for $t < 0$. For $t \geq 0$ it is subjected to a time-dependent but spatially uniform force in the x -direction

$$\vec{F}(t) = F_0 e^{-\frac{t}{\tau}} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \quad (8.161)$$

1. Using TDPT to first order, obtain the probability of finding the oscillator in its first excited state for $t > 0$. Show that in the limit of $t \Rightarrow \infty$ the result is independent of time. Is this reasonable or surprising?

First I convert the perturbative force into a perturbative potential which I can add to the Hamiltonian. I use that $\vec{F} = -\vec{\nabla} \cdot V(t)$. Hereby

$$V(t) = H'(t) = -F_0 x e^{-\frac{t}{\tau}}. \quad (8.162)$$

The transition probability

$$P_{0 \Rightarrow 1}(t) \simeq |c_1^{(1)}(t)|^2, \quad (8.163)$$

where

$$\begin{aligned} c_1^{(1)}(t) &= \frac{1}{i\hbar} \int_0^t \langle 1 | H'(t') | 0 \rangle e^{i\omega_{10} t'} dt' \\ &= -\frac{\langle 1 | x | 0 \rangle F_0}{i\hbar} \int_0^t e^{t'(i\omega_{10} - \frac{1}{\tau})} dt' \\ &= -\frac{\langle 1 | x | 0 \rangle F_0}{i\hbar} \frac{e^{t(i\omega_{10} - \frac{1}{\tau})} - 1}{(i\omega_{10} - \frac{1}{\tau})}. \end{aligned} \quad (8.164)$$

By insertion

$$\begin{aligned} P_{0 \Rightarrow 1}(t) &\simeq \left| -\frac{\langle 1 | x | 0 \rangle F_0}{i\hbar} \frac{e^{t(i\omega_{10} - \frac{1}{\tau})} - 1}{(i\omega_{10} - \frac{1}{\tau})} \right|^2 \\ &= \frac{|\langle 1 | x | 0 \rangle|^2 F_0^2}{\hbar^2} \left| \frac{e^{t(i\omega_{10} - \frac{1}{\tau})} - 1}{(i\omega_{10} - \frac{1}{\tau})} \right|^2 \\ &= \frac{|\langle 1 | x | 0 \rangle|^2 F_0^2}{\hbar^2} \frac{e^{t(i\omega_{10} - \frac{1}{\tau})} - 1}{(i\omega_{10} - \frac{1}{\tau})} \frac{e^{t(-i\omega_{10} - \frac{1}{\tau})} - 1}{(-i\omega_{10} - \frac{1}{\tau})} \\ &= \frac{|\langle 1 | x | 0 \rangle|^2 F_0^2}{\hbar^2} \frac{e^{-\frac{2t}{\tau}} - e^{-\frac{t}{\tau}} 2\cos(\omega_{10}t) + 1}{\omega_{10}^2 + \frac{1}{\tau^2}}. \end{aligned} \quad (8.165)$$

In the limit of $t \Rightarrow \infty$

$$\lim_{t \Rightarrow \infty} (P_{0 \Rightarrow 1}(t)) \simeq \frac{|\langle 1 | x | 0 \rangle|^2 F_0^2}{\hbar^2} \frac{1}{\omega_{10}^2 + \frac{1}{\tau^2}}. \quad (8.166)$$

This makes sense since the perturbation will vanish in this limit.

2. Can higher states be the destination of the perturbation?

Not to first order since $\langle n' | x | 0 \rangle = \sqrt{\frac{\hbar}{2M\omega}} \delta_{n',1}$. To second order (cf. equation (8.45))

$$c_{b_0}^{(2)}(t) = -\frac{1}{\hbar^2} \int_0^t \langle b_0 | H'(t'') | 1 \rangle e^{i\omega_{b_0,1} t''} \int_0^{t''} \langle 1 | H'(t') | 0 \rangle e^{i\omega_{n_0,0} t'} dt' dt'', \quad (8.167)$$

where I used that only $\langle 1 | H'(t') | 0 \rangle \neq 0$ which also requires $b_0 = 2(!)$. This expression is in general non-zero, and so there is a finite transition probability to the second excited state.

Example 8.12.

Consider a particle bound in a SHO potential. Initially ($t < 0$) it is in the GS. At $t = 0$ a perturbation is turned on. the perturbation is given by

$$H'(x, t) = Ax^2 e^{-\frac{t}{\tau}}. \quad (8.168)$$

calculate the probability that after a long time, $t \gg \tau$, the system will have made a transition to a given excited state. Consider all final states.

Since $c_n^{(1)} \propto \langle n^0 | x^2 | 0^0 \rangle$ I consider these matrix elements. Cf. equation (8.76)

$$\begin{aligned}\langle n^0 | x^2 | 0^0 \rangle &= \frac{\hbar}{2M\omega} \left(\langle n^0 | a^\dagger a | 0^0 \rangle + \langle n^0 | a^\dagger a^\dagger | 0^0 \rangle + \langle n^0 | aa^\dagger | 0^0 \rangle + \langle n^0 | aa | 0^0 \rangle \right) \\ &= \frac{\hbar}{2M\omega} \left(\langle n^0 | a^\dagger a^\dagger | 0^0 \rangle \delta_{n,2} + \langle n^0 | aa^\dagger | 0^0 \rangle \delta_{n,0} \right).\end{aligned}\quad (8.169)$$

Here only the GS and the second excited state reveals any results. So, the relevant probability here is given by

$$P_{0 \Rightarrow 2}(t) = |\langle 2 | \alpha(t) \rangle|^2 = |c_2^{(1)}(t)|^2, \quad (8.170)$$

where

$$\begin{aligned}c_2^{(1)}(t) &= \frac{1}{i\hbar} \int_0^t \langle 2 | H'(t') | 0 \rangle e^{i\omega_{20} t'} dt' \\ &= -\frac{\langle 2 | x^2 | 0 \rangle A}{i\hbar} \int_0^t e^{t'(i\omega_{20} - \frac{1}{\tau})} dt' \\ &= -\frac{\langle 2 | x^2 | 0 \rangle A}{i\hbar} \frac{e^{t(i\omega_{20} - \frac{1}{\tau})} - 1}{(i\omega_{20} - \frac{1}{\tau})}.\end{aligned}\quad (8.171)$$

By insertion

$$\begin{aligned}P_{0 \Rightarrow 2}(t) &\simeq \left| -\frac{\langle 2 | x^2 | 0 \rangle A}{i\hbar} \frac{e^{t(i\omega_{20} - \frac{1}{\tau})} - 1}{(i\omega_{20} - \frac{1}{\tau})} \right|^2 \\ &= \frac{|\langle 2 | x^2 | 0 \rangle|^2 A^2}{\hbar^2} \left| \frac{e^{t(i\omega_{20} - \frac{1}{\tau})} - 1}{(i\omega_{20} - \frac{1}{\tau})} \right|^2 \\ &= \frac{|\langle 2 | x^2 | 0 \rangle|^2 A^2}{\hbar^2} \frac{e^{t(i\omega_{20} - \frac{1}{\tau})} - 1}{(i\omega_{20} - \frac{1}{\tau})} \frac{e^{t(-i\omega_{20} - \frac{1}{\tau})} - 1}{(-i\omega_{20} - \frac{1}{\tau})} \\ &= \frac{|\langle 2 | x^2 | 0 \rangle|^2 A^2}{\hbar^2} \frac{e^{-\frac{2t}{\tau}} - e^{-\frac{t}{\tau}} 2\cos(\omega_{20}t) + 1}{\omega_{20}^2 + \frac{1}{\tau^2}}.\end{aligned}\quad (8.172)$$

In the limit of $t \Rightarrow \infty$

$$\lim_{t \Rightarrow \infty} (P_{0 \Rightarrow 2}(t)) \simeq \frac{|\langle 2 | x^2 | 0 \rangle|^2 A^2}{\hbar^2} \frac{1}{\omega_{20}^2 + \frac{1}{\tau^2}}. \quad (8.173)$$

This makes sense since the perturbation will vanish in this limit.

Example 8.13.

The unperturbed Hamiltonian of a two-state system is represented by

$$H_0 \doteq \begin{bmatrix} E_1^{(0)} & 0 \\ 0 & E_2^{(0)} \end{bmatrix}. \quad (8.174)$$

The two state system is perturbed by

$$H'(t) \doteq \begin{bmatrix} 0 & \lambda \cos(\omega t) \\ \lambda \cos(\omega t) & 0 \end{bmatrix}, \quad (8.175)$$

where λ is real.

1. At $t = 0$ the system is known to be in the first state, represented by

$$|1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (8.176)$$

Using TDPT and assuming that $E_1^{(0)} - E_2^{(0)}$ is not close to $\pm\hbar\omega$, derive an expression for the probability that the system is found in the second state represented by

$$|2\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (8.177)$$

as a function of time.

I use that

$$P_{1 \Rightarrow 2}(t) = |\langle 2|\alpha(t)\rangle|^2, \quad (8.178)$$

where $|\alpha(t)\rangle$ is a general state vector described by

$$|\alpha(t)\rangle = c_1(t)e^{-\frac{iE_1^{(0)}t}{\hbar}}|1\rangle + c_2(t)e^{-\frac{iE_2^{(0)}t}{\hbar}}|2\rangle. \quad (8.179)$$

Since $|1\rangle$ and $|2\rangle$ are orthogonal the probability reduces to

$$P_{1 \Rightarrow 2}(t) = |\langle 2|\alpha(t)\rangle|^2 = |c_2(t)e^{-\frac{iE_2^{(0)}t}{\hbar}}|^2 = |c_2(t)|^2. \quad (8.180)$$

To first order

$$\begin{aligned} c_2^{(1)}(t) &= \frac{1}{i\hbar} \int_0^t \langle 2|H'(t')|1\rangle e^{i\omega_{21}t'} dt' \\ &= \frac{\lambda}{i\hbar} \int_0^t \cos(\omega t) e^{i\omega_{21}t'} dt'. \end{aligned} \quad (8.181)$$

From this the probability can be found; use partial integration or see section 8.9.

Example 8.14.

Consider a two-level system with $E_1 < E_2$. There is a time-dependent potential that connects the two levels as follows

$$H' \doteq \begin{bmatrix} 0 & \gamma e^{i\omega t} \\ \gamma e^{-i\omega t} & 0 \end{bmatrix}, \quad (8.182)$$

where γ is a real constant. At $t = 0$, it is known that only the lower level is populated (i.e. $c_1(0) = 1$, $c_2(0) = 0$).

- Find $|c_1(t)|^2$ and $|c_2(t)|^2$ for $t > 0$ by exactly solving the coupled differential equation

$$i\hbar \frac{\partial c_k}{\partial t} = \sum_{n=1}^2 \langle k|H'(t)|n\rangle e^{i\omega_{kn}t} c_n, \quad (8.183)$$

where $k = 1, 2$.

I use that

$$i\hbar \frac{\partial c_1}{\partial t} = \gamma e^{i(\omega+\omega_{12})t} c_2, \quad (8.184)$$

$$i\hbar \frac{\partial c_2}{\partial t} = \gamma e^{i(\omega_{21}-\omega)t} c_1. \quad (8.185)$$

I isolate c_1 from the second equation, differentiate it and insert it in the first equation. I find

$$\frac{\partial^2 c_2}{\partial t^2} + i(\omega - \omega_{21}) \frac{\partial c_2}{\partial t} + \frac{\gamma^2}{\hbar^2} e^{i(\omega_{12}+\omega_{21})t} c_2 = 0. \quad (8.186)$$

Using now that $\omega_{12} = -\omega_{21}$

$$\frac{\partial^2 c_2}{\partial t^2} + i(\omega - \omega_{21}) \frac{\partial c_2}{\partial t} + \frac{\gamma^2}{\hbar^2} c_2 = 0. \quad (8.187)$$

Using the auxiliary equation I identify

$$\begin{aligned} a &= 1, \\ b &= i(\omega - \omega_{21}), \\ c &= \frac{\gamma^2}{\hbar^2}. \end{aligned} \quad (8.188)$$

Using then that

$$D = b^2 - 4ac = -(\omega - \omega_{21})^2 - \frac{4\gamma^2}{\hbar^2} < 0, \quad (8.189)$$

the solution is on the form

$$c_2(t) = A \cos(\omega_0 t) + B \sin(\omega_0 t), \quad (8.190)$$

where from $c_2(0) = 0 \Rightarrow A = 0$, and ω_0 is found from

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = k \pm i\omega_0. \quad (8.191)$$

Evaluating r

$$\begin{aligned} r &= 0 \pm \frac{i}{2} \left(\sqrt{(\omega - \omega_{21})^2 + \frac{4\gamma^2}{\hbar^2}} - (\omega - \omega_{21}) \right) \\ &\Rightarrow \omega_0 = \frac{1}{2} \left(\sqrt{(\omega - \omega_{21})^2 + \frac{4\gamma^2}{\hbar^2}} - (\omega - \omega_{21}) \right). \end{aligned} \quad (8.192)$$

So

$$c_2(t) = B \sin(\omega_0 t). \quad (8.193)$$

From using this in equation (8.185)

$$c_1(t) = \frac{i\hbar}{\gamma e^{i(\omega_{21}-\omega)t}} \frac{\partial c_2(t)}{\partial t} = \frac{iB\omega_0\hbar}{\gamma} e^{i(\omega-\omega_{21})t} \cos(\omega_0 t), \quad (8.194)$$

since $c_1(0) = 1 \Rightarrow B = -\frac{i\gamma}{\omega_0\hbar}$. So

$$\begin{aligned} c_1(t) &= e^{-i(\omega+\omega_{21})t} \cos(\omega_0 t), \\ c_2(t) &= -\frac{i\gamma}{\omega_0\hbar} \sin(\omega_0 t). \end{aligned} \quad (8.195)$$

The probabilities

$$\begin{aligned} |c_1(t)|^2 &= \cos(\omega_0 t)^2, \\ |c_2(t)|^2 &= \frac{\gamma^2 \sin(\omega_0 t)^2}{\hbar^2 \omega_0^2}. \end{aligned} \quad (8.196)$$

From which certain restrictions on γ can be made.

2. Do the same problem using time-dependent perturbation theory to lowest non-vanishing order. Compare the two approaches for small values of γ . Treat the following two cases separately: (i) ω very different from ω_{21} and (ii) ω close to ω_{21} .

From equation (8.44)

$$\begin{aligned} c_2^1(t) &= \frac{1}{i\hbar} \int_0^t \langle 2 | H'(t') | 1 \rangle e^{i\omega_{21}t'} dt' \\ &= \frac{\gamma}{i\hbar} \int_0^t e^{i(\omega_{21}-\omega)t'} dt' \\ &= \frac{\gamma}{\hbar(\omega_{21}-\omega)} (1 - e^{i(\omega_{21}-\omega)t}). \end{aligned} \quad (8.197)$$

The probability

$$|c_2^1(t)|^2 = \frac{\gamma^2}{\hbar^2 (\frac{\omega_{21}-\omega}{2})^2} \sin^2 \left(\frac{\omega_{21}-\omega}{2} t \right). \quad (8.198)$$

From the above $|c_2(t)|^2$ can be found from $|c_2(t)|^2 = 1 - |c_2(t)|^2$. Comparing the perturbative result with the exact result it is clear that the condition is

$$\omega_0 = \frac{1}{2} \left(\sqrt{(\omega - \omega_{21})^2 + \frac{4\gamma^2}{\hbar^2}} - (\omega - \omega_{21}) \right) \Rightarrow \frac{\omega_{21} - \omega}{2}, \quad (8.199)$$

where the arrow indicates "should be equal to". This is satisfied only when

$$(\omega - \omega_{21})^2 \gg \frac{4\gamma^2}{\hbar^2}. \quad (8.200)$$

In the case where $\omega \simeq \omega_{21}$ the exact and approximate probability reduces to

$$\omega \simeq \omega_{21}, \text{exact: } |c_2(t)|^2 \simeq \sin^2 \left(\frac{\gamma}{\hbar} t \right), \quad (8.201)$$

$$\omega \simeq \omega_{21}, \text{approx: } |c_2^{(1)}(t)|^2 \simeq \frac{\gamma^2 t^2}{\hbar^2}. \quad (8.202)$$

As is apparent from above the approximate solution is not bound to below one, and as time goes it can go above 1 and become unphysical. The two are however equivalent at $t \sim 0$.

For ω very different from ω_{21} the relation $(\omega - \omega_{21})^2 \gg \frac{4\gamma^2}{\hbar^2}$ is satisfied and the two results will be equal.

Example 8.15.

A hydrogen atom in its ground state $[(n,l,m=(1,0,0))]$ is placed between the plates of a capacitor. A time-dependent but spatially uniform electric field (not potential!) is applied as follows

$$\vec{E} = \begin{cases} 0 & \text{for } t < 0 \\ \vec{E}_0 e^{-\frac{t}{\tau}} & \text{for } t \geq 0 \end{cases}. \quad (8.203)$$

Using first-order time-dependent perturbation theory, compute the probability for the atom to be found at $t \gg \tau$ in each of the three $2p$ states: $(n,l,m) = (2,1,\pm 1 \text{ or } 0)$. Repeat the problem for the $2s$ state: $(n,l,m) = (2,0,0)$. You need not attempt to evaluate radial integrals, but perform all other integrations (with respect to angles and time).

Assuming the electric field is applied in the z -direction, the perturbative term added to the Hamiltonian is given from the linear Stark effect

$$H'(t) = -ez|\vec{E}_0|e^{-\frac{t}{\tau}}. \quad (8.204)$$

Thereby the first order correction

$$\begin{aligned} c_f^1(t) &= \frac{1}{i\hbar} \int_0^t \langle f | H'(t') | 1, 0, 0 \rangle e^{i\omega_0 t'} dt' \\ &= \frac{-e|\vec{E}_0|}{i\hbar} \int_0^t \langle f | z | 1, 0, 0 \rangle e^{(i\omega_0 - \frac{1}{\tau})t'} dt', \end{aligned} \quad (8.205)$$

where $|f\rangle \in \{|2, 1, 1\rangle, |2, 1, -1\rangle, |2, 1, 0\rangle\}$ for the $2p$ states. Cf. the arguments made in section 8.6¹⁸ the probability for the system to end up in $|2, 1, 1\rangle$ and $|2, 1, -1\rangle$ is zero. Thereby on the case of $1s \Rightarrow 2p$ there is only one non-vanishing probability to end up in one of the given states. This is given as the following coefficient squared

$$\begin{aligned} c_{2,1,0}^1(t) &= \frac{-e|\vec{E}_0|}{i\hbar} \int_0^t \langle 2, 1, 0 | z | 1, 0, 0 \rangle e^{(i\omega_0 - \frac{1}{\tau})t'} dt' \\ &= \frac{-e|\vec{E}_0| \langle 2, 1, 0 | z | 1, 0, 0 \rangle}{i\hbar} \int_0^t e^{(i\omega_0 - \frac{1}{\tau})t'} dt' \\ &= \frac{-e|\vec{E}_0| \langle 2, 1, 0 | z | 1, 0, 0 \rangle}{i\hbar(i\omega_0 - \frac{1}{\tau})} \left(e^{(i\omega_0 - \frac{1}{\tau})t} - 1 \right). \end{aligned} \quad (8.206)$$

So, the probability for a transition from $|1, 0, 0\rangle$ to $|2, 1, 0\rangle$ is given by

$$\begin{aligned} |c_{2,1,0}^1(t)|^2 &= \frac{e^2 |\vec{E}_0|^2 |\langle 2, 1, 0 | z | 1, 0, 0 \rangle|^2}{\hbar^2 \omega_0^2 + \frac{\hbar^2}{\tau^2}} \left(e^{(i\omega_0 - \frac{1}{\tau})t} - 1 \right) \left(e^{(-i\omega_0 - \frac{1}{\tau})t} - 1 \right) \\ &= \frac{e^2 |\vec{E}_0|^2 |\langle 2, 1, 0 | z | 1, 0, 0 \rangle|^2}{\hbar^2 \omega_0^2 + \frac{\hbar^2}{\tau^2}} \left(1 + e^{-\frac{2t}{\tau}} - e^{-\frac{t}{\tau}} \cos(\omega_0 t) \right). \end{aligned} \quad (8.207)$$

In the limit of $t \gg \tau$ the exponential terms will go to zero. Therefore

$$t \gg \tau : \quad |c_{2,1,0}^1(t)|^2 \simeq \frac{e^2 |\vec{E}_0|^2 |\langle 2, 1, 0 | z | 1, 0, 0 \rangle|^2}{\hbar^2 \omega_0^2 + \frac{\hbar^2}{\tau^2}}. \quad (8.208)$$

The transition $|1, 0, 0\rangle \Rightarrow |2, 0, 0\rangle$ is also forbidden (probability is zero) since, cf. (Sakurai, 1994, p.322)

$$\langle n', l', m' | z | n, l, m \rangle = 0 \quad \text{unless } \begin{cases} l' = l \pm 1 \\ m' = m \end{cases}. \quad (8.209)$$

The criterion $m = m'$ comes from the fact that z can be written as a spherical tensor, $T_0^{(1)} = z$, and the matrix element of one such is zero unless $m = m' \pm q$ where $q = 0$ in this case. The criterion $l' = l \pm 1$ comes from parity. The above criterion are called *selection rules*.

Example 8.16.

I consider a bumpy potential in the bottom of an infinitely deep potential well. The potential function is given by

$$V(x) = \begin{cases} V(x) & \text{for } 0 < x < a \\ \infty & \text{otherwise} \end{cases}. \quad (8.210)$$

¹⁸Regarding the matrix element of angular momentum eigenstates with respect to z . Expressing z as im terms of spherical tensors, and seeing that unless $m = m' \pm q$ the matrix element is zero. Now, since $T_0^{(1)} = z$ only $m = m'$ is non-zero.

The potential is shown in figure 8.7.

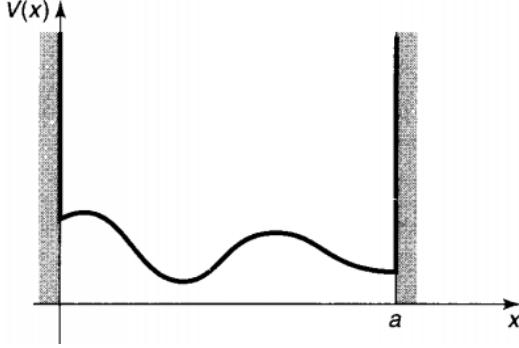


Figure 8.7: The imagined potential in the bottom of an infinite well.

I assume that $p(x)$ is real, the classical case, and thus $E > V(x)$ from equation (8.49). From equation (8.57) I find

$$\psi(x) \simeq \frac{1}{\sqrt{p(x)}}(C_+ e^{i\phi(x)} + C_- e^{-i\phi(x)}), \quad (8.211)$$

where

$$\phi(x) = \frac{1}{\hbar} \int_0^x p(x') dx'. \quad (8.212)$$

Using that $e^{\pm ix} = \cos(x) \pm i\sin(x)$ I rewrite equation (8.211)

$$\psi(x) \simeq \frac{1}{\sqrt{p(x)}}((C_+ + C_-)\cos(\phi(x)) + (C_+ - C_-)\sin(\phi(x))). \quad (8.213)$$

Since $\psi(x) \Rightarrow 0$ at $x = 0$ $C_+ + C_- = 0$. Likewise, $\psi(x) \Rightarrow 0$ at $x = a$, which results in $\phi(x) = n\pi$. Thus

$$\psi(x) \simeq \frac{1}{\sqrt{p(x)}}(C_+ - C_-)\sin(\phi(x)), \quad (8.214)$$

where

$$\int_0^a p(x) dx = n\pi\hbar. \quad (8.215)$$

The quantization determines the approximate energy eigenvalues. They are determined by evaluating $p = \sqrt{2M(E - V(x))}$ and $\int_0^a p(x) dx = n\pi\hbar$ for $p(x)$ and equating the results. With $\phi(x)$ evaluated wave-function in equation (8.214) is also determined.

Example 8.17.

In this example I consider quantum tunneling, which corresponds to the non-classical case where $E < V$ and thereby $p(x)$ is imaginary. By setting $E < V$ I find from equation (8.57)

$$\psi(x) \simeq \frac{C}{\sqrt{|p(x)|}} e^{\pm \frac{1}{\hbar} \int |p(x)| dx}. \quad (8.216)$$

Quantum tunneling is non-classical because a particle of lower energy travels through a larger potential barrier. The probability of traveling through the barrier can be found by considering the wave function hitting the(a rectangular) barrier. I consider the scenario shown in figure 8.8.

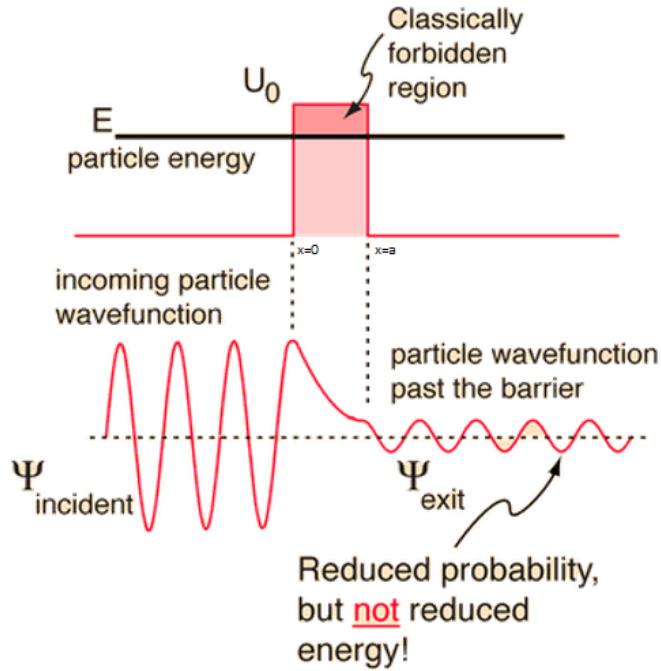


Figure 8.8: Quantum tunneling.

Analog to classical wave-physics, the involved waves will be an incident wave (I) a reflected wave(R) and a transmitted wave(T). I define $x = 0$ where the wall starts, and $x = a$ where it ends. At $x < 0$ the the incident and reflected wave is described by the wave function¹⁹

$$\psi_{x<0}(x) = I e^{ik_1 x} + R e^{-ik_1 x}, \quad (8.217)$$

where the subscript " $x < 0$ " denotes that this is the wave function for $x < 0$, I and R are the amplitudes of the incident and reflected wave, respectively, and²⁰ $k_1 = \frac{\sqrt{2M(E - U_0)}}{\hbar}$. I write the wave function in the tunneling region, $0 < x < a$, as (with constant potential, E_0)

$$\psi_{0 < x < a}(x) = C_1 e^{ik_2 x} + C_2 e^{-ik_2 x}, \quad (8.218)$$

where $k_2 = \frac{\sqrt{2M(E_0 - E)}}{\hbar}$. I write the transmitted wave at $x > a$ as

$$\psi_{x>a}(x) = T e^{ik_1 x}. \quad (8.219)$$

From my definitions the transmission probability is given by

$$P_{trans} = \frac{|T|^2}{|I|^2}. \quad (8.220)$$

I find P_{trans} by evaluating the amplitudes T and I via imposing boundary conditions. I require continuity at the boundaries

$$\psi_{x<0}(0) = \psi_{0 < x < a}(0), \quad (8.221)$$

$$\frac{\partial \psi_{x<0}(0)}{\partial x} = \frac{\partial \psi_{0 < x < a}(0)}{\partial x}, \quad (8.222)$$

$$\psi_{0 < x < a}(a) = \psi_{x>a}(a), \quad (8.223)$$

$$\frac{\partial \psi_{0 < x < a}(a)}{\partial x} = \frac{\partial \psi_{x>a}(a)}{\partial x}. \quad (8.224)$$

¹⁹The potential is constant, so the integral just evaluates to x .

²⁰At $x < 0$ $V(x) = 0$. I have used this in $p = \sqrt{2M(E - V(x))}$ along with $k = \frac{p}{\hbar}$.

From the BC I find

$$R = \frac{1}{2} \left(C_2 \left(1 - \frac{ik_2}{k_1} \right) + C_1 \left(1 + \frac{ik_2}{k_1} \right) \right), \quad (8.225)$$

$$I = \frac{1}{2} \left(C_2 \left(1 + \frac{ik_2}{k_1} \right) + C_1 \left(1 - \frac{ik_2}{k_1} \right) \right), \quad (8.226)$$

$$C_2 = \frac{B_3}{2} \left(1 - \frac{ik_1}{k_2} \right) e^{-(ik_1+k_2)a}, \quad (8.227)$$

$$C_1 = \frac{B_3}{2} \left(1 + \frac{ik_1}{k_2} \right) e^{(k_2-ik_1)a}. \quad (8.228)$$

Thus, I find the transmission probability (P_{trans}) as (Martin, 2006, p.340)

$$P_{trans} = \left| \frac{2k_1 k_2 e^{-k_1 a}}{2k_1 k_2 \cosh(k_2 a) - i(k_1^2 - k_2^2) \sinh(k_2 a)} \right|^2. \quad (8.229)$$

For large $k_2 a$, which corresponds to small penetrations, I can make the small angle approximations

$$\sinh(k_2 a) \approx \cosh(k_2 a) \approx \frac{e^{k_2 a}}{2}. \quad (8.230)$$

Thus, equation (8.229) reduces to

$$P_{trans} \approx \left(\frac{4k_1 k_2}{k_1^2 + k_2^2} \right)^2 e^{-2k_2 a}. \quad (8.231)$$

The first factor is down to reflection losses at the two boundaries ($x = 0$ and $x = a$). The exponential describes the exponential decay of the amplitude of the wave function inside the barrier. The first factor varies slowly with energy and is therefore often neglected. Thus, the transmission probability is written as

$$P_{trans} \approx e^{-2k_2 a}. \quad (8.232)$$

Taking into account that the potential could be non-constant in the tunneling region, I find

$$P_{trans} \approx e^{-\frac{2}{\hbar} \int_0^a |p(x)| dx} = e^{-\frac{2}{\hbar} \int_0^a |\sqrt{2M(V(x)-E)}| dx}, \quad (8.233)$$

where $V(x)$ and E has switched places because of the imaginary unit is "pulled out" and I consider the magnitude. From equation (8.233) it is clear that the incoming wave decays approximately exponentially inside the barrier, so the relationship between the incident and outgoing wave is the exponential decay. This squared gives the (approximate) transmission probability. The transmission probability is often used eg. in α particle decay or in nuclear fusion.

Part III

Relativistic Mechanics

Introduction

The subject of relativistic mechanics is the realm of classical fields in the relativistic limit. Classical because the theory describes macroscopic objects and fields because the laws of nature must be *local*. The non-relativistic physical laws, of eg. Newton and Coulomb, employ "action at a distance". The principle of action at a distance approximates the speed of light to be infinite, and so information is mediated instantaneously between objects. This approximation is good when considering systems in the non-relativistic limit, but not in the relativistic limit. In the relativistic limit fields are needed to mediate the interactions, and besides that special relativity, or general relativity, is needed to describe the physics in a Lorentz covariant way. SR is closely related to group theory, and so the place to begin is really with an introduction to group theory ending with an introduction of the group used in special relativity; the Lorentz group.

9 | Group Theory and Special Relativity

Group theory concerns the classification of mathematical similarities between sets of elements and connects these to physical symmetries.

Definition of a group: A set of elements, $a, b, c, \dots \in G$, is said to form a group if there, between two arbitrary elements of the group, exists a well defined product that satisfies:

1. The product of two elements of the group is another element the group $a \cdot b = c, \forall a, b : \exists c : a, b, c \in G$.
2. The group product between several elements is associative, i.e. $a \cdot (b \cdot c) = (a \cdot b) \cdot c, \forall a, b, c \in G$.
3. The identity element exists within the group, i.e. $e \in G$. The identity is defined by the operation on other, generic elements via $a \cdot e = e \cdot a = a, \forall a \in G, e \in G$.
4. There is an inverse to each group element, such that $a^{-1} \cdot a = a \cdot a^{-1} = e, \forall a \in G, e \in G$.

An abelian group: A group is said to be abelian if the product of group elements is commutative, i.e. $a \cdot b = b \cdot a, \forall a, b \in G$.

Order of a group: The order of a group is the number of elements in the group.

Subgroup: If a subset, H , of a group, G , fulfill the requirements for a group, then the subset forms a subgroup.

Continuous group: If the group elements carry a label, which is a continuous parameter, the group is said to be continuous. A continuous group of special interest to physics is the Lie group. A Lie group is a group which is also a differentiable manifold. In mathematics, a manifold is a topological space that locally resembles Euclidean space near each point. This means that a geometrical structure can be associated with Lie groups and the group can be associated with such a structure in some abstract space. Further more, the elements of a Lie group must be parameterized smoothly and continuously.

Example 9.1.

1. The general linear group, of dimension n , $GL(n)$, can be represented as the set of all invertible $n \times n$ matrices.
2. The unitary group, of dimension n , $U(n)$, can be represented as the set of $n \times n$ unitary matrices. Ie. matrices satisfying $UU^\dagger = U^\dagger U = I$.
3. The special unitary group, of dimension n , $SU(n)$, can be represented as the set of unitary matrices with unit determinant.

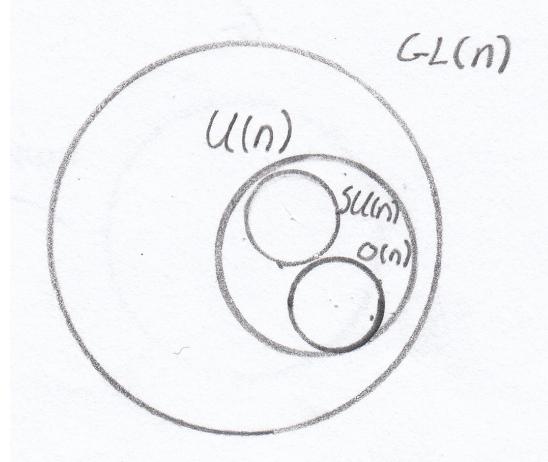


Figure 9.1: The relationship between the different groups mentioned above.

4. The orthogonal group, of dimension n , $O(n)$, can be represented as the set of real, orthogonal matrices. Ie. matrices satisfying $M^T = M^{-1}$. The different groups are connected via the relation

$$SU(n) \wedge O(n) \in U(n) \in GL(n). \quad (9.1)$$

The relationship is also seen illustrated in figure 9.1.

Homomorphism: A homomorphism is a mapping between two groups, G and G' , that conserve group multiplication. The map does not have to be one-to-one. An illustration of the mapping is seen in figure 9.2.

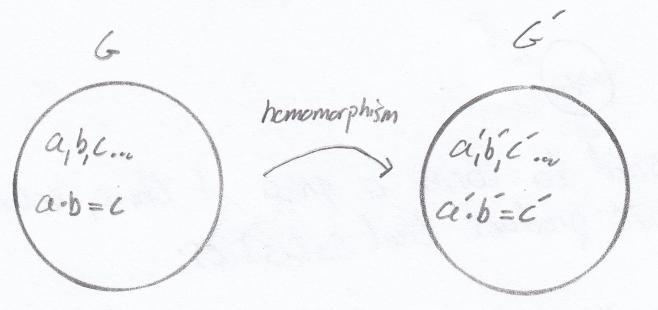


Figure 9.2: An illustration of a homomorphism between two groups.

Isomorphism: An isomorphism is a homomorphism between two groups, G and G' , with a one-to-one correspondence. Hence, not only group multiplication is preserved, but also

$$a \rightleftharpoons a', \quad b \rightleftharpoons b', \quad c \rightleftharpoons c', \dots \quad (9.2)$$

Conjugate elements of a group: Elements of a group, G , $a, b \in G$, is said to be conjugate to each other if there exist another element, $p \in G$, such that $b = pap^{-1}$.

Direct product group: Let H_1 and H_2 be subgroups of G which satisfies:

1. All elements of H_1 and H_2 commute, i.e. $h_1 h_2 = h_2 h_1, \forall h \in H_1 \wedge h_2 \in H_2$.
2. All elements of G can be described as the product of an element from H_1 and an element from H_2 . Ie. $h_1 h_2 = g \in G, \forall g$.

In this case G is said to be the direct product of H_1 and H_2 . This is denoted by $G = H_1 \otimes H_2$.

Group representation: If there is a homomorphism¹ from a group, G , to a set of operators, $U(g)$ (where g are the group elements of G), defined on a linear vector space, V , then $U(g)$ is a representation² of G . The concept of a representation is shown in figure 9.3. In terms of representations of Lie groups it is important

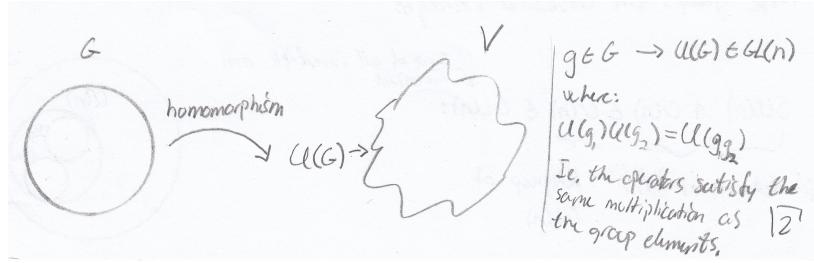


Figure 9.3: An illustration of a representation of a group.

to remember that a Lie group is defined in terms of the associated manifold. For this reason one can talk about different representations of groups made up by matrices³ (eg. $SU(2)$). The representation of a Lie group is a representation of the associated manifold.

Invariant subspace: Consider a representation, $U(g)$, of a group, G , defined on a linear vector space V . $V' \subseteq V$ is called an invariant subspace if for $v \in V'$; $U(g)v \in V', \forall g \in G$. This means that if a vector in the invariant subspace V' act on an arbitrary group element the transformed vector will always be again a part of the invariant subspace, V' . For an invariant subspace a representation $U'(g)$ of G on V' , called a sub-representation of $U(g)$, can be defined viz

$$U'(g)v = U(g)v, \quad \forall v \in V'. \quad (9.3)$$

This leads to the notion that the representation $U(g)$ is not a fundamental but instead composed of smaller building blocks.

Irreducible representations: An irreducible representation is a representation of a group G on a vector space V that has no invariant subspaces besides 0 and V itself. Such representations can be thought of as truly fundamental since they are not made up of smaller representations. The irreducible representations of a group are the building blocks from which any other representation of the group can be built. A different approach to irreducible representations can be found by considering similarity transformations. Given a matrix, R , and an invertible matrix, S , the similarity transformation is defined viz

$$R \Rightarrow R' = S^{-1}RS. \quad (9.4)$$

If R is a representation of the group G , then so is R' . An irreducible representation can be defined as a representation that cannot be transformed into block diagonal form via a similarity transformation.

Generators of a discrete group: A collection, c , of elements of a discrete group, G , is said to be generators of the group if multiplication among the generators yield all elements of the group.

¹A homomorphic representation is called unfaithful whereas an isomorphic representation is called faithful.

²Informally the representation of a group is a way of writing the group as matrices.

³It is a confusing thought to think of representing matrices of a given dimension by other matrices with different dimensions. However, one should think of it like this; a given Lie group has an associated manifold. This manifold can have several representations. For example; $SU(2)$ is the group of 2×2 unitary matrices with unit determinant. The manifold of this group is a circle. The manifold can be represented by either 2×2 matrices or complex numbers (as is commonly done with the unit circle in complex space).

Generators of a continuous group: In the case of a continuous group the generators are defined different, compared to the discrete group, since there are no single elements in a continuous group, but a continuum of elements. Consider a set of elements, R , that depends on some real, continuous parameters

$$R = R(x, y, z, \dots). \quad (9.5)$$

These elements form a continuous group if R fulfil the requirements for a group. If R is differentiable, the group is a Lie group. Differentiability means that $R(dx)$ only differs infinitesimally from $R(0) = e$ by a term of order dx . Therefore $R(dx)$ can be parametrized as

$$R(dx) = e - idxJ, \quad (9.6)$$

where e is the identity element, the " $-i$ " is included by convention (to make J Hermitian) and J is *defined* as the generator of the group. By going from $dx \Rightarrow x$

$$R(x) = \lim_{N \rightarrow \infty} \left(e - i \frac{dxJ}{N} \right)^N = e^{-ixJ}. \quad (9.7)$$

In the considered case, used for illustrative purposes, there is only one parameter (x) and one corresponding generator (J). In general, a group can have several parameters (α_i) and corresponding generators (t_i). The different generators make up a vector space. This vectors space is called a Lie algebra.

Lie algebra: A Lie algebra is a vector space, \mathfrak{g} , equipped with a binary operation (called the Lie bracket⁴), $[,] : \mathfrak{g} \times \mathfrak{g} \Rightarrow \mathfrak{g}$. The binary operation satisfies the following axioms:

1. Bilinearity: $[aX + bY, Z] = a[X, Z] + b[Y, Z]$ and $[Z, aX + bY] = a[Z, X] + b[Z, Y]$, for arbitrary numbers a, b and $\forall X, Y, Z \in \mathfrak{g}$.
2. Anticommutativity: $[X, Y] = -[Y, X] \forall X, Y \in \mathfrak{g}$.
3. The Jacobi identity: $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0 \forall X, Y, Z \in \mathfrak{g}$.

The Lie algebra is therefore a vector space consisting of the generators of a Lie group. However, it is important to note that the definition of the Lie algebra makes no reference to any Lie group. Therefore the definition stands on its own and therefore one can talk about different Lie groups having the same Lie algebra. This leads to the definition of a distinguished Lie group.

Distinguished Lie group: There is precisely one distinguished Lie group for each Lie algebra. The distinguished group has the property of being simply connected. This means that any closed curve on the manifold, corresponding to the Lie group, can be shrunk smoothly to a point. The simply connected group can be thought of as the "mother" of all those groups having the same Lie algebra, because there are maps to all other groups with the same Lie algebra from the simply connected group, but not vice versa. An intuitive name could be the "mother" group of a given Lie algebra. Formally the group is called the *covering group*. All other groups with the same Lie algebra is said to be covered by the simply connected group.

Example 9.2.

Consider the group $SO(2)$. It is a one dimensional group describing rotations in 2 dimensions. The group is one dimensional since only one continuous parameter, θ , characterizes a rotation. The elements of the group can be represented by 2×2 matrices which commute, and so the group is an abelian group. The group has one generator, L , which can be used to generate all elements of the group via

$$R(\theta) = e^{-i\theta L}. \quad (9.8)$$

Since the group is designated by $SO(2)$ each matrix is orthogonal and have unit determinant.

⁴The Lie bracket is often called the Lie algebra.

Example 9.3.

This example will consider the isomorphism between $U(1)$ and $SO(2)$. As mentioned in example 4.2 $SO(2)$ describes rotations in two dimensions and the group elements can be represented, in terms of the generator L , as

$$R(\theta) = e^{-i\theta L}. \quad (9.9)$$

$U(1)$ is the manifold of unitary 1×1 matrices, i.e. unit complex numbers. The generator of the groups is the identity, e , and so the elements can be represented as

$$R(x) = e^{-ix} \quad (9.10)$$

There is an isomorphism between the two groups. This can be shown by considering

$$R(\theta_1)R(\theta_2) = e^{-i\theta_1 L}e^{-i\theta_2 L} = e^{-i(\theta_1 + \theta_2)L} = R(\theta_1 + \theta_2), \quad (9.11)$$

$$f(x_1)f(x_2) = e^{-ix_1}e^{-ix_2} = e^{-i(x_1 + x_2)} = f(x_1 + x_2). \quad (9.12)$$

It is clear that there is a one-to-one correspondence between x and θ , and so $U(1)$ and $SO(2)$ are isomorphic to each other. The manifold associated with unit complex numbers is the unit circle in the complex plane. Since there is an isomorphism between $U(1)$ and $SO(2)$, the manifold of $SO(2)$ is likewise the unit circle in the complex plane. Hence, one can think of the Lie group as being the unit circle and the manifestation of the elements in complex numbers ($U(1)$) and matrices ($SU(2)$) as representations of the manifold.

9.1 $SO(3)$

The group $SO(3)$ manifold corresponding to orthogonal 3×3 matrices with unit determinant. The geometric shape of the manifold will be considered in the next section (on $SU(2)$). The reason for this is that $SU(2)$ is the simply connected group of the Lie algebra which it shares with $SO(3)$. Hence, the manifold of $SO(3)$ is a part of the manifold of $SU(2)$.

The matrices describe rotations in $3d$ space. Formally; the $SO(3)$ group consists of all continuous, linear transformations in $3d$ Euclidean space which leave the length of coordinate vectors invariant. The elements of $SO(3)$ does not, in general, commute and so $SO(3)$ is, as opposed to eg. $U(1)$ or $SO(2)$, a non-abelian Lie group.

A generic rotation in $3d$ is parametrized by 3 rotations, by 3 continuous angles, around the three coordinate axis. Corresponding to each continuous parameter is a generator, and so $SO(3)$ has 3 basis generators. The rotations around a given axis form a subgroup of $SO(3)$ that is isomorphic to $SO(2)$. Since there are 3 coordinate axis' there are 3 basis generators, denoted by J_i . the Lie bracket of the generators is given by

$$[J^i, J^j] = i\epsilon^{ijk} J^k. \quad (9.13)$$

From the basis generators a generic element of $SO(3)$ can be described as

$$R(\alpha, \beta, \gamma) = R_3(\gamma)R_2(\beta)R_1(\alpha) = e^{-iJ_3\gamma}e^{-iJ_2\beta}e^{-iJ_1\alpha}, \quad (9.14)$$

where α, β, γ are *Euler angles* and the basis matrices R_i are given by

$$\begin{aligned} R_1(\alpha) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{bmatrix}, \\ R_2(\beta) &= \begin{bmatrix} \cos(\beta) & 0 & -\sin(\beta) \\ 0 & 1 & 0 \\ \sin(\beta) & 0 & \cos(\beta) \end{bmatrix}, \\ R_3(\gamma) &= \begin{bmatrix} \cos(\gamma) & \sin(\gamma) & 0 \\ -\sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned} \quad (9.15)$$

where R_i rotate the coordinate system, not the vector. It is a choice of convention. Rotating the vector is called an active transformation whereas rotating the coordinate system is called a passive transformation. Since the placement of a coordinate system is entirely up to the physicist, i.e. the coordinate system is not a physical thing, it makes more intuitive sense to alternate your definition than actually going in and rotating the physical vector. Hence, here the approach of passive transformations is used.

9.2 $SU(2)$

The group $SU(2)$ is manifold corresponding to unitary 2×2 matrices with unit determinant. The manifold of $SU(2)$ is found by extending the method used to discover the manifold of $SO(2)$. In discovering the manifold of $SO(2)$ it was used that $SO(2)$ is isomorphic to $U(1)$, i.e. sharing manifold, and that $U(1)$ consists of two dimensional complex, unitary numbers which are defined on the unit circle in the complex plane. Extending this analogy the manifold of $SU(2)$ can be found as the manifold of 4 dimensional complex, unitary numbers; quaternions. To describe 4 dimensional complex numbers three imaginary units are needed - As opposed to one imaginary unit for two dimensional complex numbers (i). The three imaginary units are defined viz

$$i^2 = j^2 = k^2 = ijk = -1. \quad (9.16)$$

Then a 4 dimensional complex number, a quaternion, can be parametrized viz

$$q = a + bi + cj + dk. \quad (9.17)$$

The set of unit quaternions are those that satisfy

$$qq^\dagger = a^2 + b^2 + c^2 + d^2 = 1. \quad (9.18)$$

Just as complex numbers can be represented as matrices by mapping the complex unit (i) to a matrix, so can quaternions. Specifically (Schwichtenberg, 2015)

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \hat{i} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \hat{j} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \hat{k} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}. \quad (9.19)$$

This means the generic quaternion can be parametrized viz

$$q = aI + b\hat{i} + c\hat{j} + d\hat{k} = \begin{bmatrix} a + ib & c + id \\ -c + id & a - ib \end{bmatrix} \Rightarrow \det(q) = a^2 + b^2 + c^2 + d^2. \quad (9.20)$$

Comparing equation (9.20) to the definition of unit quaternions ($qq^\dagger = 1$), it is clear that the matrices must be unitary and have unit determinant. That is the matrices make up the group $SU(2)$. Hence, unit quaternions are isomorphic to $SU(2)$ just like $U(1)$ is isomorphic to $SO(2)$. Hence, the manifold of $SU(2)$ can be described by the unit quaternions. The unit quaternions are characterized by $a^2 + b^2 + c^2 + d^2 = 1$, analogous to how 2 dimensional unit complex numbers are characterized by $z = a + ib \Rightarrow a^2 + b^2 = 1$. $a^2 + b^2 = 1$ describes the unit circle, S^1 . Hence, the manifold of two dimensional unit complex numbers is the unit circle in the complex plane. Likewise $a^2 + b^2 + c^2 + d^2 = 1$ describes the unit three sphere, S^3 . Hence, the manifold of four dimensional unit complex numbers is the unit three sphere. The unit three sphere is a simply connected manifold. Hence, $SU(2)$ is simply connected.

9.2.1 Relationship between $SO(3)$ and $SU(2)$

To relate the two groups it will be investigated how quaternions can be used to rotate a vector. Parametrizing a vector viz

$$v = v_x\hat{i} + v_y\hat{j} + v_z\hat{k} = \begin{bmatrix} iv_x & v_y + iv_z \\ -v_y + iv_z & -iv_x \end{bmatrix} \Rightarrow \det(v) = x^2 + y^2 + z^2. \quad (9.21)$$

The determinant of v is the length of the vector. Since the length of a vector is conserved during a rotation, the transformation must preserve determinants. This means restricting ourselves to unit quaternions as a means to transform the vector

$$v \Rightarrow v' = q^{-1}vq, \quad (9.22)$$

where once again the approach of passive transformations has been applied. The inverse transformation, qvq^{-1} , is the corresponding active rotation⁵. In order to relate the transformation of equation (9.22) to rotations the rotation angles must be introduced. In this regard it is convenient to parametrize the unit quaternion viz

$$q = \cos(\theta) + \sin(\theta)||\vec{e}||, \quad (9.23)$$

where $||\vec{e}||$ is the normalized rotation axis. For transparency consider a rotation of $\frac{\vec{x}}{||\vec{x}||} \equiv \hat{x}$ around the z -axis. In this case

$$q = \cos(\theta)I + \sin(\theta)k = \begin{bmatrix} \cos(\theta) & i\sin(\theta) \\ i\sin(\theta) & \cos(\theta) \end{bmatrix}. \quad (9.24)$$

Using equation (9.24)

$$q^{-1}\hat{i}q = \begin{bmatrix} \cos(\theta) & -i\sin(\theta) \\ -i\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \begin{bmatrix} \cos(\theta) & i\sin(\theta) \\ i\sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} i\cos(2\gamma) & -\sin(2\theta) \\ \sin(2\theta) & -i\cos(2\theta) \end{bmatrix}. \quad (9.25)$$

Comparing equation (9.25) to equation (9.21)

$$\begin{aligned} v_x &\Rightarrow v'_x = \cos(2\theta), \\ v_y &\Rightarrow v'_y = -\sin(2\theta), \\ v_z &\Rightarrow v'_z = 0. \end{aligned} \quad (9.26)$$

Rotating a vector in 3-space reveals, cf. equation (9.15)

$$R_3(\gamma)\hat{x} \begin{bmatrix} \cos(\gamma) & \sin(\gamma) & 0 \\ -\sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos(\gamma) \\ -\sin(\gamma) \\ 0 \end{bmatrix}. \quad (9.27)$$

By comparing equation (9.26) and (9.27) it is clear that $\gamma = 2\theta$. That is, in rotating \hat{x} by θ the vector is rotated by 2γ in 3-space. This means that there is a two-to-one map from the manifold of $SU(2)$ to the manifold of $SO(3)$, i.e. there are two elements in the manifold of $SU(2)$ that corresponds to the same element in the manifold of $SO(3)$. Hence given one element of the manifold of $SU(2)$ one can find the corresponding element in the manifold of $SO(3)$ but not vice versa. This confirms that the manifold of $SU(2)$ is indeed simply connected. Since there is a two-to-one correspondence between the manifolds of $SU(2)$ and $SO(3)$, and the manifold of $SU(2)$ is the unit three sphere, the manifold of $SO(3)$ must be half a unit three sphere. Since the manifold of $SU(2)$ covers the manifold of $SO(3)$ twice it is called the double cover of the manifold of $SO(3)$.

9.3 The Lorentz group

The Lorentz group, denoted by⁶ $O(3, 1)$, is the group representing the coordinate transformations of SR; boosts and rotations that is. Because the Lorentz group describes the transformations of SR, which are transformations of four vectors, the Lorentz groups is often analysed in a certain representation; the vector representation. In the vector representation an event in SR is described by a four vector on the form

$$x^\mu \equiv \begin{bmatrix} t \\ \vec{x} \end{bmatrix}. \quad (9.28)$$

⁵ qvq^{-1} is just a the inverse transformation of $q^{-1}vq$. This makes intuitively sense since rotating a vector in one direction is equivalent to rotating the coordinate system in the opposite direction. Hence, the definition of active and passive rotations is related to the definition of positive and negative angles also.

⁶Three space-like dimensions and one time-like dimension.

This vector is called contravariant (index up) and it is defined by its transformation properties

$$dx^\mu \Rightarrow dx'^\mu = \underbrace{\frac{\partial x'^\mu}{\partial x^\nu}}_{\Lambda^\mu_\nu} dx^\nu. \quad (9.29)$$

Any vector that transforms in this way is called a contravariant vector. Likewise, a covariant vector is defined by

$$dx_\mu \Rightarrow dx'_\mu = \underbrace{\frac{\partial x'_\mu}{\partial x_\nu}}_{(\Lambda^{-1})^\nu_\mu = \Lambda_\mu^\nu} dx_\nu. \quad (9.30)$$

The contra-variant and covariant vectors are related by the metric via

$$x^\mu = g^{\mu\nu} x_\nu. \quad (9.31)$$

During the coordinate transformations, the length of the vectors, $x^\mu x_\mu \equiv x \cdot x = x^2$, is invariant. The length of a vector can be expressed via the metric

$$x^\mu x_\mu = g_{\mu\nu} x^\mu x^\nu = g_{\mu\nu} x'^\mu x'^\nu = g_{\mu\nu} \Lambda^\nu_\alpha \Lambda^\mu_\beta x^\alpha x^\beta. \quad (9.32)$$

Rearranging the dummy indices

$$g_{\mu\nu} x^\mu x^\nu = g_{\alpha\beta} \Lambda^\alpha_\nu \Lambda^\beta_\mu x^\nu x^\mu \Rightarrow g_{\mu\nu} = g_{\alpha\beta} \Lambda^\alpha_\nu \Lambda^\beta_\mu, \quad (9.33)$$

which can be written $g = \Lambda^T g \Lambda$ in matrix notation. Taking the determinant, and using that $g_{\mu\nu} = \eta_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ for Minkowski space

$$\det(\eta) = \det(\Lambda^T) \det(\eta) \det(\Lambda) \Rightarrow \det(\Lambda^T) \det(\Lambda) = 1. \quad (9.34)$$

Use now that $\det(\Lambda^T) = \det(\Lambda)$ to find $\det(\Lambda)^2 = 1 \Rightarrow \det(\Lambda) = \pm 1$. Transformations for which $\det(\Lambda) = 1$ are called proper. Transformations for which $\det(\Lambda) = -1$ are called improper. Another classification comes from

$$g_{00} = 1 = \Lambda^0_0 \Lambda^0_0 g_{\alpha\beta} = (\Lambda^0_0)^2 - \sum_{i=1}^3 (\Lambda^i_0)^2 \Rightarrow |\Lambda^0_0| > 1. \quad (9.35)$$

This allows $\Lambda^0_0 > 1$ or $\Lambda^0_0 < -1$. The transformations characterized by $\Lambda^0_0 > 1$ are called orthochronous whereas transformations characterized by $\Lambda^0_0 < -1$ are called non-orthochronous. Hence, the Lorentz transformations can be grouped into 4 categories⁷ shown in figure 9.4.

	$\Lambda^0_0 \geq 1$ orthochronous	$\Lambda^0_0 \leq -1$ non-orthochronous
$\det \Lambda = +1$ proper	$\Lambda^{(po)}$	$TP\Lambda^{(po)}$
$\det \Lambda = -1$ improper	$P\Lambda^{(po)}$	$T\Lambda^{(po)}$

Figure 9.4: The different categories of Lorentz transformations.

As shown in figure 9.4 the different elements of the categories can all be expressed in terms of proper, orthochronous transformations followed by a time-inversion and/or parity operation. Parity is called the "mirror-transformation". Until 1956 physicists believed all physical systems should be invariant under parity.

⁷That this is correct can be seen by considering the $TP\Lambda^{po}$ -term; the determinant must be positive since it is proper. The determinant of Λ^{po} is defined as positive and since the determinant is distributive and negative for both T, P , the table must be as shown. The table is confirmed by Srednicki (2007).

However, in 1956 it was found that the weak interaction violates parity. Mathematically, parity is expressed viz

$$\Lambda_P x = \begin{bmatrix} t \\ -\vec{x} \end{bmatrix}, \quad \text{and} \quad \Lambda_P p = \begin{bmatrix} p^0 \\ -\vec{p} \end{bmatrix}, \quad (9.36)$$

where $\Lambda_P = \text{diag}(1, -1, -1, -1)$ in the vector representation⁸. Similarly, time-inversion is given by

$$\Lambda_T x = \begin{bmatrix} -t \\ \vec{x} \end{bmatrix}, \quad \text{and} \quad \Lambda_T p = \begin{bmatrix} -p^0 \\ \vec{p} \end{bmatrix}, \quad (9.37)$$

where $\Lambda_T = \text{diag}(-1, 1, 1, 1)$ in the vector representation⁹. Since the subsets that are not elements of the proper orthochronous subgroup can be determined via applying parity and/or time-inversion, the elements of the proper, orthochronous subgroup is of special interest. Since all elements of the Lorentz group is enclosed in $O(3, 1)$, and proper means $\det(\Lambda) = 1$, the proper, orthochronous transformations (indeed all proper transformations) are members of $SO(3, 1)$ - in this case a subgroup of $O(3, 1)$. $SO(3, 1)$ is also called the proper Lorentz group, but most often it is just referred to as the Lorentz group. $SO(3)$ consists of all boost, rotations and combinations of the two. A generic element of the proper, orthochronous subgroup of the Lorentz group can therefore be expressed as

$$\Lambda^\mu_\nu = \Lambda(R')^\mu_\alpha \Lambda(\gamma)^\alpha_\beta \Lambda(R)^\beta_\nu. \quad (9.38)$$

The Λ -matrices form a representation of the Lorentz group.

Example 9.4. —

Prove that J^μ is a 4-vector, where

$$J^\mu = \begin{bmatrix} \rho \\ J_x \\ J_y \\ J_z \end{bmatrix} = \begin{bmatrix} \rho \\ \rho v_x \\ \rho v_y \\ \rho v_z \end{bmatrix} = \begin{bmatrix} \rho \\ \bar{J} \end{bmatrix}. \quad (9.39)$$

An n -dimensional vector is a physical object that can be represented as a $n \times 1$ matrix. Because a vector can be represented as a matrix, it must obey the rules of linear algebra, including scalar multiplication, vector addition, dot product, scalar and vector projections. Further more a vector is a rank 1 tensor, which means that it must transform in a well defined way during a coordinate transformation; it must conserve its length during a coordinate transformation. This last definition will be used to verify that J^μ is in fact a 4-vector. For a 4-vector the length is called the spacetime interval (ds^2). Consider a volume (V_1) of charge moving with a uniform velocity (v_{1x}) as observed from the reference frame S_1 ¹⁰. In this frame the spacetime interval will be given by

$$ds_1^2 = \rho_1^2(v_{1x}^2 - 1), \quad (9.40)$$

where the y and z components of the velocity is zero and the subscript on ds_1 is introduced because current density has not been proven a 4-vector yet (and therefore that ds for different reference frames is the same). From S_1 the volume containing the charge is contracted via

$$V_1 = \gamma^{-1}(v_{1x})V_0, \quad (9.41)$$

where $\gamma = \gamma(v) = (1 - v^2)^{-\frac{1}{2}}$. The volume is contracted in the direction parallel to the direction of motion, and hence the volume is diminished as seen from S_1 . Considering the scenario from the reference frame of S_2 , which moves with a uniform velocity v with respect to S_1 and parallel to the x -axis of S_1 , the volume is seen to be moving with v_{2x} . Here the volume is observed to be

$$V_2 = \gamma^{-1}(v_{2x})V_0. \quad (9.42)$$

⁸Not in general, only in the vector representation.

⁹Not in general, only in the vector representation.

¹⁰I only consider movement along the x -axis to illustrate the principle. A general treatment is rather cumbersome.

Using that

$$v_{2x} = \frac{v_{1x} - v}{1 - vv_{1x}}. \quad (9.43)$$

The volumes in the two reference frames can be related in terms of the velocity in one reference frame. For example

$$\frac{V_1}{V_2} = \gamma(v)(1 - vv_{1x}). \quad (9.44)$$

This relation can be used to determine how the charge density transforms. There must be conservation of charge ($Q = \rho_i V_i$) between reference frames. Therefore

$$V_1 \rho_1 = V_2 \rho_2 \Rightarrow \rho_2 = \rho_1 \frac{V_1}{V_2} = \rho_1 \gamma(v)(1 - vv_{1x}). \quad (9.45)$$

Using this to express ds_2^2

$$ds_2^2 = \rho_2^2 (v_{2x}^2 - 1) = \left(\rho_1 \gamma(v)(1 - vv_{1x}) \right)^2 \left(\left(\frac{v_{1x} - v}{1 - vv_{1x}} \right)^2 - 1 \right) = \rho_1^2 (v_{1x}^2 - 1) = ds_1^2. \quad (9.46)$$

The space time interval is conserved ($ds_1 = ds_2$), and J^μ is therefore a 4-vector. One can also say that ρ , as has been shown for the x -component, transforms via the Lorentz-transformation and therefore the length of the 4-vector transforms from reference frame to reference frame via

$$J'^\mu J'_\mu = \Lambda_x(v)_\nu^\mu \Lambda_x^{-1}(v)^\nu_\mu J^\nu J_\nu = J^\nu J_\nu, \quad (9.47)$$

where L is the Lorentz matrix

$$\Lambda_x(v) = \begin{bmatrix} \gamma & -v\gamma & 0 & 0 \\ -v\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (9.48)$$

As is apparent the length of the vector is conserved as it should be in order for J^μ to be a vector.

Example 9.5.

A spaceship travels with a velocity, $\vec{v}^T = \frac{1}{2} [1 \ 1 \ 1]$, when it is hit by a cosmic ray carrying a four momentum, $p^T = [300 \ 299 \ 0 \ 0] \cdot 10^{-27} \text{kg}$.

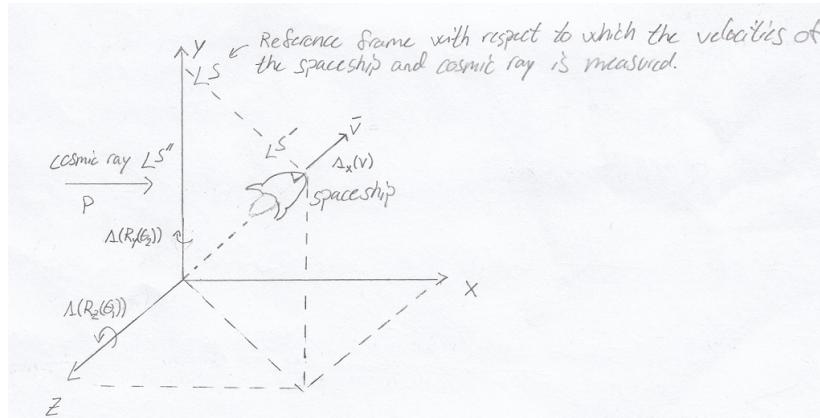


Figure 9.5: The scenario with the three reference frames involved.

1. Find the four velocity, u , of the spaceship.

From S

$$u = \gamma(v) \begin{bmatrix} 1 \\ \vec{v} \end{bmatrix} = \begin{bmatrix} 2 \\ 2\vec{v} \end{bmatrix}, \quad (9.49)$$

where $v = |\vec{v}|$.

2. Find the energy of the cosmic ray in S' .

I know the energy in S from p , so I will create a Lorentz transformation to transform p from S to S' . To this end I will use a general Lorentz transformation which will consist of two rotations, one boost and the inverse rotations. The idea is to rotate the coordinate-system so it aligns with the velocity of the space-ship, boost so the coordinate system, in rotated form, moved with the spaceships, and lastly rotate the coordinate system back. The result will be a boost of S such that it follows the spaceship, i.e. $S \Rightarrow S'$. First, I will rotate (counter clockwise, i.e. positive angle) around the z -axis such that the \vec{v} lies in the zx plane. Hereafter I will rotate (clockwise, i.e. negative angle) around the (new, since I have rotated around the z -axis) y -axis such that the x -axis is aligned with \vec{v} . Hereafter I will boost along the x -axis and rotate back. The Lorentz transformation

$$\Lambda_{tot} = (\Lambda(R_y(\theta_2))\Lambda(R_z(\theta_1)))^{-1}\Lambda_x(\gamma(v))\Lambda(R_y(\theta_2))\Lambda(R_z(\theta_1)), \quad (9.50)$$

where

$$\Lambda(R_i(\theta_j)) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & R_i(\theta_j) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \Lambda_x(\gamma(v)) = \begin{bmatrix} \gamma(v) & -v\gamma(v) & 0 & 0 \\ -v\gamma(v) & \gamma(v) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (9.51)$$

Since I know v , all I need is the two rotation angles. I find these by considering the criterion for each rotation. After the first rotation the velocity should be in the zx -plane, and therefore the transformed velocity should have no y -component. Hence, I can find the first rotation angle from

$$\begin{aligned} R_z(\theta_1)\vec{v} &\propto \begin{bmatrix} \cos(\theta_1) & \sin(\theta_1) & 0 \\ -\sin(\theta_1) & \cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta_1) + \sin(\theta_1) \\ \sin(\theta_1) - \cos(\theta_1) \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \dots \\ 0 \\ 1 \end{bmatrix}. \end{aligned} \quad (9.52)$$

In order for the second component of the transformed velocity to vanish $\sin(\theta_1) = \cos(\theta_1) \Rightarrow \theta_1 = \frac{\pi}{4}$. From the second rotation the x -axis is to align with \vec{v} , so the third component of the transformed velocity is to vanish. Hence, I can find the second rotation angle from

$$\begin{aligned} R_y(\theta_2)R_z(\theta_1)\vec{v} &\propto \begin{bmatrix} \cos(\theta_2) & 0 & -\sin(\theta_2) \\ 0 & 1 & 0 \\ \sin(\theta_2) & 0 & \cos(\theta_2) \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{2}\cos(\theta_2) - \sin(\theta_2) \\ 0 \\ \cos(\theta_2) + \sqrt{2}\sin(\theta_2) \end{bmatrix} \\ &= \begin{bmatrix} |\vec{v}| \\ 0 \\ 0 \end{bmatrix}. \end{aligned} \quad (9.53)$$

In order for the third component of the transformed velocity to vanish; $\cos(\theta_2) + \sqrt{2}\sin(\theta_2) = 0 \Rightarrow \theta_2 = -35.264\text{deg}$. By using the two rotation angles I can now transform p and find it in S'

$$p' = \Lambda_{tot}p = \begin{bmatrix} 301 \\ 296/3 \\ -601/3 \\ -601/3 \end{bmatrix} \cdot 10^{-27}\text{kg}. \quad (9.54)$$

Hence, the energy (now in SI units) as seen from S' is $301 \cdot 10^{-27}\text{kg} \cdot c^2 \simeq 2.7 \cdot 10^{-8}\text{J}$.

3. Calculate the energy in S' from $-g_{\mu\nu}u^\mu p^\nu$. Why does this work?

I can also calculate the energy from $-g_{\mu\nu}u^\mu p^\nu$. $-g_{\mu\nu}u^\mu p^\nu$ has all indices contracted and so is a Lorentz scalar. In the rest frame of the spaceship $u^T = \begin{bmatrix} 1 \\ \vec{0} \end{bmatrix}$ and so it works to pick out the zeroth component of the momentum vector; the energy. So, the quantity is the energy in S' and it is invariant between reference frame. Hence, I can calculate it in S and thereby obtain the energy in S' . I find

$$\begin{aligned} -g_{\mu\nu}u^\mu p^\nu &= u^0 p^0 - u^1 p^1 - u^2 p^2 - u^3 p^3 \\ &= (2 \cdot 300 - 299) \cdot 10^{-27} kg \cdot c^2 \\ &= 301 \cdot 10^{-27} kg \cdot c^2. \end{aligned} \quad (9.55)$$

9.4 The Poincare' group

The Poincare' group, $ISO(1,3)$, is the group of Lorentz transformations and translations. The Poincare' group encompasses the transformations the physical theories, in the absence of general relativistic effects, should obey, i.e. the physical theory should be invariant under translation, rotation and boosts - Physical theories should be invariant under Poincare' transformations. A generic Poincare' transformation is on the form

$$x^\mu \Rightarrow x'^\mu = \Lambda^\mu_\nu x^\nu + a^\mu. \quad (9.56)$$

where a^μ is a constant four vector. From which it is clear that the Poincare' group contains Lorentz transformations and translations. The generators of the Poincare' group are the generators of the Lorentz group, J_i, K_i and the generator of translation in flat space-time (Minkowski space); p^μ . In terms of the generators, the algebra reads

$$\begin{aligned} [J_i, J_j] &= i\epsilon_{ijk}J_k, & [J_i, K_j] &= i\epsilon_{ijk}K_k, & [K_i, K_j] &= -i\epsilon_{ijk}J_k, \\ [J_i, p_j] &= i\epsilon_{ijk}p_k, & [j_i, p_0] &= 0, & [k_i, p_j] &= i\delta_{ij}p_0, & [k_i, p_0] &= -ip_i. \end{aligned} \quad (9.57)$$

10 | Classical field theory

Field theory is one of the cornerstones of classical physics. The most notable examples of classical fields are the force fields that one encounters in the description of gravitational and electromagnetic phenomena. These fields are caused by the presence of masses and electric charges, respectively. In this chapter I present a framework for classical field theory, which is known as the Lagrangian formulation, alongside a vital theorem relating mathematical symmetries of the Lagrangian to conserved quantities; Noethers theorem.

10.1 The Lagrangian formulation of classical field theory

The framework of classical field theory is the generalization of Lagrangian mechanics to fields. The formulation of Lagrangian mechanics made for classical mechanics was for a finite number of degrees of freedom - fields incorporate an infinite number of degrees of freedom and as such the Lagrangian formalism must be generalized to encompass this scenario. In order to do so the classical Lagrangian is expressed, not as a function of dynamical variables that are functions of time (i.e. $q_i(t)$), but as a function of functions (fields) of space-time¹. For a local field theory, a generic Lagrangian (L) can be written in terms of the Lagrangian density (\mathcal{L}) which is a function of the generic fields ($\phi_a(x)$) and their derivatives ($\partial_\mu \phi_a(x)$), i.e.

$$L[\phi_a(x), t] = \int d^3x \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x)). \quad (10.1)$$

The Lagrangian density is related to the Hamiltonian density cf.

$$\mathcal{H} = \sum_a \Pi_a \partial_0 \phi_a - \mathcal{L}, \quad (10.2)$$

which is in complete analogy with the definition of the Hamiltonian in classical mechanics (see equation (3.31)). In both classical and quantum field theory the theories are formulated in terms of a Lagrangian density because the Lagrangian density is manifestly Lorentz invariant. In general the physical laws are required to be covariant under Poincare' transformations (translations, rotations and boosts). Poincare' covariance is a statement of the familiar principle from classical mechanics; physics should be the same in inertial reference frames. The fields in the action are inherently independent of translations². Hence, the action is only "forced" to be covariant under Lorentz transformations. Since the action is a Lorentz scalar, the Lagrangian density is also the be Lorentz invariant.

Lorentz covariance: *A quantity is called Lorentz covariant if the form of the equation does not change under a Lorentz transformation. An object is Lorentz covariant if it transforms under an irreducible representation of the Lie algebra of the proper, orthochronous Lorentz group.*

¹A function of a function is a functional. Hence, in generalizing the Lagrangian formalism from classical mechanics to classical field theory the Lagrangian goes from being a function to being a functional.

²The action is invariant under a constant coordinate transformation since this can be absorbed by the integration measure. Take $S = \int d^4x \mathcal{L}(\phi(x), \dots) \Rightarrow S' = \int d^4x \mathcal{L}(\phi(x+a), \dots) = \int d^4y \mathcal{L}(\phi(y), \dots) = S$, where $y = x + a$. Hence, the action is inherently invariant under coordinate translations.

Lorentz invariance: A term is Lorentz invariant if it does not change during a proper, orthochronous Lorentz transformation. An object that does not change under proper, orthochronous Lorentz transformations is called a Lorentz scalar.

To recap; Lorentz covariance is directly derived from the principle that physics should be the same in inertial reference frames. Enforcing Lorentz covariance means that the fields in the Lagrangian density should transform under irreducible representations of the Lie algebra of the proper, orthochronous Lorentz group. In the next section the representations will be derived. Subsequently these representations can be arranged such that the resulting objects are invariant under parity - Which must physical systems are observed to be³!

10.2 Irreducible (finite) representations of the Lorentz group

Here the irreducible representations of the Lie algebra of the proper, orthochronous Lorentz group (or equivalently; the irreducible representations of the covering group of the proper, orthochronous Lorentz group) will be determined. The proper, orthochronous Lorentz group is not a simply connected group and as such there is not a one-to-one correspondence between the irreducible representations of the Lie algebra and the representations of the proper, orthochronous Lorentz group. Instead, there is a one-to-one correspondence between the irreducible representations of the Lie algebra and the covering group of the proper, orthochronous Lorentz group. Hence, when one talks about the irreducible (finite⁴) representations of the Lorentz group, one really talks about the irreducible representations of the covering group (double cover) of the proper, orthochronous Lorentz group. Recall that a covering group is a group which manifold is simply connected. For this reason, groups which have identical Lie algebra must have a manifold that is contained inside the manifold of the covering group. An example is $SU(2)$ and $SO(3)$ (the two groups have the same Lie algebra); the manifold of $SU(2)$ is the unit three sphere whereas the manifold of $SO(3)$ is half the unit three sphere. Since the unit three sphere is simply connected, $SU(2)$ is the simply connected group of this Lie algebra and the manifolds of groups having identical Lie algebra is contained/covered by the manifold of $SU(2)$. The covering group of the proper, orthochronous Lorentz group is $SL(2, \mathbb{C})$; the group of non-degenerate complex matrices with unit determinant. Hence the irreducible representations obtained from the Lie algebra of the proper, orthochronous Lorentz group are the irreducible representations⁵ of $SL(2, \mathbb{C})$, and it is these representations that the objects in the Lagrangian density is required to transform under in order to be Lorentz covariant.

To determine the irreducible representations of the covering group of the proper, orthochronous Lorentz group I will begin by considering a field (not necessarily hermitian) that carries a generic Lorentz index, $\phi_a(x)$. Under a Lorentz transformation on the form $x^\mu \Rightarrow x'^\mu = \Lambda^\mu_\nu x^\nu$ the generic field transforms viz

$$\phi_a(x^\mu) \Rightarrow \phi'_a(x'^\mu) = \phi'_a(\Lambda^\mu_\nu x^\nu) = D[\Lambda]_a^b \phi_b(x^\mu), \quad (10.3)$$

where⁶ $D[\Lambda]_a^b$ are matrices responsible for the transformation of the field (for a vector field; $D[\Lambda]_a^b = \Lambda^\mu_\nu$). The matrices $D[\Lambda]_a^b$ form a representation of the covering group of the proper, orthochronous Lorentz group. The proper, orthochronous Lorentz group is a Lie group. Recall from the chapter on group theory: A group is a set of elements $\{g_i\}$ and a rule $g_i \otimes g_j \Rightarrow g_k$ which informs how each pair of elements is multiplied to get a third. A Lie group is a group which is also a differentiable manifold. The manifold of the group can

³Specifically the spinors are not invariant under parity. Hence, they are arranged in the Dirac spinor which is.

⁴"Finite" is used because there is also an infinite dimensional representation of the cover of the proper, orthochronous Lorentz group. These are the matrices responsible for Lorentz transforming x . When a field is transformed the field itself is transformed via a finite dimensional representation of the cover of the proper, orthochronous Lorentz group whereas the argument of the field is transformed via the infinite dimensional representation of the cover of the proper, orthochronous Lorentz group. However, how to transform the position four vector in special relativity has already been covered and as such the infinite dimensional representation will not be considered here.

⁵Only some of the irreducible representations of the Lie algebra of the proper, orthochronous Lorentz group will be irreducible representations of the proper orthochronous Lorentz group (eg. the spinor representations are not irreducible representations of the proper, orthochronous Lorentz group).

⁶The last equality can be intuitively justified by the fact that a scalar field is invariant, i.e. $\phi'(x') = \phi(x)$.

be represented by several representations. A representation is a particular embedding of the group elements into matrices. In this case the transformation matrices, $D[\Lambda]_a^b$, form a representation of the covering group of the proper, orthochronous Lorentz group. Since $D[\Lambda]_a^b$ form a representation of the covering group of the proper, orthochronous Lorentz group, they must obey

$$D[\Lambda_1]D[\Lambda_2] = D[\Lambda_1\Lambda_2], \quad (10.4)$$

where $D[\Lambda^{-1}] = D[\Lambda]^{-1}$ and $D[I] = I$. The generic transformation $\phi_a(x^\mu) \Rightarrow \phi'_a(x'^\mu)$ transforms both the fields and coordinates. However, in both classical and quantum field theory *active* coordinate transformations are considered. An active coordinate transformation is a transformation which shifts the field but not the coordinates. That is an active transformation is of the form $\phi_a(x^\mu) \Rightarrow \phi'_a(x^\mu)$. In order to obtain an active transformation from equation (10.3), let $x^\nu \Rightarrow (\Lambda^{-1})^\mu_\nu x^\nu$. Thereby

$$\phi_a(x^\mu) \Rightarrow \phi'_a(x^\mu) = D[\Lambda]_a^b \phi_b(\Lambda^{-1}x). \quad (10.5)$$

Equation (10.5) is the form of transformations (active) in both classical and quantum field theory. The covering group of the proper, orthochronous Lorentz group itself is a mathematical object (manifold) independent of any particular representation. To study the group, independent of the representations, a generic element of the group must be expressed in terms of the basis of generators of the group. The covering group of the proper, orthochronous Lorentz group (being a Lie group) has basis generators defined in terms of the deviation (of an infinitesimal, generic element of $SO(3, 1)$) from the unity element (see equation (9.6)). That is, the unitary operator must be on the (infinitesimal) form

$$\Lambda \sim R(dx) = e - idxJ. \quad (10.6)$$

In the case of the covering group of the proper, orthochronous Lorentz group there are several continuous parameters (analogous to dx), eg. the continuous parameters representing the boosts and rotations. These parameters must be collected in some way. To determine the way to perform this (the collection of the parameters), consider an infinitesimal Lorentz transformation parametrized by

$$\Lambda^\mu_\nu = \delta^\mu_\nu + \omega^\mu_\nu. \quad (10.7)$$

To gain information about ω^μ_ν use the parametrization of Λ^μ_ν in the expression for the metric

$$\begin{aligned} \eta_{\mu\nu} &= \eta_{\alpha\beta}\Lambda^\alpha_\nu\Lambda^\beta_\mu \\ &\simeq \eta_{\alpha\beta}(\delta^\alpha_\nu + \omega^\alpha_\nu)(\delta^\beta_\mu + \omega^\beta_\mu) \\ &= \eta_{\alpha\beta}(\delta^\alpha_\nu\delta^\beta_\mu + \delta^\alpha_\nu\omega^\beta_\mu + \delta^\beta_\mu\omega^\alpha_\nu + \omega^\alpha_\nu\omega^\beta_\mu) \\ &= \eta_{\mu\nu} + \eta_{\nu\beta}\omega^\beta_\mu + \eta_{\alpha\mu}\omega^\alpha_\nu + \omega_{\beta\nu}\omega^\beta_\mu \\ &= \eta_{\mu\nu} + \omega_{\nu\mu} + \omega_{\mu\nu} + \mathcal{O}(\omega^2). \end{aligned} \quad (10.8)$$

For the above to hold to first order in ω , ω must be anti-symmetric, i.e. $\omega_{\nu\mu} = -\omega_{\mu\nu}$. As an antisymmetric tensor $\omega_{\mu\nu}$ has six independent parameters, corresponding to 3 rotational angles and 3 boosts. To each of the six parameters corresponds a generator (analogous to J in equation (9.6)). These generators are listed in the antisymmetric tensor; $J_{\mu\nu}$. Hence, the usual product dxJ takes the form $\omega^{\mu\nu}J_{\mu\nu}$ in this case. Because of the summation over two anti-symmetric tensors will lead to identical terms, a factor of $\frac{1}{2}$ is included, in the expression for a generic element, in order to account for this "over counting". Thereby a generic element of the covering group of the proper, orthochronous Lorentz group can be written in terms of the generators as

$$\Lambda = e - \frac{i}{2}\omega^{\mu\nu}J_{\mu\nu}. \quad (10.9)$$

Compounding these infinitesimal transformations to a finite one

$$\Lambda = e^{-\frac{i}{2}\omega^{\mu\nu}J_{\mu\nu}}. \quad (10.10)$$

The explicit form of the generators, $J^{\mu\nu}$, depends in the representation. However, since the generators are rank 2 tensors, the generators must transform viz

$$J^{\mu\nu} \Rightarrow J'^{\mu\nu} = \Lambda^{-1} J^{\mu\nu} \Lambda = \Lambda^\mu_\alpha \Lambda^\nu_\beta J^{\alpha\beta}. \quad (10.11)$$

Next use equation (10.9) and expand $\Lambda^{-1} J^{\mu\nu} \Lambda$ to first order in Ω

$$\begin{aligned} \Lambda^{-1} J^{\mu\nu} \Lambda &= \left(e - \frac{i}{2} \omega^{\alpha\beta} J_{\alpha\beta} \right) J^{\mu\nu} \left(e - \frac{i}{2} \omega^{\sigma\rho} J_{\sigma\rho} \right) \\ &= J^{\mu\nu} + \frac{i}{2} \omega_{\alpha\beta} [J^{\mu\nu}, J^{\alpha\beta}] + \mathcal{O}(\omega^2). \end{aligned} \quad (10.12)$$

Using equation (10.7) to expand $\Lambda^\mu_\alpha \Lambda^\nu_\beta J^{\alpha\beta}$

$$\begin{aligned} \Lambda^\mu_\alpha \Lambda^\nu_\beta J^{\alpha\beta} &= (\delta^\mu_\nu + \omega^\mu_\alpha)(\delta^\mu_\nu + \omega^\mu_\beta) J^{\alpha\beta} \\ &= J^{\mu\nu} + \omega^\mu_\alpha J^{\alpha\nu} + \omega^\nu_\beta J^{\mu\beta} + \mathcal{O}(\omega^2) \\ &= J^{\mu\nu} + \frac{1}{2} \left(\omega^\mu_\alpha J^{\alpha\nu} + \omega^\mu_\beta J^{\beta\nu} + \omega^\nu_\beta J^{\mu\beta} + \omega^\nu_\alpha J^{\mu\alpha} \right) + \mathcal{O}(\omega^2) \\ &= J^{\mu\nu} + \frac{1}{2} \left(g^{\mu\beta} \omega_{\beta\alpha} J^{\alpha\nu} + g^{\mu\alpha} \omega_{\alpha\beta} J^{\beta\nu} + g^{\nu\alpha} \omega_{\alpha\beta} J^{\mu\beta} + g^{\nu\beta} \omega_{\beta\alpha} J^{\mu\alpha} \right) + \mathcal{O}(\omega^2) \\ &= J^{\mu\nu} + \frac{1}{2} \omega_{\alpha\beta} \left(-g^{\nu\beta} J^{\mu\alpha} - g^{\mu\beta} J^{\alpha\nu} + g^{\nu\alpha} J^{\mu\beta} + g^{\mu\alpha} J^{\beta\nu} \right) + \mathcal{O}(\omega^2). \end{aligned} \quad (10.13)$$

Comparing equation (10.12) and (10.13) it is clear that all generators of the Lorentz group should obey the commutator relation

$$[J^{\mu\nu}, J^{\alpha\beta}] = i \left(g^{\nu\beta} J^{\mu\alpha} + g^{\mu\beta} J^{\alpha\nu} - g^{\nu\alpha} J^{\mu\beta} - g^{\mu\alpha} J^{\beta\nu} \right). \quad (10.14)$$

This is the Lie algebra of $SO(3, 1)$. In order to analyse the representations of the covering group of the proper, orthochronous proper, orthochronous Lorentz group it is convenient to introduce the rotation generators, J^i , and the boost generators, K^j , viz

$$J^i \equiv \frac{1}{2} \epsilon^{ijk} J^{jk}, \quad K^i \equiv J^{i0}. \quad (10.15)$$

In terms of these definitions, the Lie algebra is given by

$$[J^i, J^j] = i \epsilon^{ijk} J^k, \quad [J^i, K^j] = i \epsilon^{ijk} K^k, \quad [K^i, K^j] = -i \epsilon^{ijk} J^k. \quad (10.16)$$

From the above it is clear J^i obey the Lie algebra of $SU(2)$, however the K^i does not. To simplify the Lie algebra further, make another redefinition viz

$$M^i \equiv \frac{1}{2} (J^i + iK^i), \quad N^i \equiv \frac{1}{2} (J^i - iK^i) = (M^i)^\dagger. \quad (10.17)$$

The redefined generators obey the commutator relations given by

$$[M^i, M^j] = i \epsilon^{ijk} M^k, \quad [N^i, N^j] = i \epsilon^{ijk} N^k, \quad [M^i, N^j] = 0. \quad (10.18)$$

These are precisely the commutator relations for the Lie algebra of $SU(2)$. It is therefore clear that the Lie algebra of the proper, orthochronous Lorentz groups consists of two copies of $SU(2)$, i.e. $SU(2) \otimes SU(2)$. This is good news since the irreducible representations the Lie algebra of $SU(2)$ are well determined (Schwichtenberg, 2015). Because the Lorentz group can be described as a $SU(2) \otimes SU(2)$, and the $SU(2)$ Lie algebra is the algebra of angular momentum, a particular representation can be specified by two quantum numbers of angular momentum; j, j' ⁷. In group theoretical terms, the representations are specified by the eigenvalues of Casimir operators. A Casimir operators is an operator built from generators that commute with all

⁷To keep notation it would have to be n, m but j, j' is conventionally taken.

generators of the group. The eigenvalue of the Casimir operator has a constant value for each representation and is hence suited for labelling the representation. The Casimir operator of $SU(2)$ which eigenvalues label the representations is J^2 . The eigenvalues are denoted by j . Hence, since the covering group of the proper, orthochronous Lorentz group is $SU(2) \otimes SU(2)$, this group is labelled by the eigenvalues of two Casimir operators; J_1^2, J_2^2 ; (j_1, j_2) and has $(2j + 1)(2j' + 1)$ degrees of freedom, i.e.

$$(j, j'), \quad \text{where: } j, j' = 0, \frac{1}{2}, 1, \dots \quad (10.19)$$

The regular rotation generators are given by $\vec{J} = \vec{M} + \vec{N}$, and each representation contains particles of spin $J = j + j', j + j' - 1, \dots |j - j'|$. This leads to the possible representations and the associated angular momentum:

Notation	Representation	Angular momentum
Scalar	$(0, 0)$	$J = 0$
Left handed spinor	$(\frac{1}{2}, 0)$	$J = \frac{1}{2}$
Right handed spinor	$(0, \frac{1}{2})$	$J = \frac{1}{2}$
Vector	$(\frac{1}{2}, \frac{1}{2})$	$J = \frac{1}{2}, 0$

where the most used representations (scalar, spinor, vector,...) have been listed. Hence, fields which transform under the above representations are allowed in Lagrangian densities that are to describe physical systems.

Example 10.1.

The lowest order representation, $(0, 0)$, called the scalar representation, is quite trivial because the vector space is one dimensional for both copies of the Lie algebra of $SU(2)$. The generators must therefore be 1×1 matrices. The only 1×1 matrices that satisfy the commutator relations of the generators, i.e. equation (10.18) are trivially 0. Hereby

$$J_{\mu\nu} = 0 \Rightarrow \Lambda = e. \quad (10.20)$$

Therefore the scalar representation acts on objects that do not change under Lorentz transformations. Therefore, the scalar field transforms viz

$$\phi(x) \Rightarrow \phi'(x) = \phi(\Lambda^{-1}x). \quad (10.21)$$

Example 10.2.

The $(\frac{1}{2}, 0)$ representation is called the left handed spinor representation. The left-handed spinor field transforms viz

$$\psi_a(x) \Rightarrow \psi'_a(x) = L[\Lambda]_a^b \psi_b(\Lambda^{-1}x), \quad (10.22)$$

where for an infinitesimal transformation

$$L[\Lambda]_a^b = \delta_a^b - \frac{i}{2} \omega^{\mu\nu} (S_{\mu\nu}^{(L)})_a^b, \quad (10.23)$$

where $(S_{(L)}^{\mu\nu})_a^b = -(S_{(L)}^{\nu\mu})_a^b$ are the representation matrices which must obey the same algebra as the generators of the Lorentz group (i.e. the commutator relation of $J^{\mu\nu}$). Using the same $\omega^{\mu\nu}$ in $L[\Lambda]$ as in $\Lambda(\omega)$ ensures that x, ψ are subject to the same Lorentz transformation.

Example 10.3.

The $(0, \frac{1}{2})$ representation is called the right handed spinor representation. Taking the hermitian conjugate swaps the $SU(2)$ Lie algebras of the Lorentz group. Therefore, the hermitian conjugate of a left-handed spinor field should be a right-handed spinor field. Indices of right handed spinors are denoted, following the notation of Srednicki (2007), with a dot viz⁸

$$[\psi_a(x)]^\dagger = \psi_a^\dagger(x). \quad (10.24)$$

⁸This notation is called the Van der Waerden Notation (Schwichtenberg, 2015).

Taking the hermitian conjugate of the transformation of a left-handed spinor

$$\psi_a^\dagger(x) \Rightarrow \psi_a'^\dagger(x) = R[\Lambda]_{\dot{a}}^{\dot{b}} \psi_b^\dagger(\Lambda^{-1}x), \quad (10.25)$$

where for an infinitesimal transformation

$$R[\Lambda]_{\dot{a}}^{\dot{b}} = \delta_{\dot{a}}^{\dot{b}} - \frac{i}{2} \omega^{\mu\nu} (S_{\mu\nu}^{(R)})_{\dot{a}}^{\dot{b}}. \quad (10.26)$$

Example 10.4.

The $(\frac{1}{2}, \frac{1}{2})$ representation is called the vector representation. This representation is the one familiar from special relativity⁹. A vector field transforms viz

$$A^\mu(x) \Rightarrow A'^\mu(x) = \Lambda^\mu_\nu A^\nu(\Lambda^{-1}x). \quad (10.27)$$

10.2.1 Construction a Lagrangian density

At this point it has been established that in field theory the primary content of the Lagrangian density is fields. The allowed fields in the Lagrangian must transform under an irreducible representation of the cover group of the proper, orthochronous Lorentz group. whereas the scalar and vector fields are invariant under parity, the spinor fields are not. The easiest thing to do to build a parity invariant quantity from the spinors is to write them into a Dirac field. The Dirac field is defined viz

$$\Psi_D(x) = \psi_L(x) \oplus \psi_R(x) \doteq \begin{bmatrix} \psi_L(x) \\ \psi_R(x) \end{bmatrix}, \quad (10.28)$$

where $\psi_{L,R}(x)$ are left handed and right handed spinor fields, respectively. Usually only Dirac fields are considered, and as such the subscript D is omitted. The left handed spinor transforms via $L[\Lambda]$ whereas the right handed spinor transforms via $R[\Lambda]$. Hence, the transformation of the Dirac field can be written viz

$$\Lambda_D = \begin{bmatrix} L[\Lambda] & 0 \\ 0 & R[\Lambda] \end{bmatrix}. \quad (10.29)$$

That is

$$\Psi(x) \Rightarrow \Psi'(x') = \Lambda_D \Psi(\Lambda^{-1}x). \quad (10.30)$$

By introducing the Dirac field, the fields available to build Lagrangians form are¹⁰

$$\text{Available fields: } \{\text{scalar } (\phi(x)), \text{ Dirac } (\Psi(x)), \text{ vector } (A^\mu(x)) \dots\}. \quad (10.31)$$

From these fields, Lorentz scalars can be constructed. In order for the Lagrangian density to be Lorentz invariant, it must be a Lorentz scalar. Hence, the terms in the Lagrangian must be Lorentz scalars, i.e.

$$\mathcal{L} = \sum_i (\text{Lorentz scalars})_i. \quad (10.32)$$

Lorentz scalars terms that transform like a scalar under a Lorentz transformation.

Example 10.5.

An example of a Lorentz scalar is $\bar{\psi}(x)\psi(x)$, where both fields are Dirac spinors. $\bar{\psi}(x)\psi(x)$ transforms viz

$$\begin{aligned} \bar{\psi}(x)\psi(x) &\Rightarrow \psi^\dagger(\Lambda^{-1}x)\Lambda_D^\dagger \gamma^0 \Lambda_D \psi(\Lambda^{-1}x) = \psi^\dagger(\Lambda^{-1}x)\gamma^0 \psi(\Lambda^{-1}x) \\ &= \bar{\psi}(\Lambda^{-1}x)\psi(\Lambda^{-1}x). \end{aligned} \quad (10.33)$$

⁹The generators can be found in Peskin and Schroeder (1995)[p. 39].

¹⁰There are higher representations available but these are the most fundamental and suffice to describe all known particles.

In general, the terms of Lagrangian densities are arranged into two categories; kinetic terms and potential terms. In general, the kinetic terms contain the derivative of the field. The potential terms are more complicated to make broad statements about. For a spin half field, which most fields are, all potential terms (they must be Lorentz scalars) can be constructed from what is called bilinears. In relation to the standard model in quantum field theory all known particles of spin half are described by Dirac fields. The Dirac bilinears are on the form

$$\text{Form of Dirac field bilinears: } \bar{\psi}(x)\Gamma\psi(x), \quad (10.34)$$

where Γ is a 4×4 constant matrix. The possible Γ 's are categorized into groups according to how they transform under the Lorentz group

Scalar:	$\Gamma = I,$	
Vector:	$\Gamma = \gamma^\mu,$	
Tensor:	$\Gamma = \sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu],$	(10.35)
Pseudo-vector:	$\Gamma = \gamma^\mu\gamma^5,$	
Pseudo-scalar:	$\Gamma = i\gamma^5,$	

where the i 's are added to make the terms Hermitian. The terms involving "pseudo" transforms as vectors and scalars, respectively, under the continuous Lorentz group (proper, orthochronous transformations), but acquire a sign change under parity transformation. The non-scalar terms contain uncontracted Lorentz indices and as such they must be contracted with terms containing corresponding Lorentz indices in order to transform as a scalar under the Lorentz group.

Example 10.6.

One common Lagrangian density is the Klein-Gordon (KG) Lagrangian density. The KG Lagrangian density comes in two versions; for a complex scalar field and for a real scalar field. Here the case of a complex scalar field will be considered. The KG equation consists of the simplest kinetic and potential terms

$$\text{KG: } \mathcal{L} = (\partial_\mu\phi)^*\partial^\mu\phi - m^2\phi^*\phi. \quad (10.36)$$

Example 10.7.

Consider the following Lagrangian density (Klein-Gordon Lagrangian density)

$$\mathcal{L} = (D_\mu\phi)^*D^\mu\phi - m^2\phi^*\phi, \quad (10.37)$$

where $D_\mu = \partial_\mu - igA_\mu$. Show that the KG Lagrangian density is invariant under the Gauge transformation given by

$$\phi \Rightarrow \phi e^{i\alpha(x)} \quad \text{and} \quad A_\mu \Rightarrow A_\mu + \partial_\mu\Lambda(x). \quad (10.38)$$

Find also a relation between $\alpha(x)$ and $\Lambda(x)$.

Begin by inserting the Gauge transformation and write the expression out

$$\begin{aligned} \mathcal{L} = & \underbrace{\left(\partial_\mu + ig(A_\mu + \partial_\mu\Lambda(x))\right)\phi^*e^{-i\alpha(x)}\left(\partial^\mu - ig(A^\mu + \partial^\mu\Lambda(x))\right)\phi e^{i\alpha(x)}}_{\mathcal{L}_1} \\ & - \underbrace{m^2\phi^*e^{-i\alpha(x)}\phi e^{i\alpha(x)}}_{\mathcal{L}_2}. \end{aligned} \quad (10.39)$$

In the last term (\mathcal{L}_2) the exponentials cancel each other out and the term reduces to the one before the transformation. Therefore it comes down to the first term

$$\begin{aligned}\mathcal{L}_1 &= \left(\partial_\mu + ig(A_\mu + \partial_\mu \Lambda(x)) \right) \phi^* e^{-i\alpha(x)} \left(\partial^\mu - ig(A^\mu + \partial^\mu \Lambda(x)) \right) \phi e^{i\alpha(x)} \\ &= \left(e^{-i\alpha(x)} \partial_\mu(\phi^*) - i\phi^* e^{-i\alpha(x)} \partial_\mu(\alpha(x)) + igA_\mu \phi^* e^{-i\alpha(x)} + ig\partial_\mu(\Lambda(x)) \phi^* e^{-i\alpha(x)} \right) \\ &\quad \cdot \left(e^{i\alpha(x)} \partial^\mu(\phi) + i\phi e^{i\alpha(x)} \partial^\mu(\alpha(x)) - igA^\mu \phi e^{i\alpha(x)} - ig\partial^\mu(\Lambda(x)) \phi e^{i\alpha(x)} \right).\end{aligned}\quad (10.40)$$

Pull out the exponentials from each term and let them cancel

$$\begin{aligned}\mathcal{L}_1 &= \left(\partial_\mu \phi^* - i\phi^* \partial_\mu \alpha(x) + igA_\mu \phi^* + ig\partial_\mu \Lambda(x) \phi^* \right) \left(\partial^\mu \phi + i\phi \partial^\mu \alpha(x) - igA^\mu \phi - ig\partial^\mu \Lambda(x) \phi \right) \\ &= \left(D_\mu \phi^* - i\phi^* \partial_\mu \alpha(x) + ig\partial_\mu \Lambda(x) \phi^* \right) \left(D^\mu \phi + i\phi \partial^\mu \alpha(x) - ig\partial^\mu \Lambda(x) \phi \right).\end{aligned}\quad (10.41)$$

If the last two terms in each parenthesis cancel the Lagrangian reduces to the Lagrangian before transformation and hence the Lagrangian is invariant to the gauge transformation. That is gauge invariance requires

$$-i\phi^* \partial_\mu \alpha(x) + ig\partial_\mu \Lambda(x) \phi^* = 0, \quad (10.42)$$

$$i\phi \partial^\mu \alpha(x) - ig\partial^\mu \Lambda(x) \phi = 0. \quad (10.43)$$

Both equations are satisfied if $\alpha(x) = g\Lambda(x)$. Which when inserted in equation 10.39 reduces it to

Example 10.8.

The perhaps most used Lagrangian density in both classical and quantum field theory is the Dirac Lagrangian density. The Dirac Lagrangian density is a function of the Dirac field (made up of spinor fields). It contains a slightly more complicated kinetic term alongside the simplest potential term

$$\text{Dirac: } \mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \quad (10.44)$$

Example 10.9.

Determine the Hamiltonian density given the Lagrangian density

$$\mathcal{L} = -\frac{1}{4\mu_0} F^{\mu\nu} F_{\mu\nu} - A_\mu J^\mu. \quad (10.45)$$

Use that the Hamiltonian is found by performing a Legendre transformation of the Lagrangian. From equation (10.2)

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\alpha)} \partial_0 A_\alpha - \mathcal{L}. \quad (10.46)$$

In order to evaluate the above the Lagrangian density must be expressed entirely in terms of A_μ (with the

index down). by doing so and subsequently taking the derivative

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial(\partial_\beta A_\alpha)} &= \frac{\partial}{\partial(\partial_\beta A_\alpha)} \left(-\frac{1}{4\mu_0} F^{\mu\nu} F_{\mu\nu} - A_\mu J^\mu \right) = -\frac{1}{4\mu_0} \frac{\partial}{\partial(\partial_\beta A_\alpha)} \left(F^{\mu\nu} F_{\mu\nu} \right) \\
&= -\frac{1}{4\mu_0} \frac{\partial}{\partial(\partial_\beta A_\alpha)} \left((\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) \right) \\
&= -\frac{1}{4\mu_0} \frac{\partial}{\partial(\partial_\beta A_\alpha)} \left(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\nu A^\mu \partial_\mu A_\nu + \partial^\nu A^\mu \partial_\nu A_\mu \right) \\
&= -\frac{1}{2\mu_0} \frac{\partial}{\partial(\partial_\beta A_\alpha)} \left(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu \right) \\
&= -\frac{g^{\mu\lambda} g^{\nu\sigma}}{2\mu_0} \frac{\partial}{\partial(\partial_\beta A_\alpha)} \left(\partial_\lambda A_\sigma \partial_\mu A_\nu - \partial_\lambda A_\sigma \partial_\nu A_\mu \right) \\
&= -\frac{g^{\mu\lambda} g^{\nu\sigma}}{2\mu_0} \left(\delta_\lambda^\beta \delta_\sigma^\alpha \partial_\mu A_\nu + \delta_\mu^\beta \delta_\nu^\alpha \partial_\lambda A_\sigma - \delta_\lambda^\beta \delta_\sigma^\alpha \partial_\nu A_\mu - \delta_\nu^\beta \delta_\mu^\alpha \partial_\lambda A_\sigma \right) \\
&= -\frac{1}{2\mu_0} \left(g^{\mu\beta} g^{\nu\alpha} \partial_\mu A_\nu + g^{\beta\lambda} g^{\alpha\sigma} \partial_\lambda A_\sigma - g^{\mu\beta} g^{\nu\alpha} \partial_\nu A_\mu - g^{\beta\lambda} g^{\alpha\sigma} \partial_\lambda A_\sigma \right) \\
&= -\frac{1}{2\mu_0} \left(g^{\mu\beta} g^{\nu\alpha} (\partial_\mu A_\nu - \partial_\nu A_\mu) + g^{\beta\lambda} g^{\alpha\sigma} (\partial_\lambda A_\sigma - \partial_\lambda A_\sigma) \right) \\
&= -\frac{1}{2\mu_0} \left(g^{\mu\beta} g^{\nu\alpha} F_{\mu\nu} + g^{\beta\lambda} g^{\alpha\sigma} F_{\lambda\sigma} \right) = -\frac{F^{\beta\alpha}}{\mu_0}.
\end{aligned} \tag{10.47}$$

Using this in the expression for the Hamiltonian density

$$\mathcal{H} = \frac{1}{4\mu_0} F^{\mu\nu} F_{\mu\nu} + A_\mu J^\mu - \frac{F^{0\alpha}}{\mu_0} \partial_0 A_\alpha. \tag{10.48}$$

10.2.2 The Euler-Lagrange equation in field theory

To obtain the equations of motion from the Lagrangian density in field theory a generalized version of the Euler-Lagrange equation from classical mechanics can be used. The generalized Euler-Lagrange equation is found by, just as in classical mechanics, varying the action. In field theory the action is given as follows (see equation (3.1) to compare with classical mechanics)

$$S[\phi_a(x)] = \int d^4x \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x)). \tag{10.49}$$

The action is a functional of the fields. Consider now the change in the action when the fields are changed (leaving the coordinates alone), i.e. consider the change in the action under

$$\delta_0 \phi_a(x) = \phi'_a(x) - \phi_a(x), \tag{10.50}$$

where the subscript 0 indicates that the space-time coordinate is fixed during the variation. The variation in the action

$$\begin{aligned}
\delta_0 S[\phi_a(x)] &= \int d^4x \delta_0 \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x)) \\
&= \int d^4x \left(\frac{\partial \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x))}{\partial \phi_a(x)} \delta_0 \phi_a(x) + \frac{\partial \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x))}{\partial(\partial_\mu \phi_a(x))} \delta_0(\partial_\mu \phi_a(x)) \right) \\
&= \int d^4x \left(\frac{\partial \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x))}{\partial \phi_a(x)} - \partial_\mu \frac{\partial \mathcal{L}(\phi_a(x), \partial_\mu \phi_a(x))}{\partial(\partial_\mu \phi_a(x))} \right) \delta_0 \phi_a(x) \\
&= 0,
\end{aligned} \tag{10.51}$$

where it has been used that δ_0 and ∂_μ commute, partial integration and that the boundary term vanishes¹¹. For equation (10.51) to be valid for every $\delta_0\phi_a(x)$

$$\frac{\partial \mathcal{L}}{\partial \phi_a} - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_a)} = 0. \quad (10.52)$$

This is the Euler-Lagrange equation generalized to field theory. From equation (10.52), and the specified Lagrangian density, the EOM for the fields can be determined.

Example 10.10.

Consider the Dirac equation

$$(i\gamma^\mu \partial_\mu - m)\psi = 0. \quad (10.53)$$

By defining $\bar{\psi} = \psi^\dagger \gamma^{(0)}$ the Dirac equation can be derived from the Lagrangian density

$$L = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \quad (10.54)$$

1. *Treating $\bar{\psi}$ and ψ as independent fields, show that the Lagrangian results in the Dirac equation.*

The equation of motion, which the Dirac equation is, can be found by using the Lagrangian density in the Euler-Lagrange equation. Since there are two field there are two equations of motion. These equations of motion are equivalent in this case. Consider the equation of motion obtained from $\bar{\psi}$

$$\frac{\partial L}{\partial \bar{\psi}} - \partial_\mu \frac{\partial L}{\partial (\partial_\mu \bar{\psi})} = 0 \Rightarrow (i\gamma^\mu \partial_\mu - m)\psi = 0. \quad (10.55)$$

Since the Lagrangian does not depend on the derivative of the $\bar{\psi}$ -field.

2. *Find the conjugate momentum densities π_ψ and $\pi_{\bar{\psi}}$ and use these to construct the Hamiltonian density.*

The conjugate momenta is given by

$$\frac{\partial L}{\partial (\partial_0 \phi_i)} = \pi_i. \quad (10.56)$$

If the field is a vector, so is the conjugate momentum. Immediately it is clear that the momentum conjugate to $\bar{\psi}$ is zero. For the ψ -field

$$\pi = \frac{\partial L}{\partial (\partial_0 \psi)} = \bar{\psi} i\gamma^0 \partial_0 \psi = i\psi^\dagger \gamma^0 \gamma^0 = i\psi^\dagger I_{4 \times 4} = i\psi^\dagger. \quad (10.57)$$

The Hamiltonian density operator

$$H = \partial_t \psi \cdot \pi - L = i\psi^\dagger \partial_t \psi - \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \quad (10.58)$$

Example 10.11.

Derive Diracs equation for a classical field in the way Dirac did it.

To derive the Dirac equation the basic idea used by Dirac was to factorize the relativistic expression for the energy ($E^2 = \vec{p}^2 + m^2$) into the form

$$\begin{aligned} H^2 &= (\vec{\alpha} \cdot \vec{p} + \beta m)^2 \\ &= (\sum_i \alpha^i p_i + \beta m)(\sum_j \alpha^j p_j + \beta m) \\ &= \sum_{i,j} \alpha^i p_i \alpha^j p_j + \sum_i \alpha^i p_i \beta m + \sum_j \beta m \alpha^j p_j + \beta^2 m^2, \end{aligned} \quad (10.59)$$

¹¹The boundary terms vanish because the endpoints of the path is fixed - just like in the case of classical mechanics.

where $\vec{\alpha}$ is a three component vector in which each component is a 4×4 matrix and β is a 4×4 matrix. Because H must have real eigenvalues (it is hermitian), and p_i is hermitian, it follows that α^i and β must be hermitian matrices¹². Note that $\vec{\alpha}$ and β cannot simply be number if the equation is to have rotational invariance. By assuming that $[\alpha^i, p_j] = 0$, $[p_i, \beta] = 0$ and utilizing the Einstein summation notation

$$\begin{aligned} E^2 &= \alpha^i \alpha^j p_i p_j + (\alpha^i \beta + \beta \alpha^i) p_i m + \beta^2 m^2 \\ &= \frac{1}{2} (\alpha^i \alpha^j + \alpha^j \alpha^i) p_i p_j + (\alpha^i \beta + \beta \alpha^i) p_i m + \beta^2 m^2. \end{aligned} \quad (10.60)$$

This should correspond to $E^2 = \vec{p}^2 + m^2$. Therefore the prefactor in the term with $p_i p_j$ should equal $2\delta_{ij}$. Also the prefactor of the $p_i m$ -term should vanish and $\beta^2 = I_{4 \times 4}$. Explicitly written

$$\frac{1}{2} \sum_{i,j} (\alpha^i \alpha^j + \alpha^j \alpha^i) = \delta^{ij} \Rightarrow \alpha^i \alpha^j + \alpha^j \alpha^i = \{\alpha^i, \alpha^j\} = 2I_{4 \times 4}, \quad (10.61)$$

$$\alpha^i \beta + \beta \alpha^i = 0 \Rightarrow \{\alpha^i, \beta\} = 0, \quad (10.62)$$

$$\beta^2 m^2 = I_{4 \times 4} m^2 \Rightarrow \beta^2 = I_{4 \times 4}. \quad (10.63)$$

The specific form of the α -and β -matrices can be found by taking the β -matrix to be diagonal. It must also be traceless, so therefore the diagonal is of an equal number of "ones" with different sign. That is

$$\beta = \begin{bmatrix} I_{2 \times 2} & 0 \\ 0 & -I_{2 \times 2} \end{bmatrix}. \quad (10.64)$$

From this β -matrix the α -matrices can be found by expressing it (the α -matrix) as a general, hermitian matrix. Hereafter the general hermitian matrix can be made to obey equation 10.62 in order to limit the components of the general matrix. It will be found that the remaining components, the off diagonal-ones in block form, must obey an anti-commutator relation to the Pauli matrices ($\{\sigma^i, \sigma^j\} = 2I_{2 \times 2}$). Therefore the α -matrices can be expressed in terms of the Pauli matrices.

$$\alpha^i = \begin{bmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{bmatrix}. \quad (10.65)$$

This set of α - and β -matrices satisfy the factorization condition given in equation 10.59. This Hamiltonian is used in Schroedingers equation. However, in order to fulfill the dimensional requirements the wave-function must now be a 4×1 matrix. By insertion into the SE

$$i\partial_t \psi = -i\alpha^i \nabla_i \psi - m\beta \psi. \quad (10.66)$$

By acting with β from the left

$$i\partial_t(\beta \psi) = -i\beta \alpha^i \nabla_i \psi - m\beta^2 \psi. \quad (10.67)$$

Next up use that $\beta^2 = I_{4 \times 4}$ and define now the γ -matrices

$$\gamma^i = \beta \alpha^i = \begin{bmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{bmatrix} \quad \text{and} \quad \gamma^0 = \beta = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}. \quad (10.68)$$

From these definitions the Schrödinger equation with the modified Hamiltonian can be written as¹³

$$(i\gamma^\mu \partial_\mu - m)\psi = 0. \quad (10.69)$$

¹²A hermitian matrix satisfies $A = (A^T)^*$.

¹³To be explicit; without Einstein notation and in SI units: $\sum_{\mu=0}^3 (i\hbar \gamma^\mu \partial_\mu - mc)\psi = 0$.

10.3 Noethers theorem

The time has come for one of the most important theorems in modern physics; Noethers theorem. Noethers theorem can be stated as follows

Noethers theorem: *If the Lagrangian density for a physical system is not affected by a continuous change (transformation) in the coordinate system used to describe it, then there will be a corresponding conservation law.*

Noethers theorem has a special place in most theoretical physicists hearts. The reason for this is that it connects a mathematical symmetry of a Lagrangian density with a physical property (symmetry) of a physical system. For example, most physical systems are invariant with respect to a space-time translation. The Lagrangian density for such systems must have the same symmetry (the Lagrangian density is to describe the physical system after all, so it better have the same symmetries). Corresponding to the symmetry must be a conserved quantity. In the case of a Lagrangian density with a space-time symmetry (invariant with respect to translations in space and time), there are two conserved quantities; time \Rightarrow energy and space \Rightarrow momentum. Likewise other conserved quantities can be connected to the physical layout of the system. The theorem therefore invites an interpretation of conservation laws as a result of the physical layout of a system. This is a profound result and allows physicists to search for conserved quantities by searching for symmetries of the Lagrangian density. Conversely, if one is to construct a Lagrangian density for a system with a given set of symmetries, it is now clear how these are implemented (via making the Lagrangian density invariant under transformations).

Corresponding to the conserved quantity is a current with vanishing divergence. This can be determined from the symmetry by investigating the variation in the action carefully. It goes as follows; consider a Lagrangian density that is invariant under the generic transformations

$$x^\mu \Rightarrow x'^\mu = x^\mu + \delta x^\mu, \quad \phi_a(x) \Rightarrow \phi'_a(x') = \phi_a(x) + \delta \phi_a(x). \quad (10.70)$$

Note that, in contrast to the derivation of the Euler-Lagrange equation for field theory, now the coordinates are varied as well as the fields. Take the transformations to be parameterized by some infinitesimal parameter, ε^ν , such that

$$x'^\mu|_{\varepsilon^\nu=0} = x^\mu, \quad \phi'_i(x')|_{\varepsilon^\nu=0} = \phi_i(x). \quad (10.71)$$

The change in the action is given by

$$\delta S[\phi_a(x)] = \int_{\Omega'} d^4x' \mathcal{L}(\partial_\mu \phi'_a(x'), \phi'_a(x')) - \int_{\Omega} d^4x \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) = 0, \quad (10.72)$$

where Ω, Ω' denotes the integration domains. In order to collect the integrals, use that (Goldstein et al., 2002)[p. 593]

$$\begin{aligned} \delta S[\phi_a(x)] &= \int_{\Omega} d^4x [\mathcal{L}(\partial_\mu \phi'_a(x), \phi'_a(x)) - \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))] + \int_S dS_\mu \delta x^\mu \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) \\ &= \int_{\Omega} d^4x \delta_0 \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) + \int_S dS_\mu \delta x^\mu \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) \end{aligned}, \quad (10.73)$$

where it has been used that x' is just a dummy variable in the first integral, so it can be relabeled to x . However, the domain in the first integral is still Ω' rather than Ω . To change this, it has been used that $\int_{\Omega'} d^4x' \mathcal{L}(\phi'_a(x')) = \int_{\Omega} d^4x \mathcal{L}(\phi'_a(x)) + \int_S dS_\mu \delta x^\mu \mathcal{L}(\phi_a(x))$, where S is the three dimensional surface of the region Ω . Next use the divergence theorem ($\int_S dS_\mu \delta x^\mu \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) = \int_{\Omega} \partial_\mu [\mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) \delta x^\mu]$) to find

$$\delta S[\phi_a(x)] = \int_{\Omega} d^4x \left[\delta_0 \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) + \partial_\mu [\mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) \delta x^\mu] \right]. \quad (10.74)$$

The integrand in the first integral in equation (10.73) is to first order the variation in the Lagrangian density due to a change in the field

$$\begin{aligned}\delta_0 \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) &= \frac{\partial \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))}{\partial \phi_a(x)} \delta_0 \phi_a(x) + \frac{\partial \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))}{\partial (\partial_\mu \phi_a(x))} \delta_0 (\partial_\mu \phi_a(x)) \\ &= \frac{\partial \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))}{\partial \phi_a(x)} \delta_0 \phi_a(x) + \left[\partial_\mu \left(\frac{\partial \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))}{\partial (\partial_\mu \phi_a(x))} \delta_0 \phi_a(x) \right) \right. \\ &\quad \left. - \partial_\mu \left(\frac{\partial \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))}{\partial (\partial_\mu \phi_a(x))} \right) \delta_0 \phi_a(x) \right] \\ &= \partial_\mu \left(\frac{\partial \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))}{\partial (\partial_\mu \phi_a(x))} \delta_0 \phi_a(x) \right)\end{aligned}\quad (10.75)$$

where the Euler-Lagrange equation was used to reach the last equality and the notation has been eased a bit. By using equation (10.75) and (10.74) together

$$\delta S[\phi_a(x)] = \int_{\Omega} d^4x \partial_\mu \left[\frac{\partial \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x))}{\partial (\partial_\mu \phi_a(x))} \delta_0 \phi_a(x) + \mathcal{L}(\partial_\mu \phi_a(x), \phi_a(x)) \delta x^\mu \right]. \quad (10.76)$$

The next step is to introduce the infinitesimal parameter. This is done by using that, from $\delta \phi_a(x) = \phi'_a(x') - \phi_a(x)$ and $\delta_0 \phi_a(x) = \phi'_a(x) - \phi_a(x)$;

$$\phi_a(x) = \phi'_a(x') - \delta \phi_a(x) = \phi'_a(x) - \delta_0 \phi_a(x) \Rightarrow \delta_0 \phi_a(x) = -[\phi'_a(x') - \phi'_a(x)] + \delta \phi_a(x). \quad (10.77)$$

Use now in equation (10.77) that $\phi'_a(x') - \phi'_a(x) \simeq \delta x^\mu \partial_\mu \phi'_a(x) \simeq \delta x^\mu \partial_\mu \phi_a(x)$. where $\phi'_a(x) = \phi_a(x) + \delta_0 \phi_a(x)$ has been used and $\delta x^\mu \partial_\mu \delta_0 \phi_a$ has been neglected based on it being of order "small squared". By using this

$$\delta_0 \phi_a(x) = \delta \phi_a(x) - \delta x^\mu \partial_\mu \phi_a(x). \quad (10.78)$$

Lastly parameterize the change in the coordinates and fields via the infinitesimal parameter ε^ν viz¹⁴

$$\delta x^\mu = \frac{\partial x'^\mu}{\partial \varepsilon^\nu} \varepsilon^\nu, \quad \delta \phi_a(x) = \frac{\partial \phi_a(x)}{\partial \varepsilon^\nu} \varepsilon^\nu, \quad (10.79)$$

where $\frac{\partial x^\mu}{\partial \varepsilon^\nu}$ denotes a generic function that describes the deviation from the infinitesimal quantity, ε^ν . By using equation (10.79) in (10.78)

$$\delta_0 \phi_a(x) = \varepsilon^\nu \left[\frac{\partial \phi_a(x)}{\partial \varepsilon^\nu} - \frac{\partial x^\mu}{\partial \varepsilon^\nu} \frac{\partial \phi_a(x)}{\partial x^\mu} \right]. \quad (10.80)$$

Using equation (10.80) and (10.79) in equation (10.76) while cleaning up the notation

$$\begin{aligned}\delta S &= \int_{\Omega} d^4x \varepsilon^\nu \partial_\mu \left[\mathcal{L} \frac{\partial x^\mu}{\partial \varepsilon^\nu} + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_a)} \left(\frac{\partial \phi_a}{\partial \varepsilon^\nu} - \frac{\partial x^\mu}{\partial \varepsilon^\nu} \frac{\partial \phi_a}{\partial x^\mu} \right) \right] \\ &\equiv \int_{\Omega} d^4x \varepsilon^\nu \partial_\mu J_\nu^\mu = 0.\end{aligned}\quad (10.81)$$

Since $\varepsilon^\nu \neq 0$ in general, and $\delta S = 0$, $\partial_\mu J_\nu^\mu = 0$ must be the case. J^μ is the conserved current corresponding to the symmetry of the Lagrangian. To be explicit

$$J_\nu^\mu \equiv \mathcal{L} \frac{\partial x^\mu}{\partial \varepsilon^\nu} + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_a)} \left(\frac{\partial \phi_a}{\partial \varepsilon^\nu} - \frac{\partial x^\mu}{\partial \varepsilon^\nu} \frac{\partial \phi_a}{\partial x^\mu} \right). \quad (10.82)$$

¹⁴It is x' in the expression for δx^μ because: $x'^\mu(\varepsilon^\nu) = x'^\mu(\varepsilon^\nu)|_{\varepsilon^\nu=0} + \frac{\partial x'^\mu(\varepsilon^\nu)}{\partial \varepsilon^\nu}|_{\varepsilon^\nu=0} + \mathcal{O}(\varepsilon^2) = x^\mu + \frac{\partial x'^\mu(\varepsilon^\nu)}{\partial \varepsilon^\nu} + \mathcal{O}(\varepsilon^2)$.

Example 10.12.

One of the most prominent results of Noethers theorem is that of the energy-momentum tensor. The energy momentum-tensor is the conserved current related to a space-time symmetry og the Lagrangian density. To show this, take the Lagrangian density to be invariant under the coordinate transformation

$$x^\nu \Rightarrow x'^\nu = x^\nu + \varepsilon^\nu, \quad (10.83)$$

where ε^ν is a constant displacement in space-time. Since the field does not change under this coordinate transformation, the relevant quantities for equation (10.81) is given by

$$\frac{\partial \phi_i(x)}{\partial \varepsilon} = 0, \quad \frac{\partial x^\mu}{\partial \varepsilon^\nu} = \delta^\mu_\nu. \quad (10.84)$$

Hereby

$$J^\mu_\nu \equiv T^{\text{asym}, \mu}_\nu = \mathcal{L} \delta^\mu_\nu - \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_i)} \partial_\nu \phi_i, \quad (10.85)$$

where T_{asym} is a symbol dedicated to the asymmetric energy-momentum tensor. One can also define a symmetric energy-momentum tensor - this is the one used in general relativity and the Einstein equation! The energy-momentum tensor can be used as an alternative means, as opposed to the Euler-Lagrange equation, to obtain the EOM's.

Example 10.13.

Consider different Lagrangians and find the conserved currents by using Noethers theorem. In this example only transformation in the fields, not coordinate transformations, are considered. Therefore, the conserved current is given by

$$J^\alpha = \sum_i \frac{\partial \mathcal{L}}{\partial (\partial_\alpha \phi_i)} \delta \phi_i, \quad (10.86)$$

where i runs over the different fields in the Lagrangian and $\delta \phi_i$ is found by considering the symmetry transformation as an infinitesimal transformation. Many different symmetries can exist for a Lagrangian, eg. symmetry under Lorentz boost, symmetry under rotation of the phase of the field, symmetry under spacetime translation and many more. Here I will focus on the easiest one; symmetry under a rotation of the phase of the field.

1. Find the conserved current for the lagrangian (Modified KG-Lagrangian)

$$\mathcal{L} = -(\partial^\mu \varphi)^* \partial_\mu \varphi - m^2 \varphi^* \varphi - \frac{\lambda}{4} (\varphi^* \varphi)^2. \quad (10.87)$$

This Lagrangian is invariant under the global symmetry transformation given by

$$\varphi \Rightarrow \varphi e^{i\lambda} \quad \text{and} \quad \varphi^* \Rightarrow \varphi^* e^{-i\lambda}. \quad (10.88)$$

To find the conserved current use that the infinitesimal symmetry transformation is given by

$$\varphi \Rightarrow \varphi(1 + i\lambda) = \varphi + \delta\varphi. \quad (10.89)$$

For the hermitian conjugate field

$$\varphi^* \Rightarrow \varphi^*(1 - i\lambda) = \varphi^* + \delta\varphi^*. \quad (10.90)$$

Using this in Noethers theorem

$$\begin{aligned} J^\alpha &= \frac{\partial \mathcal{L}}{\partial (\partial_\alpha \phi)} \delta\phi + \frac{\partial \mathcal{L}}{\partial (\partial_\alpha \phi^*)} \delta\phi^* \\ &= -g^{\mu\nu} \partial_\nu \varphi^* \delta_\mu^\alpha (i\lambda \varphi) + (g^{\mu\nu}) \partial_\nu \varphi \delta_\mu^\alpha (-i\lambda \varphi^*) \\ &= i\lambda \left(\varphi^* \partial^\alpha \varphi - \varphi \partial^\alpha \varphi^* \right). \end{aligned} \quad (10.91)$$

2. Find the conserved current for the Lagrangian (Dirac Lagrangian)

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \quad (10.92)$$

This Lagrangian is invariant under the global symmetry transformation given by

$$\psi \Rightarrow \psi e^{i\lambda} \quad \text{and} \quad \bar{\psi} \Rightarrow \bar{\psi} e^{-i\lambda}. \quad (10.93)$$

Since $\bar{\psi} = \psi^* \gamma^{(0)}$. The same infinitesimal transformations apply, so just go straight for the current

$$\begin{aligned} J^\alpha &= \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \psi)} \delta\psi + \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \bar{\psi})} \delta\bar{\psi} \\ &= -\lambda \bar{\psi} \gamma^\mu \psi. \end{aligned} \quad (10.94)$$

3. Find the conserved current for the lagrangian (KG with EM interaction)

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + [(\partial^\mu + ieA^\mu) \varphi^*][(\partial_\mu - ieA_\mu) \varphi] - m^2 \varphi^* \varphi. \quad (10.95)$$

This Lagrangian is invariant under the global symmetry transformation given by

$$\varphi \Rightarrow \varphi e^{iq\theta} \quad \text{and} \quad \varphi^* \Rightarrow \varphi^* e^{-iq\theta}. \quad (10.96)$$

The infinitesimal transformations is then

$$\varphi \Rightarrow \varphi(1 + iq\theta) \quad \text{and} \quad \varphi^* \Rightarrow \varphi^*(1 - iq\theta). \quad (10.97)$$

Using this in Noethers theorem

$$\begin{aligned} J^\alpha &= \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi)} \delta\varphi + \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \varphi^*)} \delta\varphi^* \\ &= iq\theta g^{\mu\nu} [\varphi \delta_\mu^\alpha (\partial_\nu + ieA_\nu) \varphi^* - \varphi^* \delta_\nu^\alpha (\partial_\mu - ieA_\mu) \varphi] \\ &= iq\theta [\varphi (\partial^\alpha + ieA^\alpha) \varphi^* - \varphi^* (\partial^\alpha - ieA^\alpha) \varphi] \\ &= iq\theta [\varphi \partial^\alpha \varphi^* + ieA^\alpha \varphi \varphi^* - \varphi^* \partial^\alpha \varphi + ieA^\alpha \varphi^* \varphi] \\ &= iq\theta [\varphi \partial^\alpha \varphi^* - \varphi^* \partial^\alpha \varphi + 2ieA^\alpha \varphi^* \varphi] \\ &= -iq\theta [\varphi^* \partial^\alpha \varphi - \varphi \partial^\alpha \varphi^* - 2ieA^\alpha \varphi^* \varphi]. \end{aligned} \quad (10.98)$$

Noethers theorem is a very useful and fast way to derive the conserved currents. It is however not the only way. An alternative method will be used in the next section so as to illustrate the power of Noethers theorem.

Example 10.14.

Derive the conserved current and charge for the KG equation with the minimal coupling.

Begin by considering the KG equation with the minimal coupling

$$\left(g^{\mu\nu} (\partial_\nu - \frac{ie}{\hbar c} A_\nu) (\partial_\mu - \frac{ie}{\hbar c} A_\mu) + \frac{m^2 c^2}{\hbar^2} \right) \psi = 0, \quad (10.99)$$

where $\partial_\mu = (\frac{1}{c}\partial_t, \vec{\nabla})$, $A_\mu = (-\phi, \vec{A})$, $g^{\mu\nu} = g_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$. To derive the conserved current use

$$\psi^* (\text{equation}) - \psi (\text{equation})^* = 0, \quad (10.100)$$

where "(equation)" is the equation in question, could be KG or SE. This equation is then rewritten to be on the form of the continuity equation

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{J} = \partial_\nu J^\nu = 0. \quad (10.101)$$

From this comparison the conserved charge and current can be found. By using equation 10.99 in equation 10.100

$$\psi^* \left(g^{\mu\nu} (\partial_\nu - \frac{ie}{\hbar c} A_\nu) (\partial_\mu - \frac{ie}{\hbar c} A_\mu) + \frac{m^2 c^2}{\hbar^2} \right) \psi - \psi \left(g^{\mu\nu} (\partial_\nu + \frac{ie}{\hbar c} A_\nu) (\partial_\mu + \frac{ie}{\hbar c} A_\mu) + \frac{m^2 c^2}{\hbar^2} \right) \psi^* = 0. \quad (10.102)$$

It is immediately clear that the m^2 -terms cancel out. The rest is written out

$$0 = g^{\mu\nu} \left(\psi^* \partial_\nu \partial_\mu (\psi) - \frac{ie}{\hbar c} \psi^* \partial_\nu (A_\mu \psi) - \frac{ie}{\hbar c} \psi^* A_\nu \partial_\mu (\psi) - \frac{e^2}{\hbar^2 c^2} \psi^* A_\nu A_\mu \psi \right) \\ - g^{\mu\nu} \left(\psi \partial_\nu \partial_\mu (\psi^*) + \frac{ie}{\hbar c} \psi \partial_\nu (A_\mu \psi^*) + \frac{ie}{\hbar c} \psi A_\nu \partial_\mu (\psi^*) - \frac{e^2}{\hbar^2 c^2} \psi A_\nu A_\mu \psi^* \right). \quad (10.103)$$

The two last terms from each line cancels, and the Minkowski metric can be taken outside a parenthesis. Hereby

$$0 = g^{\mu\nu} \left(\psi^* \partial_\nu \partial_\mu (\psi) - \psi \partial_\nu \partial_\mu (\psi^*) - \frac{ie}{\hbar c} \left(\psi^* \partial_\nu (A_\mu \psi) + \psi^* A_\nu \partial_\mu (\psi) + \psi \partial_\nu (A_\mu \psi^*) + \psi A_\nu \partial_\mu (\psi^*) \right) \right) \\ = g^{\mu\nu} \left(\partial_\nu \left(\psi^* \partial_\mu (\psi) - \psi \partial_\mu (\psi^*) \right) - \frac{2ie}{\hbar c} \partial_\nu \left(\psi^* A_\mu \psi \right) \right) \\ = g^{\mu\nu} \partial_\nu \left(\psi^* \partial_\mu (\psi) - \psi \partial_\mu (\psi^*) - \frac{2ie}{\hbar c} \psi^* A_\mu \psi \right). \quad (10.104)$$

So, by comparison between equation 10.101 and equation 10.104 it is clear that

$$J^\mu = g^{\mu\nu} J_\nu \propto g^{\mu\nu} \left(\psi^* \partial_\nu (\psi) - \psi \partial_\nu (\psi^*) - \frac{2ie}{\hbar c} \psi^* A_\nu \psi \right). \quad (10.105)$$

To compare with equation 10.98 the current is taken in natural units

$$J^\nu \propto \psi^* \partial^\nu \psi - \psi \partial^\nu \psi^* - 2ie A^\nu \psi^* \psi. \quad (10.106)$$

It is clear by comparison between equation 10.98 and 10.106 that the results are in agreement and the Noethers theorem was easier to use! Continuing with the current density to find the conserved charge and current use that by definition

$$J_\mu = \frac{i\hbar e}{2m} \left(\psi^* \partial_\mu (\psi) - \psi \partial_\mu (\psi^*) - \frac{e^2}{2m} A_\mu \psi \psi^* \right). \quad (10.107)$$

From the definition of $j_\nu = (c\rho, -\vec{j})$

$$\rho = \frac{i\hbar e}{2mc^2} \left(\psi^* \partial_t (\psi) - \psi \partial_t (\psi^*) + \frac{e^2}{2m} \phi \psi \psi^* \right), \quad (10.108)$$

where the sign on the last term can vary from definition to definition.

$$\vec{J} = -\frac{i\hbar e}{2m} \left(\psi^* \vec{\nabla} (\psi) - \psi \vec{\nabla} (\psi^*) + \frac{e^2}{2m} \vec{A} \psi \psi^* \right). \quad (10.109)$$

11 | General Relativity

The theory that describes the distortion of space-time is general relativity (GR). GR describes the physics of accelerated reference frames (equivalent to reference frames with gravity), and introduces the notion of space-time as a merged entity¹ which is distorted by mass (and energy). In the distorted space-time, particles follow the shortest distance between two points²; the geodesic. The geodesic equation can be derived in three different ways, each introducing new, central principles to GR. The geodesic equation in turn depends on the gravitational field which can be determined from the Einstein equation.

11.1 The geodesic equation via the principle of equivalence

The first way to derive the geodesic equation is from the principle of equivalence. The principle of equivalence is the central principle upon which GR is built, and it informs how to relate SR to GR - i.e. how to map SR onto GR. Knowing the equations of SR, these can be generalized to GR via the principle of equivalence. The principle of equivalence is split into two; the weak principle of equivalence and the strong principle of equivalence. The weak principle of equivalence is the statement of the experimental fact that the gravitational mass is equal to the inertial mass. The equality of these masses mean that the effect of a homogeneous, static gravitational field can be transformed away by a coordinate transformation. To illustrate this, consider the resulting force on a particle in a homogeneous, static gravitational field in the Newtonian limit

$$m_I \frac{d^2x}{dt^2} = m_G g + \sum_i F_i. \quad (11.1)$$

The gravitational force can be transformed away by making a non-Galilean coordinate transformation on the form

$$x' = x - \frac{1}{2}gt^2 \quad , \quad t = t'. \quad (11.2)$$

From this transformation

$$m_I \left(\frac{d^2x'}{dt'^2} + g \right) = m_G g + \sum_i F'_i. \quad (11.3)$$

Taking $m_I = M_G$ it is clear that the gravitational field-terms cancel and leaves

$$m_I \frac{d^2x'}{dt'^2} = \sum_i F'_i. \quad (11.4)$$

Hence, the effect of the homogeneous, static gravitational field can be transformed away by an appropriate coordinate transformation. If the gravitational field is inhomogeneous and time-dependent the effects of the gravitational field cannot be transformed away in a macroscopic free falling elevator. The effect of the gravitational field can however be transformed away in a local reference frame; a reference frame which is confined to a sufficiently small region in space-time. This is the strong equivalence principle.

¹Space and time is merged into space-time.

²Not necessarily a straight - contrary to intuition.

The strong equivalence principle: *In an inhomogeneous, time-dependent gravitational field, there exists a local inertial reference frame (LIRF), i.e. a reference frame confined to a sufficiently small region in space-time, in which the effect of the gravitational field is transformed away and the laws of SR are valid.*

The strong equivalence principle intuitively makes especially much sense from a geometrical view point; in GR space-time is modeled as a four dimensional Lorentzian manifold. A manifold is a topological space that locally resembles Euclidean space. Another way to state this is to say that in GR energy and mass curves space-time and the curved space-time is what is called a four dimensional Lorentzian manifold. For an arbitrary curved geometry one can always confine oneself to a sufficiently small region in which the geometry is asymptotically flat. Flat space-time is Minkowski space; i.e. the space of SR.

As mentioned; from the strong equivalence principle the equations of SR can be generalized to GR, and amongst those equations is the EOM for a particle in free fall in a gravitational field; the geodesic equation. Denoting the coordinates in the LIRF by ξ^α , the relativistic force in the LIRF can be written as

$$\frac{d^2\xi^\alpha}{d\tau^2} = 0, \quad (11.5)$$

where $d\tau^2 = -ds^2$ (natural units) is the proper time and $ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu$ is the square of the infinitesimal length measure in curved space-time. This equation can be transformed to another reference frame by taking $\xi^\alpha = \xi^\alpha(x)$. Hereby

$$\begin{aligned} 0 &= \frac{d^2\xi^\alpha}{d\tau^2} \\ &= \frac{d}{d\tau} \left(\frac{\partial\xi^\alpha}{\partial x^\mu} \frac{dx^\mu}{d\tau} \right) \\ &= \underbrace{\frac{\partial^2\xi^\alpha}{\partial x^\mu \partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}}_{\equiv \Gamma_{\mu\nu}^\lambda} + \frac{\partial\xi^\alpha}{\partial x^\mu} \frac{d^2x^\mu}{d\tau^2}. \end{aligned} \quad (11.6)$$

To get rid of the $\frac{\partial\xi^\alpha}{\partial x^\mu}$ -factor multiply with $\frac{\partial x^\lambda}{\partial\xi^\alpha}$ and use that $\frac{\partial x^\lambda}{\partial\xi^\alpha} \frac{\partial\xi^\alpha}{\partial x^\mu} = \delta_\mu^\lambda$. Hereby

$$\underbrace{\frac{\partial x^\lambda}{\partial\xi^\alpha} \frac{\partial^2\xi^\alpha}{\partial x^\mu \partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}}_{\equiv \Gamma_{\mu\nu}^\lambda} + \frac{d^2x^\lambda}{d\tau^2} = 0. \quad (11.7)$$

$\Gamma_{\mu\nu}^\lambda$ (symmetric in the lower indices) is defined as the *affine connection*, also called the Christoffel symbol. From equation (11.7) it is clear that the last term is the term that appears in the LIRF and so the first term must be a result of the gravitational field. Therefore, the Christoffel symbol can be expressed in terms of the metric tensor - the metric tensor which represents the gravitational field. In the LIRF the EOM must reduce to the ordinary derivative, i.e. the first term must vanish, and so the Christoffel symbol must vanish here. Since the spacetime manifold is locally flat in the LIRF, and the metric of flat spacetime is the Minkowski metric, the metric in the LIRF can be expanded cf.

$$g_{\mu\nu} = \eta_{\mu\nu} + \mathcal{O}(x^2), \quad (11.8)$$

where the first derivative vanish since the spacetime manifold is to be locally flat. Since the first derivatives vanish in the LIRF, and the Christoffel symbol should vanish here as well, it makes sense that the Christoffel symbol should be given in terms first derivatives of the metric tensor. That is

$$\Gamma_{\mu\nu}^\lambda \sim \partial g_{\mu\nu}. \quad (11.9)$$

In particular it can be found that (Weinberg, 1972)

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2}g^{\lambda\alpha} \left(\frac{\partial g_{\mu\alpha}}{\partial x^\nu} + \frac{\partial g_{\nu\alpha}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \right), \quad (11.10)$$

11.2 The geodesic equation via the principle of general covariance

The second way to derive the geodesic equation is from the principle of general covariance. The principle of general covariance states that an equation is valid in a generic gravitational field if two conditions are fulfilled

1. The equation is valid in the LIRF, i.e. the equation is valid in SR.
2. The equation is generally covariant, that is it preserves its form under a coordinate transformation.

By using the principle of general covariance the equations of GR can be obtained by going to the LIRF, formulate the equations cf. SR, make them covariant and promote them to be valid in a generic reference frame. Making an equation covariant means expressing it in terms of scalars, vectors, tensors, covariant derivatives and introducing the invariant integration measure into integrals. In SR the equations are already expressed in terms of scalars, vectors and tensors, so what remains is to introduce the covariant derivative and the invariant integration measure. The covariant derivative is introduced because the ordinary differentiation of a tensor, in general, does not result in another tensor. This needs to be so if the equations are to be covariant. Therefore a new definition of the derivative is introduced. The expression for the covariant derivative can be deduced in many ways. The one I like the most is from the geodesic equation. The geodesic equation is obtained from the derivative of the velocity in the LIRF along the curve $\xi^\alpha(\tau)$. Making this covariant means that the derivative must go to a covariant derivative along the curve. Therefore

$$\begin{aligned} \frac{DU^\mu}{D\tau} &= \frac{dx^\beta}{d\tau} U^\mu_{;\beta} \\ &= \frac{dx^\beta}{d\tau} \left(\frac{\partial U^\mu}{\partial x^\beta} + \Gamma_{\alpha\beta}^\mu U^\alpha \right). \end{aligned} \quad (11.11)$$

The covariant derivative along a curve is the GR-generalization of the material derivative known from fluid mechanics. From the covariant derivative along a curve the covariant derivative of a generic covariant vector is defined viz

$$A^\mu_{;\beta} \equiv \frac{dA^\mu}{dx^\beta} + \Gamma_{\alpha\beta}^\mu A^\alpha. \quad (11.12)$$

Example 11.1.

Determine Newtons second law in the GR limit and take the NR limit.

I use that the physical laws can be transformed from the SR regime to GR by replacing a time-derivative by the covariant derivative along a curve. Hereby³

$$\vec{F} = m\vec{a} \Rightarrow F^\mu = m \frac{Du^\mu}{D\tau}, \quad (11.13)$$

where $u^\mu \doteq \gamma(v) [1 \ \vec{v}]^T$ is the relativistic velocity and $v = |\vec{v}|$ is the relative velocity which makes the considered frame of reference relativistic. Non-relativistic velocities imply that $\frac{dx^i}{dt} \ll 1 \Leftrightarrow dx^i \ll dt$ (natural units). This inequality implies that

$$\frac{dx^i}{d\tau} \ll \frac{dt}{d\tau}. \quad (11.14)$$

This means that in the the non-relativistic limit

$$\begin{aligned} \frac{Du^\mu}{D\tau} &= \frac{dx^\beta}{d\tau} \left(\frac{\partial u^i}{\partial x^\beta} + \Gamma_{\alpha\beta}^i u^\alpha \right) \\ &= \frac{d^2x^\mu}{d\tau^2} + \Gamma_{\alpha\beta}^\mu \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} \\ &\simeq \frac{d^2x^\mu}{d\tau^2} + \Gamma_{00}^\mu \frac{dx^0}{d\tau} \frac{dx^0}{d\tau}. \end{aligned} \quad (11.15)$$

³Note however that in GR gravity is the structure of space-time and not a forces. However, the structure of space-time effects the movement of particles freely falling and as such it appears (to us as naive observers) as a force.

Hence

$$\begin{aligned}
\frac{Du^\mu}{D\tau} &= \frac{dx^\beta}{d\tau} \left(\frac{\partial u^i}{\partial x^\beta} + \Gamma_{\alpha\beta}^i u^\alpha \right) \\
&= \frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\alpha\beta}^\mu \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} \\
&\simeq \frac{d^2 x^\mu}{d\tau^2} + \Gamma_{00}^\mu \frac{dx^0}{d\tau} \frac{dx^0}{d\tau} \\
&\simeq \left(\frac{d^2 x^\mu}{dt^2} + \Gamma_{00}^\mu \right) \left(\frac{dx^0}{d\tau} \right)^2.
\end{aligned} \tag{11.16}$$

Hereby, in the non-relativistic limit

$$F^i \simeq m \left(\frac{d^2 x^i}{dt^2} + \Gamma_{00}^i \right) \left(\frac{dx^0}{d\tau} \right)^2. \tag{11.17}$$

The Γ_{00}^i -term is the force due to the gravitational field. In GR the gravity is the structure of space-time and not a force. Therefore gravity is not included in the force terms above. In the non-relativistic limit gravity is interpreted as a force and here $dt = \gamma d\tau \simeq d\tau$. Hence

$$F^i \simeq ma^i, \tag{11.18}$$

where a^i can be identified as follows

$$a^i = \frac{d^2 x^i}{dt^2} + \Gamma_{00}^i. \tag{11.19}$$

In order to obtain the Newtonian limit from the non-relativistic limit the gravitational field is taken to be static, i.e. $\frac{\partial g_{\mu\nu}}{\partial x^0} = 0$, and weak, i.e. $g_{\mu\nu} \simeq \eta_{\mu\nu} + h_{\mu\nu}$. Hereby $\Gamma_{00}^0 = 0$ and $\Gamma_{00}^i = -\frac{1}{2} \frac{\partial h_{00}}{\partial x^i}$. Using the geodesic equation (i.e. setting the relativistic force to zero) and identifying with the Newtonian result the Newtonian expression for the gravitational field can be determined (Weinberg, 1972)[p.77].

The covariant derivative can also be derived from the transformation properties of the covariant derivative (Weinberg, 1972). Such a derivation motivates the definition of the covariant derivative of a contra-variant vector. Since the contra-variant vector transforms, cf. equation (9.29), inversely as compared to the covariant vector, and the first order approximation to the inverse is a minus sign on the second term⁴, the covariant derivative of a contra-variant vector is given by

$$A_{\mu;\nu} = \frac{\partial A_\mu}{\partial x^\nu} - \Gamma_{\mu\nu}^\alpha A_\alpha. \tag{11.20}$$

As for the integration measure, it transforms cf.

$$d^4 x \Rightarrow d^4 x' = \left| \det \left(\frac{\partial x'}{\partial x} \right) \right| d^4 x, \tag{11.21}$$

where $\det(\frac{\partial x'}{\partial x})$ is the Jacobian of the transformation and $\left| \det(\frac{\partial x'}{\partial x}) \right|$ is the magnitude of the Jacobian. Since then ordinary measure transforms we need to introduce something that will eliminate this transformation. Use now

$$g_{\mu\nu} \Rightarrow g'_{\mu\nu} = \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} g_{\alpha\beta}. \tag{11.22}$$

Defining $-\det(g_{\mu\nu}) = g$ and taking the determinant of the above equation

$$g \Rightarrow g' = \det \left(\frac{\partial x}{\partial x'} \right)^2 g. \tag{11.23}$$

This can be used to construct the invariant integration measure

$$\sqrt{g} d^4 x \Rightarrow \sqrt{g'} d^4 x' = \sqrt{g} d^4 x. \tag{11.24}$$

⁴To illustrate this, consider: $g_{\alpha\mu} g^{\mu\beta} = (\eta_{\alpha\mu} + h_{\alpha\mu})(\eta^{\mu\beta} + h^{\mu\beta}) = \eta_{\alpha\mu} \eta^{\mu\beta} + \mathcal{O}(h^2)$.

11.3 The geodesic equation via the variational principle

The third way to derive the geodesic equation is via the variational principle. In the distorted space-time of GR, particles follow the shortest distance between two points; the geodesic. The principle of the geodesic is similar to the principle of stationary action familiar from Lagrangian mechanics, and so the path of a particle can be determined from the same variational calculus. To determine the equations of motion (EOM's) for particles in curved space-time the variation of the infinitesimal length measure in curved space-time, ds , is considered. ds can be described in terms of the metric tensor, a tensor describing the curvature of space-time, via⁵

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu, \quad (11.25)$$

where Einstein summation has been employed and dx^μ, dx^ν are the coordinates in some generic coordinate system - not in general an Euclidean coordinate system. In the case of a flat space-time, as is considered in SR, the metric tensor reduces to the Minkowski tensor⁶

$$\eta^{\mu\nu} \doteq diag(-1, 1, 1, 1). \quad (11.26)$$

In analogy with Lagrangian mechanics the infinitesimal length element, ds , which plays the role of the action in Lagrangian mechanics, is expressed in terms of a (by analogy) Lagrangian⁷

$$\begin{aligned} ds &= \sqrt{g_{\mu\nu}(x)dx^\mu dx^\nu} \\ &= \int ds \\ &= \int \frac{ds}{d\lambda} d\lambda \\ &= \int L(x, \dot{x}) d\lambda, \end{aligned} \quad (11.27)$$

where λ in general is some parameterization of the path of a particle through space-time, the position dependence of $L = L(x, \dot{x})$ comes from the metric and $\dot{x} = \frac{dx}{d\lambda}$. In most cases the path will be parameterized in terms of the proper time (τ), defined as the time experienced by an observer in his rest frame. By identification

$$L(x, \dot{x}) = \sqrt{g_{\mu\nu}(x) \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}}. \quad (11.28)$$

However, since the EOM's obtained in the end are independent of the square root in equation (11.28), the Lagrangian is conventionally defined as (Cheng, 2010)

$$\begin{aligned} L(x, \dot{x}) &\equiv \left(\frac{ds}{d\tau} \right)^2 \\ &= g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \\ &= g_{\mu\nu}(x) \dot{x}^\mu \dot{x}^\nu \\ &= -1, \end{aligned} \quad (11.29)$$

where $\lambda \Rightarrow \tau$ has been employed and it has been used that $ds^2 = -d\tau^2$ for the last equality. The last equality illustrates that the Lagrangian in GR is a constant of the motion. To determine the shortest path through

⁵A notational note: x is both used for the 4-vector and the Cartesian coordinate. A rule of thumb is that whenever y, z enters, x is the Cartesian coordinate, otherwise the 4-vector.

⁶Note that there is a freedom in choosing the Minkowski metric to correspond to a time-inversion or parity operation. Since the physical theory under consideration should be invariant under both these operations, the definition of the metric has some freedom. In GR it is convention to use $\eta^{\mu\nu} \doteq diag(-1, 1, 1, 1)$ where in QFT it is convention to use $\eta^{\mu\nu} \doteq diag(1, -1, -1, -1)$.

⁷Often in GR one needs to include a factor of $\sqrt{\det(g)}$ in order to make the integral measure covariant. This is however not necessary in this case since the infinitesimal line elements is already covariant.

in the curved space-time the length element, expressed in terms of the Lagrangian of equation (11.29), is varied whilst keeping the end points fixed

$$\begin{aligned}
\delta s &= \int \delta L(x, \dot{x}) d\tau \\
&= \int_{\tau_1}^{\tau_2} \left(\frac{\partial L}{\partial x^\mu} dx^\mu + \frac{\partial L}{\partial \dot{x}^\mu} d\dot{x}^\mu \right) d\tau \\
&= \int_{\tau_1}^{\tau_2} \underbrace{\left(\frac{\partial L}{\partial x^\mu} - \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{x}^\mu} \right) \right)}_{=0} dx^\mu d\tau \\
&= 0,
\end{aligned} \tag{11.30}$$

where integration by parts has been used alongside the fact that the boundary term vanishes when the endpoints of the path are kept fixed and the parenthesis that has been indicated to be zero has been done so in retrospect - since the entire expression is to be zero for arbitrary dx^μ . The part of the integrand that has been required to vanish is the analogue to the Euler-Lagrange equation of classical mechanics. In GR these equations are coined the geodesic equations (GE's). To be explicit

$$\text{The GE's: } \frac{\partial L}{\partial x^\mu} - \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{x}^\mu} \right) = 0. \tag{11.31}$$

The GE's describe the EOM's for an observer in free fall in a gravitational field. By using equation (11.29) in equation (11.31) the GE's can be written out in terms of the metric tensor

$$g_{\mu\alpha} \ddot{x}^\mu + \frac{\partial g_{\mu\alpha}}{\partial x^\beta} \dot{x}^\beta \dot{x}^\mu - \frac{1}{2} \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \dot{x}^\mu \dot{x}^\nu = 0. \tag{11.32}$$

By introducing the Christoffel symbol ($\Gamma_{\mu\nu}^\alpha$) equation (11.32) can be written as

$$\ddot{x}^\alpha + \Gamma_{\mu\nu}^\alpha \dot{x}^\mu \dot{x}^\nu = 0, \tag{11.33}$$

where

$$\Gamma_{\mu\nu}^\alpha = \frac{1}{2} g^{\alpha\beta} \left(\frac{\partial g_{\beta\nu}}{\partial x^\mu} + \frac{\partial g_{\beta\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\beta} \right). \tag{11.34}$$

Example 11.2.

Consider Alice and Bob near a black hole. Bob falls into the black hole and Alice observes from afar. How does Bob see experience his journey? How does Alice experience Bob's journey?

I take Bob to be a particle in free-fall in a gravitational field. Hence, his EOM is determined from the geodesic equation. Bob will experience his journey as $r(\tau)$ (I take him to fall in radially) and Alice will experience Bob's journey as $r(t)$.

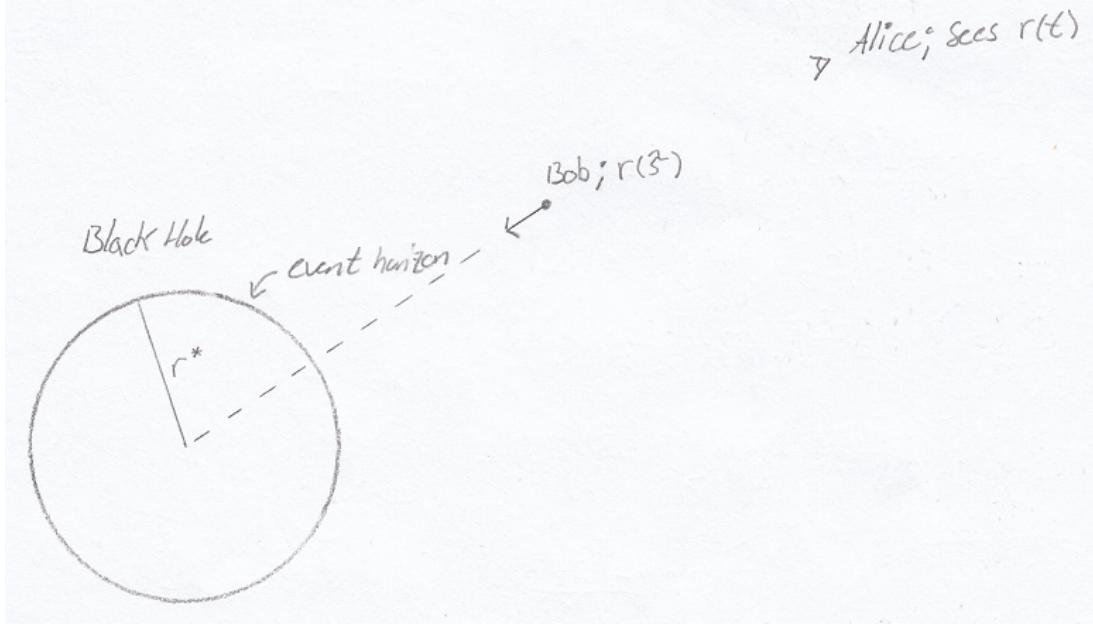


Figure 11.1: The scenario with the black hole, Bob and Alice.

I can find $r(\tau)$ and $r(t)$ by using the variational principle. To this end I need the Lagrangian of the gravitational field. The Lagrangian is found from the metric, cf. equation (11.29), which is the Schwarzschild metric. The Schwarzschild metric describes the gravitational field from a spherically symmetric source, e.g. a planet, star, black holes,.... It is given by

$$ds^2 = -f(r)dt^2 + f(r)^{-1}dr^2 + r^2d\Omega^2, \quad (11.35)$$

where

$$f(r) = 1 - \frac{2MG}{r}. \quad (11.36)$$

Hereby the Lagrangian can be determined

$$L = -f(r)\dot{t}^2 + f(r)^{-1}\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin(\theta)^2\dot{\varphi}^2 = -1. \quad (11.37)$$

Since I take Bob to fall in radially; $\dot{\varphi} = \dot{\theta} = 0$. Hereby

$$L = -f(r)\dot{t}^2 + f(r)^{-1}\dot{r}^2 = -1. \quad (11.38)$$

From the Lagrangian I can determine the EOM of t, r via the Euler-Lagrange equation. However, in this any most other cases, it is easier to combine the use of EOM's from the Euler-Lagrange equation and $L = -1$ - this last property is unique to GR. In this case it is (seen in hindsight) wise to begin with the EOM for the time-component

$$\frac{\partial L}{\partial t} - \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{t}} \right) = 0 \Rightarrow \frac{d}{d\tau} \left(\underbrace{f(r)\dot{t}}_{\equiv \tilde{L}} \right) = 0. \quad (11.39)$$

From the above I can determine \dot{t}^2 and insert it in L . Hereby

$$\tilde{L}^2 - \dot{r}^2 = f(r). \quad (11.40)$$

To solve this equation, I use that $L = const$ and that it can be evaluated in the limit of $r \Rightarrow \infty$. In this limit $f(r) \Rightarrow 1$ and $\dot{t} \Rightarrow 1$, and so $\tilde{L} = 1$. Hereby

$$\tilde{L}^2 - \dot{r}^2 = f(r) \Rightarrow \frac{dr}{d\tau} = -\sqrt{\frac{2MG}{r}}, \quad (11.41)$$

where the minus sign comes from the fact that Bob is falling into the black hole - so it is directional. From the above I find via separation of variables

$$\sqrt{r}dr = -\sqrt{2MG}d\tau \Rightarrow r(\tau) \propto \tau^{\frac{2}{3}}. \quad (11.42)$$

The prefactor of $r(\tau)$ is not important in this case. What is important is the fact that there is no singularity at the event horizon. Bob experience himself falling through the event horizon without any changes. The question is now whether Alice will agree with Bob! To find $r(t)$ I use that

$$\frac{dr}{dt} = \frac{\dot{r}}{\dot{t}} = -\sqrt{\frac{2MG}{r}} \left(1 - \frac{2MG}{r}\right). \quad (11.43)$$

In the limit of $r \Rightarrow 2MG$ the above differential equation has a solution on the form⁸

$$r \approx 2MG + \text{const} \cdot e^{-\frac{t}{2MG}}. \quad (11.44)$$

What is noteworthy here is that as $t \Rightarrow \infty$ Alice will see Bob approach $r = 2MG$, but never cross it! Hence, Bob and Alice will disagree on what transpires. This disagreement is due to time-dilation.

11.4 The Einstein equation

The Einstein equation is the equation of motion for the metric, i.e. the gravitational field. The derivation of the Einstein equation revolves around the generalization of the classical equation that relates the gravitational field to the mass distribution; i.e. the Poisson equation

$$\nabla^2\phi = 4\pi G_N\rho. \quad (11.45)$$

In the relativistic limit the gravitational field and mass-energy is represented by the metric and energy-momentum tensor, respectively. Therefore, the starting point is an equation on the form

$$f(g_{\mu\nu}) = g(T_{\mu\nu}), \quad (11.46)$$

where f and g are *some* functions of the metric and the (symmetric) energy-momentum tensor, respectively. Because $T^{00} = \rho$; $g \propto T_{\mu\nu}$. As for f ; because of the classical limit it is expected that $f \supset \partial^2 g^{\mu\nu}$. Furthermore, since the gravitational field can store energy, it is expected that field strength terms are included in f , i.e. $f \supset (\partial g_{\mu\nu})^2$. To be even more specific use that the mass-energy distribution curves space-time, so it is expected that f is related to curvature. The curvature is measured by the Riemann curvature tensor that is defined in terms of the commutator of the covariant derivatives

$$[D_\mu, D_\nu]V_\alpha \equiv R^\sigma_{\mu\beta\alpha}V_\sigma, \quad (11.47)$$

where

$$R^\sigma_{\alpha\mu\beta} = \frac{\partial\Gamma^\sigma_{\alpha\mu}}{\partial x^\beta} - \frac{\partial\Gamma^\sigma_{\alpha\beta}}{\partial x^\mu} + \Gamma^\eta_{\alpha\mu}\Gamma^\sigma_{\beta\eta} - \Gamma^\eta_{\alpha\beta}\Gamma^\sigma_{\mu\eta}. \quad (11.48)$$

Because of Lorentz covariance f cannot be directly proportional to $R^\sigma_{\alpha\mu\beta}$. Instead $f \sim g_{\sigma\mu}R^\sigma_{\alpha\mu\beta} = R^\mu_{\alpha\mu\beta}$ - where $R^\mu_{\alpha\mu\beta} = R_{\alpha\beta}$ is called the Ricci tensor. Therefore, a first guess about the form of the Einstein equation is

$$R^\mu_{\alpha\mu\beta} \sim T_{\alpha\beta}. \quad (11.49)$$

However, as will be shown later on, the energy momentum tensor is the conserved Noether current associated with space-time translations in SR. This means⁹

$$\text{SR: } T^{\mu\nu}_{,\mu} = 0 \Rightarrow \text{GR: } T^{\mu\nu}_{;\mu} = 0. \quad (11.50)$$

⁸Found by Taylor-expanding around $r = 2MG$.

⁹, μ denotes a partial derivative w.r.t. μ ; μ denotes a covariant derivative w.r.t. μ .

This means that the covariant derivative of the Einstein equation should vanish. However, $R^{\mu\nu}{}_{;\mu} \neq 0$, so $R^{\mu\nu} \propto T^{\mu\nu}$ is not the form of the Einstein equation. However, f should still be related to the curvature in the form of the Ricci tensor. This conundrum is fixed by evaluating the covariant derivative of the Ricci tensor and subtracting the result - hence, a quantity related to the curvature will with a vanishing covariant derivative is obtained. The covariant derivative of the Ricci tensor can be evaluated by brute force, however this approach is a bit tedious since the Ricci tensor has four terms and the covariant derivative has three terms (an extra term since the covariant derivative is of a tensor with two indices). Instead use the Bianchi identity

$$R_{\alpha\mu\nu\beta;\eta} + R_{\alpha\mu\eta\nu;\beta} + R_{\alpha\mu\beta\eta;\nu} = 0. \quad (11.51)$$

Multiply this with $g^{\alpha\nu}$ and use that the covariant derivative of the metric vanishes. Hereby the metric can be pulled inside the covariant derivative to act on the Riemann curvature tensor. The result is given by

$$R_{\mu\beta;\eta} - R_{\mu\eta;\beta} + R^\nu{}_{\mu\beta\eta;\nu} = 0. \quad (11.52)$$

where the minus sign comes from permuting the outer indices of the Riemann curvature tensor. Multiplying once more with $g^{\mu\beta}$

$$R_{;\eta} - R^\beta{}_{\eta;\beta} - R^\nu{}_{\eta;\nu} = 0. \quad (11.53)$$

Factorizing the covariant derivative and rewriting a bit

$$\left(\frac{1}{2} \delta^\beta_\eta R - R^\beta_\eta \right)_{;\eta} = 0 \Leftrightarrow \left(R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R \right)_{;\eta} = 0, \quad (11.54)$$

where R is the Ricci scalar - the contracted version of the Ricci tensor. The last equality is what is needed for the Einstein equation! Therefore

$$R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R \propto T^{\mu\nu}, \quad (11.55)$$

where the LHS is defined as the Einstein tensor

$$G^{\mu\nu} \equiv R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R. \quad (11.56)$$

The proportionality constant can be found from the classical limit¹⁰. Hereby the Einstein equation is written as

$$G^{\mu\nu} = -8\pi G_N T^{\mu\nu}, \quad (11.57)$$

where the minus sign comes from the definition of the Ricci tensor. The Einstein equation, as it stands in equation (11.57), is however not completely generic, since the constraint that the LHS only contain second derivatives is a bit too strong. Higher derivatives would be suppressed too much, but a term proportional to the metric would be allowed as long as it vanishes in the classical limit. So, in the most generic form, to second order in the derivatives, the Einstein equation is given by

$$G^{\mu\nu} + g^{\mu\nu} \Lambda = -8\pi G_N T^{\mu\nu}. \quad (11.58)$$

Λ is called the cosmological constant and it can be interpreted as a vacuum energy if it is moved to the RHS and taken as a contribution to the energy-momentum tensor. Equation (11.58) is the end product; it is the Einstein equation in its final form. The Einstein equation can be used to obtain the gravitational field from a given mass-energy distribution. The gravitational field can then be used in the geodesic equation to obtain the trajectories of particles in free fall in the gravitational field.

¹⁰Be careful here since there is a little conversion from ϕ to g^{00} - As it turns out to be. See Cheng (2010).

11.5 Standard Cosmology

The standard model of cosmology is built upon the FRW metric and the Friedmann equations derives therefrom. The FRW metric describes the gravitational field on cosmological scales, that is it describes the evolution of the universe as a whole. It is built from what is called the cosmological principle which approximates the universe as a perfect, spatially homogeneous, isotropic, fluid, as seen from the reference frame of a comoving observer¹¹, on very large scales. The particles in the perfect fluid are taken to be the gravitationally bound objects in the universe. From these assumptions the metric is developed by writing down the most generic metric and then applying the physics of the cosmological principle to simplify the metric.

In general the metric depends on the choice of reference system. Taking the reference frame of the comoving observer the universe is approximated as isotropic. Since an isotropic universe has spherical symmetry the metric will be diagonal in these coordinates. Lastly, the homogeneous condition can be enforced in order to find (Weinberg, 1972)

$$ds^2 = -dt^2 + R(t)^2 \left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2 \right], \quad (11.59)$$

where $d\Omega \equiv r^2 d\theta^2 + r^2 \sin(\theta)^2 d\phi^2$, r, θ, ϕ, t denotes radial coordinate, polar angle, azimuthal angle and time coordinate, respectively, $R(t)$ is a time-varying function called the scale factor and k is a constant which sign is related to the curvature of the universe. To see how, consider different characteristics for $k, R(t)$:

1. $k = 0, R(t) = \text{const} \Rightarrow ds^2 = -dt^2 + R^2(dr^2 + r^2d\Omega^2)$. Since $R = \text{const}$ it can be transformed away by a suitable coordinate transformation and the metric will thereby reduce to the Minkowski metric. The Minkowski metric results in a vanishing Riemann curvature tensor so, cf. Einsteins equation, the energy-momentum tensor will likewise vanish and there will be no matter in the universe. Hence, this is the case of an empty universe.
2. $k = 0, R(t) \neq \text{const} \Rightarrow ds^2 = -dt^2 + R^2(t)(dr^2 + r^2d\Omega^2)$. Here $R(t)$ cannot be transformed away by a coordinate transformation, and even though the spatial part of the metric is flat, space-time is curved. Hence, the curvature tensor does not vanish, and the universe is not empty in this case. From observations $k \simeq 0$ and so this is a good approximation of the universe.
3. $k > 0, R(t) \neq \text{const} \Rightarrow ds^2 = -dt^2 + R(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2 \right)$. Once more $R(t)$ cannot be transformed away and space has a positive curvature. The spatial part of the metric resembles that of a 3D hypersphere. Since the hypersphere has a finite surface area, this universe is classified as a closed universe - referring to the finite size of the universe.
4. $k < 0, R(t) \neq \text{const} \Rightarrow ds^2 = -dt^2 + R(t)^2 \left(\frac{dr^2}{1+|k|r^2} + r^2 d\Omega^2 \right)$. Here the spatial part of the metric resembles that of a 3D pseudo hypersphere. Since the pseudo hypersphere has infinite volume, the universe is classified as open - referring to the infinite size of the universe.

By using the metric in the Einstein equation the Friedmann equations can be found. Neglecting the vacuum energy, equation (11.56) and (11.57) can be combined to write the Einstein equation as follows

$$R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = -8\pi G_N T^{\mu\nu}. \quad (11.60)$$

Because the Ricci tensor includes a lot of terms the Einstein equation is often rewritten by eliminating Ricci scalar and instead introducing the trace of the energy-momentum tensor. This is done by taking the trace on both sides of equation (11.60). Hereby

$$R^{\mu\nu} = -8\pi G_N (T^{\mu\nu} - \frac{1}{2}g^{\mu\nu}T_\alpha^\alpha). \quad (11.61)$$

¹¹A comoving observer is one which is in free-fall in space-time, and for which the proper time is equal to the coordinate time, ie $g_{00} = -1$.

To obtain the equations for $R(t)$ use equation (11.59) in equation (11.61) to determine the Ricci tensor and use that the energy-momentum tensor for a perfect fluid is given by

$$T^{\mu\nu} = Pg^{\mu\nu} + (P + \rho)u^\mu u^\nu, \quad (11.62)$$

where P is the pressure of the fluid, ρ is the energy density of the fluid and u is the four-velocity. The energy-momentum tensor is the conserved Noether current associated with space-time translations in SR. Therefore the relativistic divergence of the energy-momentum tensor must vanish in SR. Generalizing from SR to GR the derivative is generalized to the covariant derivative. Therefore

$$T^{\mu\nu}_{;\nu} = 0. \quad (11.63)$$

The $\mu = 0$ component of equation (11.63) reveals

$$d(\rho R^3) = -Pd(R^3) \Leftrightarrow \dot{\rho} + 3H(\rho + P) = 0, \quad (11.64)$$

where $H \equiv \frac{\dot{R}}{R}$. Equation (11.64) implies that the expansion of the universe can lead to local changes in the energy density. Note that the space-time geometry can hold energy and so energy can flow between it and matter - hence the energy of the space-time geometry has to be counted if one is to consider energy conservation. Taking $E = \rho V$ and $R^3 \sim V$ equation (11.64) can be rewritten as follows

$$dE + pdV = 0. \quad (11.65)$$

Equation (11.65) is the first law of thermodynamics with no heat exchange ($dQ = 0$). Hence, cf. equation (11.65), the FRW metric implies that the universe expands adiabatically ($dQ = 0$). Approximating the processes that occur to be reversible means the expansion is also isentropic, i.e. $dQ_{rev} = TdS = 0 \Rightarrow dS = 0$. The approximation of reversible processes is equivalent to the approximation of local thermal equilibrium in the universe. It is important to emphasize this approximation since it is key in the established formalism that treats DM.

Using $P = w\rho$ with $w = \text{const}$ in equation (11.64) reveals $\rho \propto R^{-3(1+w)}$. w depends on which component dominates the universe. The popular considerations are radiation ($w = \frac{1}{3}$), matter ($w = 0$) and vacuum energy ($w = -1$)

$$\begin{aligned} \text{Radiation: } & \rho \propto R^{-4}, \\ \text{Matter: } & \rho \propto R^{-3}, \\ \text{Vacuum energy: } & \rho = \text{const}. \end{aligned} \quad (11.66)$$

In this thesis the early universe, which was radiation dominated, will primarily be considered and so $\rho \propto R^{-3}$ will often be assumed. By evaluating the different components of equation (11.61); One from the 00-component and one from the ij -components

$$\left(\frac{\dot{R}}{R}\right)^2 \equiv H^2 = \frac{8\pi G_N}{3}\rho - \frac{k}{R^2}, \quad \frac{\ddot{R}}{R} = -\frac{4\pi G_N}{3}(\rho + 3P). \quad (11.67)$$

Equation (11.67) defines the Friedmann equations (the first to the left) which describe the relationship between curvature, matter density, pressure and the expansion of the universe. The first Friedmann equation is often rewritten by defining $\rho_c \equiv \frac{3H^2}{8\pi G_N}$. Hereby

$$H^2 = \frac{8\pi G_N}{3}\rho - \frac{k}{R^2} \Rightarrow \frac{k}{H^2 R^2} = \Omega - 1, \quad (11.68)$$

where $\Omega \equiv \frac{\rho}{\rho_c}$. Since $k \simeq 0$ is observed it is clear from equation (11.68) that $\Omega \simeq 1$.

Part IV

Quantum Field Theory applied to particle physics

Introduction

Elementary particle physics is the area of physics concerning the behavior of the elementary particles. An elementary particle is a particle that is assumed not to consist of any smaller particles. Hence, the proton (which consists of quarks) is not an elementary particle whereas the quarks are (since these are assumed not to be "separable"). Elementary particle physics are used in several ways and has several faces; one area of physics in which elementary particle physics are used is nuclear physics. Nuclear physics considers the interaction between particles that can be observed in the lab¹² in the non-relativistic limit. In the non-relativistic limit the elementary particles are confined in larger particles, and so the work with elementary particles are only indirectly. To examine the elementary particles themselves the larger particles need to be split apart. This is done by smashing the particles together at high velocities. Hence, the real at which elementary particle are examined is the relativistic limit. Since the elementary particles are quantum particles the proper formalism to describe elementary particles is quantum field theory (QFT). QFT is a versatile theoretical construct that can be used in several areas of physics. In this chapter I will focus on how it can be used to describe the elementary particles that appear in the standard model.

¹²Protons, neutrons, muons, electrons... The quarks go together to form particles and the leptons exist for themselves. Hence nuclear physics consider the interaction between such particles.

12 | Lagrangian Density of the SM

The SM is the theory that describes the physics of elementary particles. The elementary particles of the SM can be grouped as follows

$$\left\{ \begin{array}{l} \text{Quarks} \\ \text{Leptons} \end{array} \right\} \Rightarrow \text{Fermions, spin } \frac{1}{2} \Rightarrow \text{Makes up the "matter".} \quad (12.1)$$

$$\left\{ \begin{array}{l} g \\ W^\pm, Z^0 \\ \gamma \end{array} \right\} \Rightarrow \text{Vector bosons, spin 1} \Rightarrow \text{Makes up "interactions".} \quad (12.2)$$

$$\left\{ \text{Higgs} \right\} \Rightarrow \text{Scalar, spin 0} \Rightarrow \text{Gives mass to the fermions and } W^\pm, Z^0 \text{ bosons.} \quad (12.3)$$

where there are 8 gluons which mediates strong interaction and couples to quarks, the W^\pm, Z^0 bosons mediate the weak interaction and couples to quarks and leptons and the photon mediates EM interaction and couples to charged particles. These are the SM particles. The relevant theoretical construction used to describe the SM particles is QFT. In QFT particles are interpreted as excitations in fields, so to each particle corresponds one quantum field. The quantum fields are in turn described in terms of a Lagrangian density. Based on the particle content of the SM one therefore expects a Lagrangian density on the form

$$\mathcal{L}_{sm} \sim \mathcal{L}_{fermions} + \mathcal{L}_{vb} + \mathcal{L}_{Higgs}, \quad (12.4)$$

where vb abbreviates "vector bosons". Since one field is needed to describe each type of particle it is expected that three types of fields are needed; scalar fields, Dirac (spin $\frac{1}{2}$) fields and vector fields (spin 1). This is however not completely correct; *four* types of fields are actually needed, since the vector fields come in two forms: Abelian and non-abelian. Further more, it is expected that, since the particles can interact, the Lagrangian density will contain all sorts of interaction terms. These interaction terms can be deduced after introducing the free fields (scalar, Dirac and abelian gauge fields) and the non-abelian gauge fields. The free field theories are defined as the ones which are given in terms of classical Lagrangian density that can be quantized via *canonical quantization*.

12.1 Canonical Quantization

Canonical quantization is a method for turning a non-interacting, classical field into a free quantum field, and as such it defines what is meant by a "free quantum field". The free field theories considered are those relevant for the standard model of elementary particle physics (SM); the free scalar field, the free Dirac field and the free photon field. So, the first questions one might ask is; what is the goal of quantization? When is the quantization process completed? Since free field theories are considered, the particles they represent cannot interact and all they do is propagate. Therefore, all the information from the free field theory is contained in the probability amplitude for a particle to be created at one point and annihilated at another point - this probability amplitude is defined as the propagator. In terms of fields it is given by

$$G \equiv \langle 0 | T\{\text{fields}\} | 0 \rangle, \quad (12.5)$$

where G denotes that the propagator is a two-point Greens function and $T\{\dots\}$ denotes the act of time ordering, i.e. the ordering of the fields with the "latest" time to the left. Since the propagator depends on the fields, there is a different propagator to each free field.

The free, real scalar field

The first free field considered is the free, real scalar field¹. The real scalar field is the simplest field and so it works well to illustrate the procedure of canonical quantization. The procedure follows five steps:

1. Write up the classical Lagrangian density in terms of the field in question:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2. \quad (12.6)$$

From the Lagrangian density find the equation of motion (EOM) via using the Euler-Lagrange equation

$$\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} = 0 \Rightarrow (\square + m^2)\phi = 0, \quad (12.7)$$

where $\square = \partial_\mu\partial^\mu$. This particular EOM is called the Klein-Gordon (KG) equation.

2. The solution to the KG equation is on the form of a plane wave, i.e. $\phi(x) \sim e^{\pm ip \cdot x}$. After quantization is complete the solution with a plus sign represents a particle with negative energy, i.e. an anti-particle. However, for the real scalar field the particle is its own anti-particle, so this aspect does not come into play here - it will however do so for the Dirac field. Because the solution is on the form of a plane wave, the solution (field) can be expanded in what is called a plane wave expansion cf.

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \left(a_{\vec{p}} e^{-ip \cdot x} + a_{\vec{p}}^* e^{ip \cdot x} \right), \quad (12.8)$$

where the factor of $\frac{1}{\sqrt{2E_{\vec{p}}}}$ is there to ensure Lorentz invariance of the integration measure and $E_{\vec{p}}^2 = \vec{p}^2 + m^2$. The field in equation (12.8) is still a classical field. Note that the field is a function of space-time and when quantized it will therefore be represented in the Heisenberg scheme - since the field is the dynamical variable in QFT, taking the place of position in QM, and time-dependent dynamical variables belong in the Heisenberg scheme.

3. Calculate the momentum conjugate to the field via

$$\begin{aligned} \Pi(x) &\equiv \frac{\partial\mathcal{L}}{\partial(\partial_0\phi)} \\ &= \partial_0\phi \\ &= \int \frac{d^3p}{(2\pi)^3} \frac{(-iE_{\vec{p}})}{\sqrt{2E_{\vec{p}}}} (a_{\vec{p}} e^{ip \cdot x} - a_{\vec{p}}^* e^{-ip \cdot x}). \end{aligned} \quad (12.9)$$

4. Promote the classical fields (ϕ and Π) to operator valued quantum fields. This is done by promoting the expansion coefficients ($a_{\vec{p}}$ and $a_{\vec{p}}^*$) to ladder operators. Subsequently to the promotion, impose the canonical commutator relations

$$[\phi(\vec{x}, t), \Pi(\vec{y}, t)] = i\delta^{(3)}(\vec{x} - \vec{y}), \quad [\phi(\vec{x}, t), \phi(\vec{y}, t)] = [\Pi(\vec{x}, t), \Pi(\vec{y}, t)] = 0. \quad (12.10)$$

These commutator relations are called equal time commutator relations. The commutator relations will impose commutator relations on the coefficients. If these can be imposed without violating any

¹The field is specified to be real because the free, complex scalar field is another field.

physical principles, then the theory is quantized. In the case of the real scalar field the commutator relations of the fields imposes on the coefficients

$$[a_{\vec{p}}, a_{\vec{q}}^\dagger] = (2\pi)^3 \delta^{(3)}(\vec{p} - \vec{q}), \quad [a_{\vec{p}}, a_{\vec{q}}] = [a_{\vec{p}}^\dagger, a_{\vec{q}}^\dagger] = 0. \quad (12.11)$$

These commutator relations are the relations of the ladder operators. The quantized field is to be understood in terms of its effect on the vacuum state

$$\phi(x) |0\rangle = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} e^{-ip \cdot x} |\vec{p}\rangle \quad (12.12)$$

= Linear superposition of single particle states with well-defined momentum..

The above is interpreted as follows $\phi(x)$ acting on the vacuum creates a particle at space-time point x . In general, $\phi(x)$ acting on a generic state, can create and/or annihilate particles at a given space-time position.

5. Determine the propagator of the free theory. The propagator of the free theory can be found by using the definition

$$D_F(x - y) = \langle 0 | T\{\phi(x)\phi(y)\} | 0 \rangle, \quad (12.13)$$

where $D_F(x - y)$ is used instead of G to denote that it is the propagator for the real scalar field. The propagator is both defined in real space and momentum space. The real space representation is found by insertion of the field into the definition, i.e.

$$\begin{aligned} D_F(x - y) &= \langle 0 | T\{\phi(x)\phi(y)\} | 0 \rangle \\ &= \langle 0 | \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \left(a_{\vec{p}} e^{-ip \cdot x} + a_{\vec{p}}^\dagger e^{ip \cdot x} \right) \int \frac{d^3 q}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{q}}}} \left(a_{\vec{q}} e^{-iq \cdot y} + a_{\vec{q}}^\dagger e^{iq \cdot y} \right) | 0 \rangle. \end{aligned} \quad (12.14)$$

The above will have four different terms with ladder operators

$$\langle 0 | a_{\vec{p}}, a_{\vec{q}} | 0 \rangle = \langle 0 | a_{\vec{p}}^\dagger, a_{\vec{q}} | 0 \rangle = \langle 0 | a_{\vec{p}}^\dagger, a_{\vec{q}}^\dagger | 0 \rangle = 0, \quad (12.15)$$

$$\begin{aligned} \langle 0 | a_{\vec{p}}, a_{\vec{q}}^\dagger | 0 \rangle &= \langle 0 | a_{\vec{q}}^\dagger, a_{\vec{p}} | 0 \rangle + \langle 0 | 0 \rangle (2\pi)^3 \delta^{(3)}(\vec{p} - \vec{q}) \\ &= (2\pi)^3 \delta^{(3)}(\vec{p} - \vec{q}). \end{aligned} \quad (12.16)$$

Hereby

$$\begin{aligned} D_F(x - y) &= \int \frac{d^3 p}{(2\pi)^3} \int \frac{d^3 q}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \frac{1}{\sqrt{2E_{\vec{q}}}} (2\pi)^3 \delta^{(3)}(\vec{p} - \vec{q}) e^{-i(p \cdot x - q \cdot y)} \\ &= \int \frac{d^3 p}{(2\pi)^3} \frac{1}{2E_{\vec{p}}} e^{-i(p \cdot x - q \cdot y)}. \end{aligned} \quad (12.17)$$

This is the propagator of the free, real scalar field in the position representation. Most often, however, the propagator is used in the momentum representation. This can be found by using that the free, real scalar field obeys the Klein-Gordon equation. Consequently the propagator is the Greens function of the Klein-Gordon equation, i.e.

$$(\square + m^2) D_F(x - y) = -i\delta^{(4)}(x - y). \quad (12.18)$$

Equation (12.18) can be used to find the momentum, and position, representation of the propagator by expressing the propagator in terms of the inverse Fourier transform of the momentum representation

$$D_F(x - y) = \int \frac{d^4 p}{(2\pi)^4} e^{-ip \cdot (x-y)} \tilde{D}_F(p). \quad (12.19)$$

By using this

$$\begin{aligned} (\square + m^2)D_F(x - y) &= \int \frac{d^4 p}{(2\pi)^4} (-p^2 + m^2) \tilde{D}_F(p) e^{-ip \cdot (x-y)} \\ &= -i\delta^{(4)}(x - y). \end{aligned} \quad (12.20)$$

Now, since $\int \frac{d^4 p}{(2\pi)^4} e^{-ip \cdot (x-y)} = \delta^{(4)}(x - y)$ the momentum representation of the propagator must be given by

$$\tilde{D}_F(p) = \frac{i}{p^2 - m^2}. \quad (12.21)$$

Using this propagator in the Fourier expansion (equation (12.19)) will result in an expression quite similar to equation (12.17). However, note that there are small differences, among them the dimensionality of the integral.

The Dirac field

The next field subject to canonical quantization is the Dirac field. The five steps are repeated:

1. The classical Lagrangian density in terms of the field(s) in question

$$\mathcal{L} = \bar{\psi}(i\cancel{\partial} - m)\psi, \quad (12.22)$$

where $\cancel{\partial} = \gamma^\mu \partial_\mu$ and γ^μ are 4×4 matrices. From the Lagrangian density the EOM is found via using the Euler-Lagrange equation for $\bar{\psi}$

$$\frac{\partial \mathcal{L}}{\partial \bar{\psi}} - \partial_\alpha \frac{\partial \mathcal{L}}{\partial (\partial_\alpha \bar{\psi})} = 0 \Rightarrow (i\cancel{\partial} - m)\psi = 0. \quad (12.23)$$

This EOM is called the Dirac equation.

2. The solution to the Dirac equation is a somewhat more involved compared to the KG equation. Firstly, as indicated by the gamma-matrices which are 4×4 , the solutions to the Dirac equation are 4×1 objects called spinors, or Dirac spinors. These can be written as

$$\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{bmatrix} = \begin{bmatrix} \psi_L \\ \psi_R \end{bmatrix}, \quad (12.24)$$

where ψ_L and ψ_R are called left-handed and right-handed Weyl spinors. A spinor differs from a vector in the way it transforms. Each component of the Dirac spinor must fulfill the KG equation, since

$$(i\cancel{\partial} + m)(i\cancel{\partial} - m)\psi = -(\square + m^2)\psi = 0. \quad (12.25)$$

Hence, the solution must once again be on the form of a plane wave, i.e. $\psi(x) \sim e^{\pm ip \cdot x}$. However, in the case of the Dirac field, which is not required to be real, the two sets of solutions will result in a particle and an anti-particle. Further more, since the solution must fulfill the Dirac equation, but have the position dependence such that it fulfills the KG equation, the solutions are parametrized cf.

$$\psi(x) = u(\vec{p})e^{-ip \cdot x} = \begin{bmatrix} u_L(\vec{p}) \\ u_R(\vec{p}) \end{bmatrix} e^{-ip \cdot x}, \quad \psi(x) = v(\vec{p})e^{ip \cdot x} = \begin{bmatrix} v_L(\vec{p}) \\ v_R(\vec{p}) \end{bmatrix} e^{ip \cdot x}. \quad (12.26)$$

The functions $u(\vec{p})$ and $v(\vec{p})$ satisfy the Dirac equations in momentum space. These are found by inserting the parametrized solutions into the Dirac equation

$$(\cancel{\partial} - m)u(\vec{p}) = 0, \quad (-\cancel{\partial} - m)v(\vec{p}) = 0. \quad (12.27)$$

To determine the functions $u(\vec{p})$, $v(\vec{p})$ the solution is considered in the reference frame at which the particle is at rest, and thereafter the solution can be boosted to an arbitrary reference frame. In the rest frame; $\vec{p}_{rf}^T = [m \ 0]$ and so

$$\begin{aligned} (\gamma^0 m - m) u(\vec{p}_{rf}) &= \begin{bmatrix} -m & m \\ m & -m \end{bmatrix} \begin{bmatrix} u_L(\vec{p}_{rf}) \\ u_R(\vec{p}_{rf}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Rightarrow u_L(\vec{p}_{rf}) &= u_R(\vec{p}_{rf}) \equiv \xi^s \\ \Rightarrow u(\vec{p}_{rf}) &= \sqrt{m} \begin{bmatrix} \xi^s \\ \xi^s \end{bmatrix}, \end{aligned} \quad (12.28)$$

where \sqrt{m} is included by convention, ξ^s is a two component spinor that is normalized cf. $\xi \xi^\dagger = 1$, and $s = 1, 2$ represents the spin states (along the z -axis) of the particle. Likewise for $v(\vec{p})$

$$\begin{aligned} (-\gamma^0 m - m) v(\vec{p}_{rf}) &= \begin{bmatrix} -m & -m \\ -m & -m \end{bmatrix} \begin{bmatrix} v_L(\vec{p}_{rf}) \\ v_R(\vec{p}_{rf}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Rightarrow v_L(\vec{p}_{rf}) &= -v_R(\vec{p}_{rf}) \equiv \eta^s \\ \Rightarrow v(\vec{p}_{rf}) &= \sqrt{m} \begin{bmatrix} \eta^s \\ -\eta^s \end{bmatrix}. \end{aligned} \quad (12.29)$$

It is clear that the solutions have two free components, the components of ξ and η , this is what is expected since the Dirac field describes spin $\frac{1}{2}$ particles - and spin $\frac{1}{2}$ particles only have two physical states; spin up and down. From the solutions in the rest frame of the particle the solution in a generic reference frame can be found, as mentioned, by boosting to a generic reference frame. Hereby

$$u^s(\vec{p}) = \begin{bmatrix} \sqrt{\vec{p} \cdot \vec{\sigma}} \xi^s \\ \sqrt{\vec{p} \cdot \vec{\sigma}} \xi^s \end{bmatrix}, \quad v^s(\vec{p}) = \begin{bmatrix} \sqrt{\vec{p} \cdot \vec{\sigma}} \eta^s \\ -\sqrt{\vec{p} \cdot \vec{\sigma}} \eta^s \end{bmatrix}, \quad (12.30)$$

where $\sigma^T = [1 \ \vec{\sigma}]$ and $\bar{\sigma}^T = [1 \ -\vec{\sigma}]$. With the solutions to the Dirac equation determined the fields can be expanded in a plane-wave expansion

$$\psi(x) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \sum_s \left(a_{\vec{p},s} u^s(\vec{p}) e^{-ip \cdot x} + b_{\vec{p},s}^\dagger v^s(\vec{p}) e^{ip \cdot x} \right), \quad (12.31)$$

$$\begin{aligned} \bar{\psi}(x) &\equiv \psi^\dagger \gamma^0 \\ &= \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \sum_s \left(b_{\vec{p},s} \bar{v}^s(\vec{p}) e^{-ip \cdot x} + a_{\vec{p},s}^\dagger \bar{u}^s(\vec{p}) e^{ip \cdot x} \right), \end{aligned} \quad (12.32)$$

where the expansion coefficients for v , i.e. b^s , denote what is to become the ladder operators of the anti-particle. Likewise for u , i.e. a^s , are to become the ladder operators of the particle.

3. In the case of the Dirac field there is only one conjugate momentum, namely the momentum conjugate to ψ . This is because the conjugate momentum is proportional, cf. the definition $\Pi_i \equiv \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi_i)}$, to the time-derivative of the field. From the Dirac equation is it clear that only the time-derivative of ψ appears, so there is only one conjugate momentum

$$\Pi = \frac{\partial \mathcal{L}}{\partial(\partial_0 \psi)} = i\bar{\psi} \gamma^0 = i\psi^\dagger. \quad (12.33)$$

4. The classical fields are promoted to operator valued quantum fields by promoting the expansion coefficients to ladder operators. Note, however, that there are now two sets; one for the particle and one for the anti-particle. Subsequently, impose the canonical commutator relations on the fields. However, now that the particles are fermions, cf. the spin statistic theorem, the commutator relations that are to be imposed are *anti-commutator relations*

$$\{\psi_a(\vec{x}, t), \Pi_b(\vec{y}, t)\} = i\{\psi_a(\vec{x}, t), \psi_b^\dagger(\vec{y}, t)\} = i\delta^{(3)}(\vec{x} - \vec{y})\delta_{ab}, \quad (12.34)$$

$$\{\psi_a(\vec{x}, t), \psi_b(\vec{y}, t)\} = \{\psi_a^\dagger(\vec{x}, t), \psi_b^\dagger(\vec{y}, t)\} = 0. \quad (12.35)$$

These commutator relations result in commutator relations for the ladder operators

$$\{a_{\vec{p}}^r, a_{\vec{q}}^{s\dagger}\} = \{b_{\vec{p}}^r, b_{\vec{q}}^{s\dagger}\} = (2\pi)^3 \delta^{(3)}(\vec{p} - \vec{q}) \delta_{rs}, \quad (12.36)$$

where all other anti-commutator-combinations of the ladder operators vanish. Analogous to the interpretation of the scalar field operator; ψ creates an anti-particle or destroys a particle and $\bar{\psi}$ does the opposite.

5. Determine the propagator of the free theory. For the Dirac field

$$S_F(x - y) = \langle 0 | T\{\psi(x)\bar{\psi}(y)\} | 0 \rangle. \quad (12.37)$$

Once again the definitions of the fields can be used to determine the form of the propagator in position space. To determine the momentum representation use that the Dirac propagator is the Greens function of the Dirac equation, i.e.

$$(i\cancel{\partial} - m)S_f(x - y) = i\delta^{(4)}(x - y). \quad (12.38)$$

The same approach as for the scalar field is used; express the field in terms of the inverse Fourier transform of the momentum representation and use it in the equation for the Greens function. The inverse Fourier transform

$$S_F(x - y) = \int \frac{d^4 p}{(2\pi)^3} e^{-ip \cdot (x-y)} \tilde{S}_F(p). \quad (12.39)$$

Inserted into the equation for the Greens function

$$\begin{aligned} (i\cancel{\partial} - m)S_f(x - y) &= \int \frac{d^4 p}{(2\pi)^3} (i\cancel{p} - m) e^{-ip \cdot (x-y)} \tilde{S}_F(p) \\ &= i\delta^{(4)}(x - y). \end{aligned} \quad (12.40)$$

Using once more that $\int \frac{d^4 p}{(2\pi)^4} e^{-ip \cdot (x-y)} = \delta^{(4)}(x - y)$ results in the propagator in momentum space

$$\tilde{S}_F(p) = \frac{i}{\cancel{p} - m} = \frac{i(\cancel{p} + m)}{p^2 - m^2}, \quad (12.41)$$

where the factor of $i\varepsilon$ often is added in the denominator for computational purposes.

The EM field

The canonical quantization of the EM field is a bit tricky because EM is an abelian gauge theory. "Abelian" refers to the fact that the generators of the group commute - in practice it means that the theory is quantized via canonical quantization (i.e. it means that the theory is free). "Gauge theory" refers to the fact that the theory may admit different configurations resulting in the same physical observables. Hence, the theory contains some inherent vagueness and so one can choose one of the many equivalent formulations. The formulations are called gauges, transformations between different formulations are called gauge transformations and the invariance of the physical theory under a gauge transformation is called gauge invariance. The EM theory is gauge invariant and so gives a redundant description. The problem is that only two of the four components of the vector field, A^μ , are physically independent, while the theory allows for all four to be independent. To solve this problem a certain formulation (gauge) is chosen to remove the redundancy. The common choice of gauge is one of two:

1. The Coulomb gauge; $A^0 = 0$ and $\vec{\nabla} \cdot \vec{A} = 0$.

Impose the Coulomb gauge from the start. This will remove the redundancy, but since this gauge differs between A^0 and \vec{A} , explicit Lorentz covariance is lost. This will have to be recovered in the end.

2. Make a "soft" Lorentz gauge choice.

Add a term to the Lagrangian density and quantize this Lagrangian density via canonical quantization. In the end impose the soft Lorentz gauge - The word "soft" is used because the Lorentz gauge, i.e. $\partial_\mu A^\mu = 0$, is not used directly, but only when applied to physical states. In quantizing the EM field in this way the explicit Lorentz covariance is not lost.

The second approach will be employed here, i.e. the soft Lorentz gauge.

1. The classical Lagrangian density in terms of the fields in question

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \underbrace{\frac{1}{2}(\partial_\mu A^\mu)^2}_{\text{Gauge fixing term}}. \quad (12.42)$$

From the Lagrangian density the EOM is found via the Euler-Lagrange equation

$$\frac{\partial \mathcal{L}}{\partial A^\alpha} - \partial_\beta \frac{\partial \mathcal{L}}{\partial (\partial_\beta A^\alpha)} = 0 \Rightarrow \square A^\alpha = 0. \quad (12.43)$$

2. The solution to $\square A^\alpha = 0$ is on the form of a plane wave. Taking the polarization into account, the field can be expanded cf.

$$A^\mu(x) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \sum_{\lambda=0}^3 \left(\varepsilon^\mu(\vec{p}, \lambda) a_{\vec{p}, \lambda} e^{-ip \cdot x} + \varepsilon^{\mu*}(\vec{p}, \lambda) a_{\vec{p}, \lambda}^* e^{ip \cdot x} \right). \quad (12.44)$$

3. The momentum conjugate to the field is calculated via

$$\Pi^\mu(x) = \frac{\partial \mathcal{L}}{\partial (\partial_0 A_\mu)}. \quad (12.45)$$

4. Again the classical fields are promoted to operator valued quantum fields by promoting the expansion coefficients to ladder operators. Subsequently, impose the canonical commutator relations, however, multiply with the metric to conserve the indices

$$[A^\mu(\vec{x}, t), \Pi^\nu(\vec{y}, t)] = ig^{\mu\nu} \delta^{(3)}(\vec{x} - \vec{y}), \quad [A^\mu(\vec{x}, t), A^\nu(\vec{y}, t)] = [\Pi^\mu(\vec{x}, t), \Pi^\nu(\vec{y}, t)] = 0. \quad (12.46)$$

These commutator relations impose commutator relations upon the ladder operators

$$[a_{\vec{p}, \lambda}, a_{\vec{q}, \lambda'}^\dagger] = -(2\pi)^3 \delta^{(3)}(\vec{p} - \vec{q}), \quad (12.47)$$

$$[a_{\vec{p}, \lambda}, a_{\vec{q}, \lambda'}] = [a_{\vec{p}, \lambda}^\dagger, a_{\vec{q}, \lambda'}^\dagger] = 0. \quad (12.48)$$

This is where the previous quantization procedures would move on the the propagator. In this case there remains work to be done. The metric's appearance in the commutator poses a problem since it results in a negative norm of the one-particle states ($| \vec{p}, t \rangle = \sqrt{2E_{\vec{p}}} a_{\vec{p}, \lambda}^\dagger | 0 \rangle$)

$$\langle \vec{p}, \lambda | \vec{p}, \lambda \rangle = 2E_{\vec{p}} \langle 0 | a_{\vec{p}, \lambda} a_{\vec{p}, \lambda}^\dagger | 0 \rangle = 2E_{\vec{p}} \langle 0 | [a_{\vec{p}, \lambda}, a_{\vec{p}, \lambda}^\dagger] | 0 \rangle = -2(2\pi)^3 \delta^{(3)}(0) < 0. \quad (12.49)$$

Because of the negative norm of the states the normal interpretation of the norm, i.e. as a probability density, is not possible. However, the soft Lorentz gauge is yet to be imposed, and this fixes the issue. So far the theory that has been considered is not that of EM, but a theory different from EM. This is because a term has been added to the EM Lagrangian density, to fix the gauge, and this term has not been constrained to vanish. The idea is now to pick out only the part of the Fock space that are of physical meaning to EM. This is done through requiring the soft Lorentz gauge

$$\langle \text{phys}' | \partial_\mu A^\mu | \text{phys} \rangle = 0. \quad (12.50)$$

So, rather than requiring $\partial_\mu A^\mu = 0$ in the Lagrangian density, the Lorentz gauge is imposed as an operator equation on physical states. To ensure the above, split $\partial_\mu A^\mu$ into its positive and negative energy parts

$$\partial_\mu A^\mu = (\partial_\mu A^\mu)^+ + (\partial_\mu A^\mu)^-, \quad (12.51)$$

where

$$(\partial_\mu A^\mu)^+ = -i \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \sum_{\lambda=0}^{\lambda} p_\mu \varepsilon^\mu(\vec{p}, \lambda) a_{\vec{p}, \lambda} e^{-ip \cdot x}, \quad (12.52)$$

$$(\partial_\mu A^\mu)^- = i \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \sum_{\lambda=0}^{\lambda} p_\mu \varepsilon^{\mu*}(\vec{p}, \lambda) a_{\vec{p}, \lambda}^\dagger e^{ip \cdot x}. \quad (12.53)$$

From the definitions above it is clear that $(\partial_\mu A^\mu)^- = ((\partial_\mu A^\mu)^+)^\dagger$. Therefore, $\langle \dots | \partial_\mu A^\mu | \dots \rangle = 0$ is fulfilled if only

$$(\partial_\mu A^\mu)^+ | \text{phys} \rangle = 0. \quad (12.54)$$

This equation is taken to define the physical subspace of the Fock space. Hereby the EM field has been quantized.

5. Determine the propagator of the free theory. Once more use that the propagator is the Greens function of the EOM, i.e.

$$\square D_F^{\mu\nu}(x - y) = ig^{\mu\nu} \delta^{(4)}(x - y). \quad (12.55)$$

The propagator is then expressed in terms of the inverse Fourier transform of the momentum representation

$$D_F^{\mu\nu}(x - y) = \int \frac{d^4 p}{(2\pi)^4} e^{-ip \cdot (x-y)} \tilde{D}_F^{\mu\nu}(p). \quad (12.56)$$

By using this in the equation for the Greens function

$$\begin{aligned} D_F^{\mu\nu}(x - y) &= \int \frac{d^4 p}{(2\pi)^4} (-p^2) \tilde{D}_F^{\mu\nu}(p) e^{-ip \cdot (x-y)} \\ &= ig^{\mu\nu} \delta^{(4)}(x - y). \end{aligned} \quad (12.57)$$

From the above the propagator in momentum space is found (in the Lorentz gauge) to be

$$\tilde{D}_F^{\mu\nu}(p) = \frac{-ig^{\mu\nu}}{p^2}. \quad (12.58)$$

12.2 Path Integral Quantization

Path integral quantization is a method for determining the propagator of a free theory - However, the method can also be used to quantize non-free (non-abelian gauge theories) theories. Path integral quantization therefore do the same job as canonical quantization. The two methods complement each other and each method have relative strengths and weaknesses; canonical quantization has a very clear physical interpretation whereas path integral quantization is applicable to a wider range of theories. Path integral quantization also has a clear analogy with statistical mechanics, and the theory is in principle non-perturbative², which means that non-perturbative effects can be determined via this method. Since path integral quantization is an alternative to canonical quantization, the goal of the quantization procedure is the same; the propagator (equation (12.5)). The propagator is obtained by generalizing the path integral scheme from QM (section 6.4). In the QM path integral scheme all possible trajectories, for a single particle traveling between two space-time points, are summed over. Going to QFT the functional integral now sums over all possible field

²As opposed to perturbative theories.

configurations that exist between two space-time points. Besides this, the coordinates are replaced by the fields and the Lagrangian for the Lagrangian density. For the real, free scalar field

$$q \Rightarrow \phi, \quad \int Dq \Rightarrow \int D\phi, \quad L \Rightarrow \int d^3x \mathcal{L}. \quad (12.59)$$

Hereby, by analogy

$$\langle \phi_f(\vec{x}), t_f | T\{\phi(x_1)\phi(x_2) \dots\} | \phi_i(\vec{x}), t_i \rangle = \int D\phi \phi(x_1)\phi(x_2) \dots e^{iS}. \quad (12.60)$$

To find the n-point functions, let the states go to the vacuum state and $t_{f,i} \Rightarrow \pm\infty$. Further more, since the contributions from the bubbling of the vacuum is not of interest (in diagrams these are called bubble diagrams) the above expression is divided by a normalizing factor. The bubbling of the vacuum is contributions with no external particles. By normalizing

$$\langle \Omega | T\{O(\phi)\} | \Omega \rangle = \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{\int D\phi O(\phi) e^{iS}}{\int D\phi e^{iS}} \right), \quad (12.61)$$

where $|\Omega\rangle$ is the vacuum state of the full (possibly interacting) theory, $O(\phi)$ denotes some number of scalar fields, the fields on the LHS are operators valued fields and the fields on the RHS are classical fields. The limit is taken over a complex time, T . This is just a mathematical trick. I will get back to this when I do perturbation theory. The generalization is just as straightforward for the Dirac field and the EM field, the only difference being the different fields

$$\langle \Omega | T\{O(\psi)O(\bar{\psi})\} | \Omega \rangle = \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{\int D\psi \int D\bar{\psi} O(\psi)O(\bar{\psi}) e^{iS}}{\int D\phi e^{iS}} \right), \quad (12.62)$$

$$\langle \Omega | T\{O(A)\} | \Omega \rangle = \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{\int DA O(A) e^{iS}}{\int D\phi e^{iS}} \right), \quad (12.63)$$

where $DA = DA^0 DA^1 DA^2 DA^3$ and $\psi, \bar{\psi}$ on the RHS are classical, anti-commuting fields known as Grassmann fields - these will be introduced just below in the subsection on quantizing the Dirac field via the path integral formalism. The number of fields in the above equations can be taken to be two, and the theory taken to be free, in order to obtain the propagators of the free theories. However, the expressions for the propagators are still complicated and unlike the expressions for the propagators in the momentum space that was found in canonical quantization. These expressions can be found by evaluating the path integrals, however, a more *neat* procedure to determine n-point functions, valid for free/non-free theories, is obtained by introducing the *generating functional*. The generating functional is the analogous object to the partition function form statistical mechanics and the n-point functions are found by differentiating the partition function. The partition function is defined as the amplitude for a vacuum to vacuum transition in the presence of source fields. The number of source fields will match the number of fields, so in scalar field theory there is a single source field, in Dirac theory there are two source fields and in the EM theory there are four source fields.

The real, free scalar field

For the real, free scalar field the generating functional is defined as (cf. the above definition)

$$Z[J] = \frac{z[J]}{z[0]} = \frac{1}{z[0]} \int D\phi e^{i \int d^4x (\mathcal{L} + J\phi)}, \quad (12.64)$$

where J is the source field for the scalar field theory and the Lagrangian density for a real, free scalar field is given by equation (12.6). The n-point functions for the real, free scalar field theory is found from the generating functional by using

$$\langle 0 | T\{O(\phi)\} | 0 \rangle = \frac{1}{i} \frac{\delta}{\delta J(x_1)} \frac{1}{i} \frac{\delta}{\delta J(x_2)} \dots \left(Z[J] \right) \Big|_{J=0}. \quad (12.65)$$

Here I will consider the propagator of the free theory, and so I will consider the two point function. To obtain a useful expression for the propagator the procedure is to use a method called "completing the square". In this method $z[J]$ is rewritten so as to factorize $z[0]$ and leave a relatively simple expression, containing the propagator, for $Z[J]$. First, I define the generating functional in terms of the free scalar field

$$Z[J] = \frac{1}{z[0]} \int D\phi e^{i \int d^4x (\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + J\phi)} \equiv \frac{1}{z[0]} \int D\phi e^I. \quad (12.66)$$

The integral defined as I is what must be massaged so $z[0]$ can be factorized

$$\begin{aligned} I &= i \int d^4x \left(\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + J\phi \right) \\ &= i \int d^4x \left(-\frac{1}{2}\phi(\square - m^2)\phi + J\phi \right), \end{aligned} \quad (12.67)$$

where partial integration has been used. To introduce perform the factorization and introduce the propagator make a linear shift in the field cf.

$$\phi(x) = \phi'(x) + i \int d^4y D_F(x-y)J(y). \quad (12.68)$$

By shifting the fields

$$\begin{aligned} I &= i \int d^4x \left(-\frac{1}{2} \left(\phi' + i \int d^4y D_F(x-y)J(y) \right) (\square - m^2) \left(\phi' + i \int d^4z D_F(x-z)J(z) \right) \right. \\ &\quad \left. + J \left(\phi' + i \int d^4y D_F(x-y)J(y) \right) \right) \\ &= i \int d^4x \left(-\frac{1}{2}\phi'(\square - m^2)\phi' + J\phi' + i \int d^4y J(x)D_F(x-y)J(y) \right. \\ &\quad \left. - \frac{1}{2}i \int d^4y D_F(x-y)J(y)(\square - m^2)\phi' - \frac{1}{2}\phi'(\square - m^2)i \int d^4z D_F(x-z)J(z) \right. \\ &\quad \left. - \frac{1}{2}i \int d^4y D_F(x-y)J(y)(\square - m^2)i \int d^4z D_F(x-z)J(z) \right). \end{aligned} \quad (12.69)$$

To calculate this, use that

$$(\square + m^2)D_F(x-y) = -i\delta^{(4)}(x-y). \quad (12.70)$$

Hereby

$$\begin{aligned} I_4 &\equiv -\frac{i}{2} \int d^4y D_F(x-y)J(y)(\square - m^2)\phi' \\ &= -\frac{1}{2}\phi'(x)J(x), \end{aligned} \quad (12.71)$$

$$\begin{aligned} I_5 &\equiv -\frac{i}{2}\phi' \int d^4z (\square - m^2)D_F(x-z)J(z) \\ &= -\frac{1}{2}\phi'(x)J(x), \end{aligned} \quad (12.72)$$

$$\begin{aligned} I_6 &\equiv \frac{1}{2} \int d^4y D_F(x-y)J(y)(\square - m^2) \int d^4z D_F(x-z)J(z) \\ &= \frac{1}{2} \int d^4y \int d^4z D_F(x-y)J(y)(\square - m^2)z D_F(x-z)J(z) \\ &= -\frac{i}{2} \int d^4y \int d^4z J(x)D_F(x-y)J(y), \end{aligned} \quad (12.73)$$

where partial integration has been used twice to evaluate I_4 . I_4 and I_5 cancels I_2 and I_6 eats half of I_3 .
Hereby

$$I = i \int d^4x \left(-\frac{1}{2}\phi'(\square - m^2)\phi' + \frac{i}{2} \int d^4y J(x) D_F(x-y) J(y) \right), \quad (12.74)$$

Since ϕ' is just a shift w.r.t. ϕ the Jacobian is unity and the generating functional is written as

$$\begin{aligned} Z[J] &= \frac{1}{z[0]} \int D\phi e^{-\frac{i}{2} \int d^4x \phi'(\square - m^2)\phi'} e^{-\frac{1}{2} \int d^4x \int d^4y J(x) D_F(x-y) J(y)} \\ &= e^{-\frac{1}{2} \int d^4x \int d^4y J(x) D_F(x-y) J(y)}. \end{aligned} \quad (12.75)$$

Now that the generating functional has been reformulated in a simpler form containing the propagator the propagator of the free field theory can be determined

$$\begin{aligned} \langle 0 | T\{\phi(x_1)\phi(x_2)\} | 0 \rangle &= \frac{1}{i} \frac{\delta}{\delta J(x_1)} \frac{1}{i} \frac{\delta}{\delta J(x_2)} \left(Z[J] \right) \Big|_{J=0} \\ &= -\frac{\delta}{\delta J(x_1)} \left(\left(-\frac{1}{2} \int d^4y D_F(x-y) J(y) - \frac{1}{2} \int d^4x J(x) D_F(x-y) \right) Z[J] \right) \Big|_{J=0} \\ &= D(x_2 - x_1), \end{aligned} \quad (12.76)$$

where it has been used that the derivative of $Z[J]$ contains J to linear order, and such terms vanish when $J = 0$ after all differentiations are completed. The propagator can be found in the momentum representation by using the Greens function approach presented in the section about canonical quantization. The above might seem a bit superficial, since I defined the derivative to be the propagator, and that is exactly what I found. The power of this formalism comes from the fact that the same procedure can be used to determine any n-point function and express it in terms of propagators. Further more, the formalism can accommodate non-free theories with minor adjustments.

The Dirac field

To determine the free field propagator for the Dirac field the generating functional for the Dirac field must be formulated. This is however not completely straightforward. The reason for this (not being straightforward that is) is because fermion fields do not commute, in particular³

$$\psi(x)\psi(y) = -\psi(y)\psi(x). \quad (12.77)$$

That is the fermion fields anti-commute! Since the fields in the path integral formalism are classical fields (on the RHS), a new type of classical, anti-commuting fields must be introduced - This new class of fields are called Grassmann fields and it is built upon the new class of numbers called Grassmann numbers. Take ξ and η to be two Grassmann numbers (called Grassmannians). These are then defined by the fact that they anti-commute, i.e.

$$\xi\eta = -\eta\xi. \quad (12.78)$$

From the above it is clear that $\eta^2 = \eta\eta = -\eta\eta = 0$ - The square of a Grassmann number is zero! Because $\eta^2 = 0$ the Taylor expansion of an arbitrary function, in terms of Grassmann numbers, is very simple since it terminates at $\mathcal{O}(x)$. For this reason, a generic function of Grassmann numbers can be written as

$$f(\eta) = a + b\eta, \quad (12.79)$$

where a, b are real numbers. Differentiation and integration of Grassmann numbers is defined as follows

$$\frac{\partial}{\partial\eta}(\eta) = 1, \quad \frac{\partial}{\partial\eta}(a) = 0, \quad \frac{\partial}{\partial\eta}(\eta\xi) = \xi, \quad \frac{\partial}{\partial\eta}(\xi\eta) = -\xi, \quad \frac{\partial}{\partial\eta}(f(\eta)) = b, \quad (12.80)$$

³This comes from the fact that $\{\psi_a(\vec{x}, t), \psi_b(\vec{y}, t)\} = \{\psi_a^\dagger(\vec{x}, t), \psi_b^\dagger(\vec{y}, t)\} = 0$.

$$\int d\eta \eta = 1, \quad \int d\eta = 0, \quad \int d\eta f(\eta) = \int d\eta a + \int d\eta b\eta = b. \quad (12.81)$$

With Grassmann numbers introduces the generating functional for the Dirac field can be defined in terms of the Grassmann fields $\psi(x), \bar{\psi}(x), \eta(x), \bar{\eta}(x)$

$$Z[\eta, \bar{\eta}] = \frac{1}{z[0, 0]} \int D\bar{\psi} \int D\psi e^{i \int d^4x (\bar{\psi}(i\partial - m)\psi + \bar{\eta}\psi + \bar{\psi}\eta)} \equiv \frac{1}{z[0, 0]} \int D\bar{\psi} \int D\psi e^I. \quad (12.82)$$

The next step is then to complete the square by rewriting I , and factorizing $z[0, 0]$ from $z[\eta, \bar{\eta}]$, once again. I begin by shifting the two fields cf.

$$\psi(x) = \psi'(x) + i \int d^4y S_F(x-y)\eta(y), \quad \bar{\psi}(x) = \bar{\psi}'(x) - i \int d^4z \bar{\eta}(z)S_F(x-z). \quad (12.83)$$

Hereby

$$\begin{aligned} I &= i \int d^4x \left((\bar{\psi}' - i \int d^4z \bar{\eta}(z)S_F(x-z))(i\partial - m)(\psi' + i \int d^4y S_F(x-y)\eta(y)) \right. \\ &\quad \left. + \bar{\eta}(\psi' + i \int d^4y S_F(x-y)\eta(y)) + (\bar{\psi}' - i \int d^4z \bar{\eta}(z)S_F(x-z))\eta \right) \\ &= i \int d^4x \left(\bar{\psi}'(i\partial - m)\psi' + \bar{\psi}'(i\partial) i \int d^4y S_F(x-y)\eta(y) - i \int d^4z \bar{\eta}(z)S_F(x-z)(i\partial - m)\psi' \right. \\ &\quad \left. - i \int d^4z \bar{\eta}(z)S_F(x-z)(i\partial - m)i \int d^4y S_F(x-y)\eta(y) + \bar{\eta}\psi' + i\bar{\eta} \int d^4y S_F(x-y)\eta(y) \right. \\ &\quad \left. + \bar{\psi}'\eta - i \int d^4z \bar{\eta}(z)S_F(x-z)\eta(x) \right). \end{aligned} \quad (12.84)$$

Now use that

$$(i\partial - m)S_F(x-y) = i\delta^{(4)}(x-y). \quad (12.85)$$

Hereby

$$I_2 = -\bar{\psi}'\eta \Rightarrow \text{Cancels } I_7, \quad (12.86)$$

$$I_4 = i \int d^4z \bar{\eta}(z)S_F(x-z)\eta(x) \Rightarrow \text{Cancels } I_8. \quad (12.87)$$

To evaluate I_3 some information is needed: S_F is not a Grassmann number, $\psi(x) = \sum_i \psi_i \phi_i(x)$. where $\psi(x)$ is a Grassmann field, ψ_i is a Grassmann number, $\psi_i(x)$ is a complex field and ∂ is an operator acting on complex valued fields. Hence

$$S_F(x-y)m\psi'(x) = \psi'(x)mS_F(x-y), \quad (12.88)$$

$$\int d^4z S_F(x-y)\partial_\mu \psi'(x) = - \int d^4z \psi'(x)\partial S_F(x-y). \quad (12.89)$$

Hereby

$$I_3 = -\bar{\eta}\psi' \Rightarrow \text{Cancels } I_5. \quad (12.90)$$

After the cancellations

$$\begin{aligned} Z[\eta, \bar{\eta}] &= \frac{1}{z[0, 0]} \int D\bar{\psi} \int D\psi e^{i \int d^4x \bar{\psi}(i\partial - m)\psi' e^{- \int d^4x \int d^4y \bar{\eta}(x)S_F(x-y)\eta(x)}} \\ &= e^{- \int d^4x \int d^4y \bar{\eta}(x)S_F(x-y)\eta(x)}. \end{aligned} \quad (12.91)$$

The propagator is then obtained via.

$$\begin{aligned} \langle 0 | T\{\psi(x_1)\bar{\psi}(x_2)\} | 0 \rangle &= \frac{1}{i} \frac{\delta}{\delta \bar{\eta}(x_1)} \frac{(-1)}{i} \frac{\delta}{\delta \eta(x_2)} \left(Z[\eta, \bar{\eta}] \right) \Big|_{\eta=\bar{\eta}=0} \\ &= \frac{\delta}{\delta \bar{\eta}(x_1)} \left(\int d^4x \bar{\eta}(x)S_F(x-x_2)Z[\eta, \bar{\eta}] \right) \Big|_{\eta=\bar{\eta}=0} \\ &= S_F(x_1 - x_2). \end{aligned} \quad (12.92)$$

Again, the propagator in momentum representation can be found by using that S_F is the Greens function of the Dirac equation.

The EM field

To determine the free field propagator for the EM field I once again formulate the generating functional. However, since the EM field theory is a gauge theory it incorporates redundant degrees of freedom that will have to be removed by a gauge choice. I will use the soft Lorentz gauge - The gauge I also used for canonical quantization. Using the Lagrangian density of equation (12.42) the generating functional is given by

$$Z[J^\mu] = \frac{1}{z[0]} \int DA e^{i \int d^4x (-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{2\xi}(\partial_\mu A^\mu)^2 + J^\mu A_\mu)} \equiv \frac{1}{z[0]} \int DA e^I. \quad (12.93)$$

I write out I before completing the square

$$\begin{aligned} I &= i \int d^4x \left(-\frac{1}{4}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) - \frac{1}{2\xi} \partial_\mu A^\mu \partial_\nu A^\nu + J^\mu A_\mu \right) \\ &= i \int d^4x \left(-\frac{1}{4}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu - \partial_\nu A_\mu \partial^\mu A^\nu + \partial_\nu A_\mu \partial^\nu A^\mu) - \frac{1}{2\xi} \partial_\mu A^\mu \partial_\nu A^\nu + J^\mu A_\mu \right) \\ &= i \int d^4x \left(-\frac{1}{2}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu) - \frac{1}{2\xi} \partial_\mu A^\mu \partial_\nu A^\nu + J^\mu A_\mu \right) \\ &= i \int d^4x \left(\frac{1}{2}(A_\nu \square A^\nu - A_\nu \partial_\mu \partial^\nu A^\mu) + \frac{1}{2\xi} A^\mu \partial_\mu \partial_\nu A^\nu + J^\mu A_\mu \right) \\ &= i \int d^4x \left(\frac{1}{2}A_\nu(g^{\mu\nu}\square + (\frac{1}{\xi} - 1)\partial^\mu \partial^\nu)A_\mu + J^\mu A_\mu \right) \\ &\equiv i \int d^4x \left(\frac{1}{2}A_\nu f^{\mu\nu}A_\mu + J^\mu A_\mu \right). \end{aligned} \quad (12.94)$$

Now make the shift in the field

$$A_\mu = A'_\mu + i \int d^4y G_{\mu\alpha}(x-y) J^\alpha(y). \quad (12.95)$$

Hereby

$$\begin{aligned} I &= i \int d^4x \left(\frac{1}{2}A'_\nu f^{\mu\nu}A'_\mu + \frac{1}{2}A'_\nu f^{\mu\nu}i \int d^4y G_{\mu\alpha}(x-y) J^\alpha(y) + \frac{1}{2}i \int d^4z G_{\nu\beta}(x-z) J^\beta(z) f^{\mu\nu} A'_\mu \right. \\ &\quad \left. + \frac{1}{2}i \int d^4z G_{\nu\beta}(x-z) J^\beta(z) f^{\nu\mu}i \int d^4y G_{\mu\alpha}(x-y) J^\alpha(y) + J^\mu A'_\mu + J^\mu i \int d^4y G_{\mu\alpha}(x-y) J^\alpha(y) \right). \end{aligned} \quad (12.96)$$

Now use that

$$f^{\mu\nu}G_{\mu\alpha} = i\delta^\nu_\alpha \delta^{(4)}(x-y). \quad (12.97)$$

By using this

$$I_2 = -\frac{1}{2}A'_\nu J^\nu, \quad (12.98)$$

$$I_3 = -\frac{1}{2}A'_\nu J^\nu, \quad (12.99)$$

$$I_4 = -\frac{i}{2} \int d^4z G_{\nu\beta}(x-z) J^\beta(z) J^\nu(x). \quad (12.100)$$

From the above $I_2 + I_3 + I_5 = 0$ and I_4 is subtracted from I_6 . Hereby

$$\begin{aligned} Z[J^\mu] &= \frac{z[J^\mu]}{z[0]} \\ &= \frac{1}{z[0]} \int DA e^{i \int d^4x \frac{1}{2}A'_\nu f^{\mu\nu}A'_\mu} e^{-\frac{1}{2} \int d^4x \int d^4y J^\mu(x) G_{\mu\alpha}(x-y) J^\alpha(y)} \\ &= e^{-\frac{1}{2} \int d^4x \int d^4y J^\mu(x) G_{\mu\alpha}(x-y) J^\alpha(y)}. \end{aligned} \quad (12.101)$$

The propagator is then obtained via.

$$\begin{aligned}\langle 0 | T\{A_\sigma(x_1)A_\rho(x_2)\} | 0 \rangle &= \frac{1}{i} \frac{\delta}{\delta J^\sigma(x_1)} \frac{1}{i} \frac{\delta}{\delta J^\rho(x_2)} \left(Z[J^\mu] \right) \Big|_{J^\mu=0} \\ &= G_{\rho\sigma}(x_2 - x_1).\end{aligned}\quad (12.102)$$

Again, the propagator in momentum representation can be found by using that $G_{\rho\sigma}(x - y)$ is the Greens function of $\square A^\mu = 0$.

12.3 Yang-Mills theory

After having quantized the free field theories the time has come for the non-free field theories, in particular non-abelian gauge theories which are used to describe all the gauge bosons beside the photon. The non-abelian gauge theories are described by what is called Yang-mills theory. Yang-Mills theory is a gauge theory based in the $SU(n)$ group. The purpose of Yang-Mills theory is to construct a Lagrangian density from the requirement of a local $SU(n)$ gauge invariance and subsequently quantize this Lagrangian density via path integral quantization. Yang-Mills theory is used to describe both QCD ($SU(3)$) and the electroweak theory ($SU(2) \otimes U(1)$), so it is the foundation of the standard model of elementary particle physics. Yang-Mills theory is derived from the concept of gauge invariance. This concept will be introduced by reviewing how to derive the Lagrangian density of QED from requiring gauge invariance. Hereafter the same principle will be used to derive the Yang-Mills Lagrangian density. Subsequently the EM lagrangian density will be quantized via path integral quantization⁴ and this method will be generalized to the Yang-Mills Lagrangian density.

EM Lagrangian density from gauge invariance

The Lagrangian density of EM can be constructed by requiring that the theory has a $U(1)$ symmetry. To construct the Lagrangian density a kinetic and mass term of the fermions and a kinetic term for the photon is needed. The kinetic and mass term of the fermions are given by the free theory - albeit with a covariant derivative instead of the ordinary derivative. The covariant derivative is a derivative associated with local gauge invariance. Take some theory with a global symmetry and promote the symmetry to be local. Hereby the kinetic terms of the Lagrangian density will change - Since the theory is not invariant to a local gauge transformation. Then define the covariant derivative to "mop" up the extra terms such that the Lagrangian density becomes invariant under a local gauge transformation - By construction. To construct a derivative invariant under a local $U(1)$ gauge transformation consider the definition of the derivative in some (n^μ) direction

$$n^\mu \partial_\mu \psi = \lim_{\varepsilon \Rightarrow 0} \left(\frac{1}{\varepsilon} (\psi(x + n\varepsilon) - \psi(x)) \right). \quad (12.103)$$

This is the ordinary, "old" derivative in a given (n^μ) direction. To make the derivative covariant it should be invariant under the local $U(1)$ transformation given by

$$\psi(x) \Rightarrow \psi(x) e^{i\alpha(x)}, \quad (12.104)$$

where $\alpha(x)$ is some arbitrary function. The problem is now evident; the local transformation depends on space-time, and $\psi(x + n\varepsilon)$ and $\psi(x)$ are evaluated at different space-time points, so the terms must transform differently. To fix this a unitary operator (U), that makes the resulting term transform like $\psi(x + n\varepsilon)$, is applied to $\psi(x)$. The covariant derivative is then defined in terms of this unitary operator

$$n^\mu D_\mu \psi = \lim_{\varepsilon \Rightarrow 0} \left(\frac{1}{\varepsilon} (\psi(x + n\varepsilon) - U(x + n\varepsilon, x) \psi(x)) \right), \quad (12.105)$$

where $U(x + n\varepsilon, x)$ must transform, in order to make $U(x + n\varepsilon, x) \psi(x)$ transform like $\psi(x + n\varepsilon)$, cf.

$$U(x + n\varepsilon, x) \Rightarrow e^{i\alpha(x+n\varepsilon)} U(x + n\varepsilon, x) \psi(x) e^{-i\alpha(x)}. \quad (12.106)$$

⁴Path integral quantization can be used for non-free fields. Even though the EM field is free, this will be quantized via path integral quantization in order to generalize to non-free fields.

Hereby

$$U(x + n\varepsilon, x)\psi(x) \Rightarrow e^{i\alpha(x+n\varepsilon)}U(x + n\varepsilon, x)\psi(x)e^{-i\alpha(x)}e^{i\alpha(x)}\psi(x) = e^{i\alpha(x+n\varepsilon)}U(x + n\varepsilon, x)\psi(x), \quad (12.107)$$

where it is clear that $U(x + n\varepsilon, x)\psi(x)$ transforms like $\psi(x + n\varepsilon)$. If $U(x + n\varepsilon, x)$ is a continuous function of the two parameters (x and $x + n\varepsilon$) it can be expanded cf.

$$U(x + n\varepsilon, x) = 1 - ie\varepsilon n^\mu A_\mu(x) + \mathcal{O}(e^2), \quad (12.108)$$

where $A_\mu(x)$ is some vector field, called a gauge field, and e is a constant extracted from A_μ . Using the definition of U in the covariant derivative

$$n^\mu D_\mu \psi = \lim_{\varepsilon \rightarrow 0} \left(\frac{1}{\varepsilon} (\psi(x + n\varepsilon) - \psi(x)) + ie\varepsilon n^\mu A_\mu(x)\psi(x) \right) = n^\mu \partial_\mu \psi + ie n^\mu A_\mu \psi \Rightarrow D_\mu = \partial_\mu + ieA_\mu. \quad (12.109)$$

The transformation law for the gauge field can be found by using the expansion of U in the transformation for U

$$1 - ie\varepsilon n^\mu A_\mu \Rightarrow (1 - ie\varepsilon n^\mu A_\mu)(1 - i(\alpha(x + n\varepsilon) - \alpha(x))) = 1 - ie\varepsilon n^\mu (A_\mu - \frac{1}{e}\partial_\mu \alpha(x)), \quad (12.110)$$

where $\alpha(x + n\varepsilon)$ has been Taylor-expanded for the second equal sign. From the above the familiar gauge transformation from EM is obtained

$$A_\mu \Rightarrow A_\mu - \frac{1}{e}\partial_\mu \alpha(x). \quad (12.111)$$

From the above it is clear that the existence of the gauge field depends on the existence of a local $U(1)$ gauge symmetry. That is interesting since it links a mathematical property to a physical interpretation. The mass term for the fermions ($m\bar{\psi}\psi$) is already invariant under a local gauge transformation, so no further action is needed there. The last step is then to determine the kinetic term of the gauge boson. To this end use that the commutator of the covariant derivative is invariant under a local gauge transformation, so this can be used to construct a "legal term" - a term which will turn out to be the kinetic term for the gauge boson. The commutator of the covariant derivative

$$\begin{aligned} [D_\mu, D_\nu]\psi &= [\partial_\mu + ieA_\mu, \partial_\nu + ieA_\nu]\psi \\ &= \underbrace{[\partial_\mu, \partial_\nu]}_{=0}\psi + ie([\partial_\mu, A_\nu] - [\partial_\nu, A_\mu])\psi + e^2 \underbrace{[A_\mu, A_\nu]}_{=0}\psi \\ &= ie(\partial_\mu A_\nu - \partial_\nu A_\mu)\psi. \end{aligned} \quad (12.112)$$

The Lagrangian density can then be constructed by requiring Lorentz covariance, imposing P, T , using operators up to mass dimension 4 and using Diracs equation. The result will be the Lagrangian familiar from QED

$$\mathcal{L} = \bar{\psi}(i\cancel{D} - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \quad (12.113)$$

Yang-Mills Lagrangian density

The procedure of the previous section can be used to construct a new Lagrangian density; the Yang-Mills Lagrangian density. The procedure is completely analogous, however, in this case $\psi(x)$ is taken to be on the form

$$\psi(x) = \begin{bmatrix} \psi_1(x) \\ \psi_2(x) \\ \vdots \\ \psi_n(x) \end{bmatrix}. \quad (12.114)$$

These n-plets transform cf.

$$\psi(x) \Rightarrow V(x)\psi(x), \quad (12.115)$$

where $V(x)$ is a $n \times n$ matrix. $V(x)$ can be expanded in infinitesimal form in terms of the generators of the group (t^a)

$$V(x) = 1 + i\alpha^a(x)t^a + \mathcal{O}(\alpha^2). \quad (12.116)$$

Compared to EM, where the generator was 1, the generators here are in general matrices. The covariant derivative is again defined by equation (12.105), however, now U is required to transform cf.

$$U(x + n\varepsilon) \Rightarrow V(x + n\varepsilon)U(x + n\varepsilon, x)V(x). \quad (12.117)$$

To determine the covariant derivative expand $U(x + n\varepsilon, x)$ in terms of generators of the group

$$U(x + n\varepsilon, x) = 1 + ig\varepsilon n^\mu A_\mu^a t^a + \mathcal{O}(\varepsilon^2). \quad (12.118)$$

By using this in the expression for the covariant derivative

$$D_\mu = \partial_\mu - igA_\mu^a t^a, \quad (12.119)$$

where t^a are the generators (matrices) of the group. They obey the commutator relation

$$[t^a, t^b] = if^{abc}t^c, \quad (12.120)$$

where f^{abc} are called the structure constant. The transformation law for A_μ^a

$$A_\mu^a t^a \Rightarrow V(x)(A_\mu^a(x) + \frac{i}{g}\partial_\mu)V^\dagger(x). \quad (12.121)$$

The kinetic term for the gauge boson

$$\begin{aligned} [D_\mu, D_\nu]\psi &= [\partial_\mu - igA_\mu^a t^a, \partial_\nu - igA_\nu^b t^b]\psi \\ &= \underbrace{[\partial_\mu, \partial_\nu]}_{=0}\psi - ig[\partial_\mu, A_\nu^b t^b]\psi - ig[A_\mu^a t^a, \partial_\nu]\psi - g^2[A_\mu^a t^a, A_\nu^b t^b]\psi \\ &= -ig([\partial_\mu, A_\nu^b]t^b + A_\nu^b \underbrace{[\partial_\mu, t^b]}_{=0})\psi - ig(A_\mu^a \underbrace{[t^a, \partial_\nu]}_{=0} + [A_\mu^a, \partial_\nu]t^a)\psi \\ &\quad - g^2(A_\mu^a \underbrace{[t^a, A_\nu^b]}_{=0} + A_\mu^a A_\nu^b \underbrace{[t^a, t^b]}_{=if^{abc}t^c} + \underbrace{[A_\mu^a, A_\nu^b]t^b t^a}_{=0} + A_\nu^b \underbrace{[A_\mu^a, t^b]t^a}_{=0})\psi \\ &= -ig \underbrace{(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c)}_{\equiv F_{\mu\nu}^a} t^a. \end{aligned} \quad (12.122)$$

Hereby, the Yang-Mills Lagrangian density can be written as

$$\mathcal{L} = \bar{\psi}(i\cancel{D} - m)\psi - \frac{1}{4}(F^a)^{\mu\nu}F_{\mu\nu}^a, \quad (12.123)$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c, \quad D_\mu = \partial_\mu - igA_\mu^a t^a. \quad (12.124)$$

Quantization of Yang-Mills theory

With the classical Yang-Mills Lagrangian density derived (equation (12.123)) the time has come to quantize it. By using the definitions of $F_{\mu\nu}^a$ and D_μ in \mathcal{L} the Feynman rules can be read of

$$\mathcal{L} = \mathcal{L}_{free} + \underbrace{gA_\mu^a \bar{\psi} \gamma^\mu t^a \psi}_{\text{From } D_\mu} - \underbrace{gf^{abc}(\partial_\mu A_\nu^a)(A^b)^\mu (A^c)^\nu - \frac{g^2}{2}(f^{eab}A_\mu^a A_\nu^b)(f^{ecd}(A^c)^\mu (A^d)^\nu)}_{\text{From F-term}} \quad (12.125)$$

$$gA_\mu^a\bar{\psi}\gamma^\mu t^a\psi \Rightarrow \text{---} \leftarrow \text{---} \equiv i\mathcal{M} = ig\gamma^\mu t^a$$
(12.126)

$$-g f^{abc} (\partial_\mu A_\nu^a) (A^b)^\mu (A^c)^\nu \Rightarrow a, \mu \rightsquigarrow b, \nu \equiv i\mathcal{M} = -ig f^{abc} (-ik^\nu) g^{\mu\alpha} \quad (12.127)$$

$$-\frac{g^2}{2}(f^{eab}A_\mu^a A_\nu^b)(f^{ecd}(A^c)^\mu(A^d)^\nu) \Rightarrow d, \beta \text{ (wavy line)} c, \alpha \equiv i\mathcal{M} = -ig^2 f^{eab} f^{ecd} g^{\mu\alpha} g^{\nu\beta} \quad (12.128)$$

The fermion field is free, so the propagator is found from canonical quantization or path integral quantization. The complications arise when considering the quantization of the gauge bosons - since these are inherently not free! The quantization procedure of the gauge bosons follow closely the path integral approach to quantizing the EM field. For this reason the path integral quantization of the photon field will be reviewed and generalized to non-abelian gauge theories.

Rigorous Path Integral Quantization of EM Field

Consider an n-point function on the form

$$\langle \Omega | T\{O(A)\} | \Omega \rangle = \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{\int D A O(A) e^{iS}}{\int D\phi e^{iS}} \right). \quad (12.129)$$

If no gauge fixing term is applied the functional integrals will run over the infinity of equivalent configurations and give a divergent result. It is sufficient to consider only the denominator since the same applies to the nominator; what is required is an additional term of the Lagrangian density - a gauge fixing term. To only integrate over the physical degrees of freedom the the generalized Lorentz gauge is used

$$G(A) = \partial_\mu A^\mu + \omega(x), \quad (12.130)$$

where $\omega(x)$ is a scalar function that changes nothing. To remove the unphysical degrees of freedom from the integral $\delta(G(A))$ should be inserted into the functional integral. However, an "excuse" to insert it is needed. Use

$$1 = \int D\alpha \delta(G(A)) \det \left(\frac{\delta G(A)}{\delta \alpha} \right). \quad (12.131)$$

By using the gauge $(A^\alpha)^\mu = A^\mu + \frac{1}{e} \partial^\mu \alpha$ the determinant will be constant, that is $\det\left(\frac{\delta G(A)}{\delta \alpha}\right) = \text{const.}$ By using this and integrating over $\omega(x)$ with a Gaussian weight⁵

$$\int DAe^{iS} = \det\left(\frac{\delta G(A)}{\delta \alpha}\right) \underbrace{\int D\alpha}_{\text{What diverges}} \int D\omega e^{-i \int d^4x \frac{\omega^2}{2\xi}} \int DAe^{iS} \propto \int DAe^{iS - i \int d^4x \frac{(\partial_\mu A^\mu)^2}{2\xi}}. \quad (12.132)$$

The above gives rise to the gauge fixing term in the Lagrangian. It is equivalent to evaluating the original integral with a gauge fixing term added to the Lagrangian density.

⁵The integrating over $\omega(x)$ with a Gaussian weight is an averaging operation over the arbitrary functions $\omega(x)$.

Path Integral Quantization for Non-Abelian Gauge Bosons

In quantizing the non-abelian gauge bosons the procedure is the same, however the gauge is given by

$$(A^\alpha)_\mu^a = A_\mu^a + \frac{1}{g} D_\mu^{adj} \alpha^a. \quad (12.133)$$

This means

$$\det\left(\frac{\delta G(A^\alpha)}{\delta \alpha}\right) = \det\left(\frac{1}{g} \partial^\mu D_\mu^{adj}\right) = \int Dc \int D\bar{c} e^{i \int d^4x \bar{c}(-\partial^\mu D_\mu^{adj})c}, \quad (12.134)$$

where the last equality is an identity valid for Grassmann numbers. The above comes about because the covariant derivative now contains A . For the EM field this was not the case, and so the determinant could be thrown into a normalization constant which cancels between nominator and denominator in equation (12.129). Since the determinant now depend on A , this is no longer the case. The new fields (c, \bar{c}) introduced can be absorbed into the Lagrangian and be interpreted physically. The new fields are called ghosts, or Fadeev-Poppov ghosts. The ghosts are not physical fields, since they are anti-commuting but scalar. This is not physical since a scalar field is spin 0 and anti-commuting fields must, cf. the spin statistics theorem, be spin $\frac{1}{2}$. The ghosts are to be interpreted as negative degrees of freedom which cancel the unphysical degrees of freedom in the gauge theory. The above adds to the Lagrangian density, so that an ordinary quantization can be performed with the Lagrangian density

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu}^a)^2 - \frac{1}{2\xi}(\partial_\mu A^\mu)^2 + \bar{\psi}(iD^\mu - m)\psi + \underbrace{\bar{c}^a(-\partial^\mu D_\mu^{ac})c^c}_{=\mathcal{L}_{ghosts}}. \quad (12.135)$$

Since the ghosts appear in the Lagrangian density they also have their own Feynman rules. These can be found by rewriting the ghost term

$$\mathcal{L}_{ghost} = \bar{c}^a \left(\underbrace{-\partial^2 \delta^{ac}}_{\text{Propagator}} - \underbrace{g \partial^\mu f^{abc} A_\mu^b}_{\text{Interaction vertex}} \right) c^c. \quad (12.136)$$

From this Lagrangian density the Feynman rules can be read of

$$\partial^2 \delta^{ab} \Rightarrow a \dots b \equiv i\mathcal{M} = \frac{i\delta^{ab}}{p^2} \quad (12.137)$$

$$\bar{c} A c \Rightarrow b, \mu \sim \begin{array}{c} \xrightarrow{p} \\ \diagup \quad \diagdown \\ b \end{array} \dots a \equiv i\mathcal{M} = -g f^{abc} p^\mu \quad (12.138)$$

The propagator of the gauge boson is given by⁶

$$\tilde{D}^{\mu\nu}(k) = \frac{-i}{k^2 + i\varepsilon} \left(g^{\mu\nu} - (1 - \xi) \frac{k^{\mu\nu}}{k^2} \right) \delta^{ab}. \quad (12.139)$$

12.4 Construction of SM Lagrangian Density

With all the fields of the SM quantized the Lagrangian density can now be constructed. To do so the different terms in the Lagrangian density of equation (12.4) will be specified. These are constructed form physical arguments regarding the needed ingredients for the SM; the SM is described by the symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y. \quad (12.140)$$

⁶This propagator is the same as the EM propagator (obtained from the generating functional method) multiplied with a delta function.

The $SU(3)_C$ part describes the strong interaction⁷ whereas the $SU(2)_L \otimes U(1)_Y$ part describes the electromagnetic and weak interactions⁸. Here "C" refers to color, "L" to left and "Y" to hypercharge. As illustrated in deriving the covariant derivatives; the number of massless gauge bosons corresponds to the number of generators of the group. The number of generators are in turn, for a generic $SU(n)$ group, given by $n^2 - 1$. Hereby $U(1)_Y$ has one generator and the group has one massless gauge boson associated to it. $SU(2)_L$ has three generators and three massless gauge bosons whereas $SU(3)_C$ has eight generators and eight massless gauge bosons. The eight massless gauge bosons of $SU(3)_C$ are the eight gluons which mediate the strong interaction. One would then be tempted to say that the $U(1)_Y$ gauge boson is the photon and the $SU(2)_L$ bosons are the W^\pm, Z^0 bosons - this would however be wrong since the W^\pm, Z^0 bosons are massive. Therefore, three of the gauge bosons must be given mass in order for the theory to describe the physical reality. This is where the Higgs boson comes in alongside the principle of spontaneous symmetry breaking (SSB). By introducing a scalar field with a non-vanishing vacuum expectation value the $SU(2)_L \otimes U(1)_Y$ symmetry can be broken to give mass to the gauge bosons (and fermions). However, the photons still needs to be massless, so the Higgs field must have a (new) $U(1)$ -symmetry (denoted $U(1)_{EM}$, the symmetry of EM). The Higgs field will mix all the gauge bosons of $SU(2)_L \otimes U(1)_Y$ and out of the mix will come three massive gauge bosons and one massless. The symmetry breaking process can be described by stating that the symmetry of $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ is spontaneously broken down to $SU(3)_C \otimes U(1)_{EM}$, i.e.

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \Rightarrow SU(3)_C \otimes U(1)_{EM}. \quad (12.141)$$

Returning to the Lagrangian density of the SM a bit can now be specified; there should be three kinetic terms representing the gauge bosons of $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$. Further more, as was illustrated in the quantization procedure of non-abelian gauge theories, gauge fixing terms and ghosts terms must be included as well. The Higgs field should also be included, and by interacting with the gauge bosons of $SU(2)_L$ and $U(1)_Y$ it must give three of the gauge bosons mass. Such mass terms come from the covariant derivative of the Higgs field, so

$$\mathcal{L}_{vb} \supset -\frac{1}{4}G_{\mu\nu}^c(G^c)^{\mu\nu} - \frac{1}{4}A_{\mu\nu}^a(A^a)^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \mathcal{L}_{Ghost} + \mathcal{L}_{gf}, \quad (12.142)$$

$$\mathcal{L}_{Higgs} \supset |D_\mu\phi|^2 = |(\partial_\mu - igA_\mu^a\tau^a - \frac{i}{2}g'B_\mu)\phi|^2, \quad (12.143)$$

where "gf" denotes "gauge fixing" and the definition of the covariant derivative comes from requiring invariance under the local gauge transformation

$$\psi \Rightarrow e^{i\alpha^a\tau^a}a^{\frac{i\beta}{2}}\phi, \quad \tau^a = \frac{\sigma^a}{2}. \quad (12.144)$$

Even further; the Higgs field must have a potential with a non-vanishing vacuum expectation value (VEV) which is responsible of the SSB of the electroweak theory. The Higgs potential is the familiar "wine bottle" potential, i.e.

$$\mathcal{L}_{Higgs} \supset V_{Higgs} = \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2. \quad (12.145)$$

The minimum of the potential can be found by parameterizing the potential by

$$\psi = \begin{bmatrix} \phi_a + i\phi_b \\ \phi_c + i\phi_d \end{bmatrix}. \quad (12.146)$$

The minimum is then found from $\frac{\partial V}{\partial \phi_i} = 0 \Rightarrow \phi_i^2 = \frac{\mu^2}{\lambda}$. Although this parametrization is well suited to determine the minimum of the potential, it is not the simplest parametrization. The simplest parametrization is found by expanding the field around the minimum. Since there is a continuum of equivalent vacuum values,

⁷The corresponding theory is called QCD.

⁸The corresponding theory is called the electroweak theory, or Weinberg-Salam theory.

corresponding to the bottom of the wine bottle, different vacuum values can be taken as the center of the expansion. The simplest choice is to parametrize the field by

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v + h(x) \end{bmatrix}, \quad (12.147)$$

where $v = \sqrt{\frac{\mu^2}{\lambda}}$ is the arbitrary, real valued VEV and $h(x)$, for which $\langle h(x) \rangle = 0$, is the fluctuations around the vacuum. Writing out the Higgs potential

$$V_{Higgs} = \frac{1}{4} \mu^2 v^2 - \mu^2 h^2 - \lambda v h^3 - \frac{\lambda}{4} h^4. \quad (12.148)$$

From the Higgs potential it is clear that the Higgs field interacts with itself (cf. the h^3 and h^4 terms). This concludes the Higgs and gauge boson terms of the SM. The only thing that remains is to specify the fermion term; $\mathcal{L}_{fermion}$. Since the W -bosons only couple to left-handed fermions, the fermions need to be arranged in a particular way; the left-handed fermions are assigned to doublets of $SU(2)$ whereas the right-handed particles are assigned to singlets of $SU(2)$. That is

$$\text{Doublets: } E_L = \begin{bmatrix} \nu_e \\ e^- \end{bmatrix}, \quad Q_L = \begin{bmatrix} u \\ d \end{bmatrix}, \quad (12.149)$$

$$\text{Singlets: } e_R^-, u_R, d_R. \quad (12.150)$$

Note here that the color index is suppressed on the quarks. In fact each quark is a color triplet, which in turn is a Dirac Spinor⁹. Therefore, written out

$$Q_L = \begin{bmatrix} u \\ d \end{bmatrix} = \begin{bmatrix} u_r \\ u_b \\ u_g \\ d_r \\ d_b \\ d_g \end{bmatrix} = \begin{bmatrix} u_{r,1} \\ u_{r,2} \\ u_{b,1} \\ u_{b,2} \\ u_{g,1} \\ u_{g,2} \\ d_{r,1} \\ d_{r,2} \\ d_{b,1} \\ d_{b,2} \\ d_{g,1} \\ d_{g,2} \end{bmatrix}. \quad (12.151)$$

The $[u \ d]^T$ doublet enables the quarks (mathematically) to interact via the weak force - which is a part of the electroweak theory which is "2 \times 2" since it is described by $SU(2)_L \otimes U(1)_Y$. The triplets enables the quarks to (again, mathematically) interact via the strong force - since QCD is a Yang-Mills theory built on $SU(3)_C$ it is "3 \times 3". The 1/2-doublet are the two left-handed components of the 4-component Dirac spinor that is the fermionic field used to describe the quarks. For the lepton doublet

$$E_L = \begin{bmatrix} \nu_e \\ e^- \end{bmatrix} = \begin{bmatrix} \nu_{e,1} \\ \nu_{e,2} \\ e_1^- \\ e_2^- \end{bmatrix}. \quad (12.152)$$

The leptonic field does not couple to the strong interaction, so there are fewer components. Similarly, the quark and lepton doublets for the other families can be written out. From these definitions, and the Dirac equation, the kinetic terms of the fermions can be constructed

$$\mathcal{L}_{fermion} \supset \sum_{generations} \bar{E}_L(iD)E_L + \bar{e}_R(iD)e_R + \bar{Q}_L(iD)Q_L + \bar{u}_R(iD)u_R + \bar{d}_R(iD)d_R, \quad (12.153)$$

⁹The Dirac spinor has four components, but can also be represented as a doublet with a left-handed and right-handed component.

where the covariant derivatives differ, depending on which particles couple to which gauge bosons. The general covariant derivative is given by

$$D_\mu = \partial_\mu - ig_s G_\mu^c T_s^c - ig A_\mu^a T^a - ig' Y B_\mu. \quad (12.154)$$

As mentioned in the beginning of this chapter; besides giving mass to the massive gauge bosons, the Higgs also give mass to the fermions. From purely dimensional analysis it is clear that some doublet of $SU(2)$ is needed in the mass terms for fermions, since the left-handed fermions are in a doublet and the right-handed fermions are in a singlet. The doublet of $SU(2)$ is the Higgs field. From the Higgs field the mass terms can be constructed cf.

$$\mathcal{L}_{fermion} \supset \sum_{generations} (-1) \lambda_e \bar{E}_L \phi e_R - \lambda_d \bar{Q}_L \phi d_R - \lambda_u \varepsilon^{ab} \bar{Q}_L^a \phi^{b\dagger} u_R + \text{h.c.}, \quad (12.155)$$

where ε^{ab} is the Levi-Civita tensor and "h.c." abbreviates hermitian conjugate. The ϕ -fields provide terms proportional to v and $h(x)$. Terms proportional to v will, since v is the constant VEV, be the mass terms of the fermions. Terms proportional to $h(x)$ will represent the interaction between the Higgs field and the fermions. It is clear that the terms involving $h(x)$, for both the covariant derivative and the fermion mass terms, will represent couplings between the Higgs field and other fields. It is the terms involving the non-vanishing VEV, i.e. v , that give rise to mass terms, that is terms on the form $\sim A^2, h^2, \dots$. This is the mathematical explanation/illustration of why the non-vanishing VEV is needed to give masses to the particles.

Recap of SM Lagrangian density

The end result is given by

$$\begin{aligned} \mathcal{L}_{sm} &= \mathcal{L}_{vb} + \mathcal{L}_{Higgs} + \mathcal{L}_{fermions} \\ &= -\frac{1}{4} G_{\mu\nu}^c (G^c)^{\mu\nu} - \frac{1}{4} A_{\mu\nu}^a (A^a)^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \mathcal{L}_{Ghost} + \mathcal{L}_{gf} \\ &\quad + |D_\mu \phi|^2 + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \\ &\quad + \sum_{generations} \bar{E}_L (i \not{D}) E_L + \bar{e}_R (i \not{D}) e_R + \bar{Q}_L (i \not{D}) Q_L + \bar{u}_R (i \not{D}) u_R + \bar{d}_R (i \not{D}) d_R \\ &\quad + \sum_{generations} (-1) \lambda_e \bar{E}_L \phi e_R - \lambda_d \bar{Q}_L \phi d_R - \lambda_u \varepsilon^{ab} \bar{Q}_L^a \phi^{b\dagger} u_R + \text{h.c..} \end{aligned} \quad (12.156)$$

13 | Observables of Particle Physics

The Lagrangian of the SM is used to obtain the Feynman rules which in turn are used to determine the observables of particle physics. There are two prominent observables; the scattering cross section (σ) and the decay rate (Γ). The differentials of the two are given by

$$d\Gamma = \frac{S}{2E_{\vec{p}}} d\Phi^{(n)} \langle |\mathcal{M}|^2 \rangle, \quad d\sigma = \frac{S}{4F} d\Phi^{(n)} \langle |\mathcal{M}|^2 \rangle, \quad (13.1)$$

where S is a symmetry factor, $E_{\vec{p}}$ in $d\Gamma$ is the energy of the decaying particle, F is called the flux factor¹, $\langle |\mathcal{M}|^2 \rangle$ is the spin/polarization-averaged, squared invariant matrix element² and $d\Phi^{(n)}$ is the n -body phase space factor given by

$$d\Phi^{(n)} = \prod_{i=1}^n \frac{d^3 p'_i}{(2\pi)^3} \frac{1}{2E_i} (2\pi)^4 \delta^{(4)}(\Sigma_{im} p - \Sigma_{fm} p'), \quad (13.2)$$

where "im" and "fm" denotes "initial momenta" and "final momenta", respectively. Even though the two expressions look very similar, they are in practice, due to the inherent differences in the physical processes, different to calculate; both the phase space factor and invariant matrix element will differ between the two. The matrix element is specific for the physical scenario whereas the phase space factor depends on the chosen reference frame. Usually, one of three reference frames is used;

1. $2 \Rightarrow 2$ scattering:

- (a) Massive scattering: If all four particles are massive the CM rest frame is often used.
- (b) Massless scattering: If two of the particles are massless, eg. often the case in QED with the photon, the CM rest frame of the massive, initial particle is often used.

2. $1 \Rightarrow 2$ decay: In decays the rest frame of the decaying particle is used.

13.1 Massive scattering

For a $2 \Rightarrow 2$ scattering the phase space factor under consideration is given by

$$\int d\Phi^{(2)} = \int \frac{d^3 p'_1}{(2\pi)^3} \frac{1}{2E'_1} \int \frac{d^3 p'_2}{(2\pi)^3} \frac{1}{2E'_2} (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p'_1 - p'_2), \quad (13.3)$$

where p_i denotes initial momenta and p'_i denotes final momenta. Considering the scattering event from the CM rest frame, the respective coordinates can, cf. figure 13.1, be described as

¹Most often $2 \Rightarrow 2$ scattering is considered. In this case $F = E_A E_B |v_A - v_B|$ - where A, B denotes the incoming particles.

²The invariant matrix element depends on the process under consideration. If there are particles with spin or polarization, sums, and possibly "a counting factor", i.e. multiples of $\frac{1}{2}$, must be included.

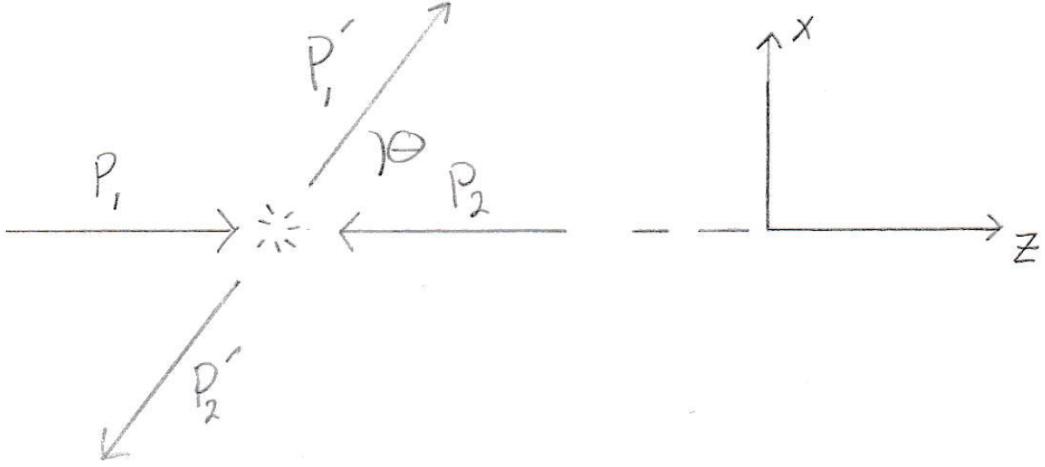


Figure 13.1: The kinematics of massive scattering.

$$p_1 = \begin{bmatrix} E_1 \\ 0 \\ 0 \\ p_{1z} \end{bmatrix}, \quad p_2 = \begin{bmatrix} E_2 \\ 0 \\ 0 \\ p_{2z} \end{bmatrix}, \quad p'_1 = \begin{bmatrix} E'_1 \\ p'_{1x} \\ 0 \\ p'_{1z} \end{bmatrix} \quad \text{and} \quad p'_2 = \begin{bmatrix} E'_2 \\ p'_{2x} \\ 0 \\ p'_{2z} \end{bmatrix}. \quad (13.4)$$

In order to evaluate the three dimensional integrals in the phase space fraction the delta function is split up as follows

$$\delta^{(4)}(p_1 + p_2 - p'_1 - p'_2) = \delta(E_1 + E_2 - E'_1 - E'_2) \delta^{(3)}(\vec{p}_1 + \vec{p}_2 - \vec{p}'_1 - \vec{p}'_2). \quad (13.5)$$

By using the three-dimensional delta-function the integral over $d^3 p'_2$ becomes trivial and it is evaluated by setting $\vec{p}'_2 = \vec{p}_1 + \vec{p}_2 - \vec{p}'_1 = -\vec{p}'_1 \Rightarrow E'_2 = E'_2(p'_1)$. By using this

$$\begin{aligned} \int d\Phi^{(2)} &= \int \frac{d^3 p'_1}{(2\pi)^3 2E'_1} \frac{1}{(2\pi)^3 2E'_2} (2\pi)^4 \delta(E_1 + E_2 - E'_1 - E'_2) \int \frac{d^3 p'_2}{(2\pi)^3 2E'_2} \frac{1}{(2\pi)^3 2E'_2} \delta^{(3)}(\vec{p}_1 + \vec{p}_2 - \vec{p}'_1 - \vec{p}'_2) \\ &= \int \frac{d^3 p'_1}{(2\pi)^3} \frac{(2\pi) \delta(E_1 + E_2 - E'_1 - E'_2)}{4E'_1 E'_2}. \end{aligned} \quad (13.6)$$

Next go to spherical coordinates such that $d^3 p'_1 \Rightarrow |\vec{p}'_1|^2 d|\vec{p}'_1| d\Omega$. Hereby

$$\int d\Phi^{(2)} = \int \frac{d\Omega}{4(2\pi)^2} \int \frac{|\vec{p}'_1|^2 d|\vec{p}'_1|}{E'_1 E'_2} \delta(E_1 + E_2 - E'_1 - E'_2). \quad (13.7)$$

In order to evaluate the integral the delta function is rewritten by using that $\delta(f(x)) = \frac{\delta(x-x_0)}{|\frac{df(x)}{dx}|}$, where x_0 is the zero³ of $f(x)$. By using this identity

$$\delta(E_1 + E_2 - E'_1 - E'_2) = \delta(|\vec{p}'_1| - p'_0) \frac{1}{|\vec{p}'_1| \left(\frac{1}{E'_1} + \frac{1}{E'_2} \right)}, \quad (13.8)$$

where is the value of $|\vec{p}'_1|$ that fulfills $E_1 + E_2 - E'_1(|\vec{p}'_1|) - E'_2(|\vec{p}'_1|) = 0$. This defines the on-shell value of $|\vec{p}'_1|$. By using the delta function

$$\int d\Phi^{(2)} = \int \frac{d\Omega}{16\pi^2} \frac{p'_0}{E'_1 + E'_2}. \quad (13.9)$$

³If there are more than one zero there should be a sum.

By using equation (13.9) and (13.1) the differential cross section can, in the CM rest frame, be written as

$$\frac{d\sigma}{d\Omega} = \frac{S}{4F} \frac{1}{16\pi^2} \frac{p'_0}{E'_1 + E'_2} \langle |\mathcal{M}|^2 \rangle. \quad (13.10)$$

For a $2 \Rightarrow 2$ scattering in the CM rest frame the flux factor can be written as

$$\begin{aligned} F &\equiv E_1 E_2 |\vec{v}_1 - \vec{v}_2| \\ &= \sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} \\ &= \sqrt{E_1^2 E_2^2 + 2E_1 E_2 |\vec{p}_1||\vec{p}_2| + |\vec{p}_1|^2 |\vec{p}_2|^2 - m_1^2 m_2^2} \\ &= |\vec{p}_1| \sqrt{\frac{(|\vec{p}_1|^2 + m_1^2)(|\vec{p}_1|^2 + m_2^2)}{|\vec{p}_1|^2} + 2E_1 E_2 + |\vec{p}_1|^2 - \frac{m_1^2 m_2^2}{|\vec{p}_1|^2}} \\ &= |\vec{p}_1|(E_1 + E_2) \\ &= |\vec{p}_1|(E'_1 + E'_2), \end{aligned} \quad (13.11)$$

where both energy and momentum conservation for the CM rest frame has been used. By using the expression of the flux factor in the differential cross section

$$\frac{d\sigma}{d\Omega} = \frac{S}{64\pi^2} \frac{\langle |\mathcal{M}|^2 \rangle}{(E'_1 + E'_2)^2} \frac{|\vec{p}'_1|}{|\vec{p}_1|}. \quad (13.12)$$

The total cross section is then found by integrating over the solid angle. To this end the invariant amplitude must however be known since it can depend on the solid angle.

13.2 Massless scattering

The phase space factor under consideration is given by equation (13.3). Considering the scattering event from the reference frame in which the massive particle is initially at rest, the respective coordinates can, cf. figure 13.2, be described as

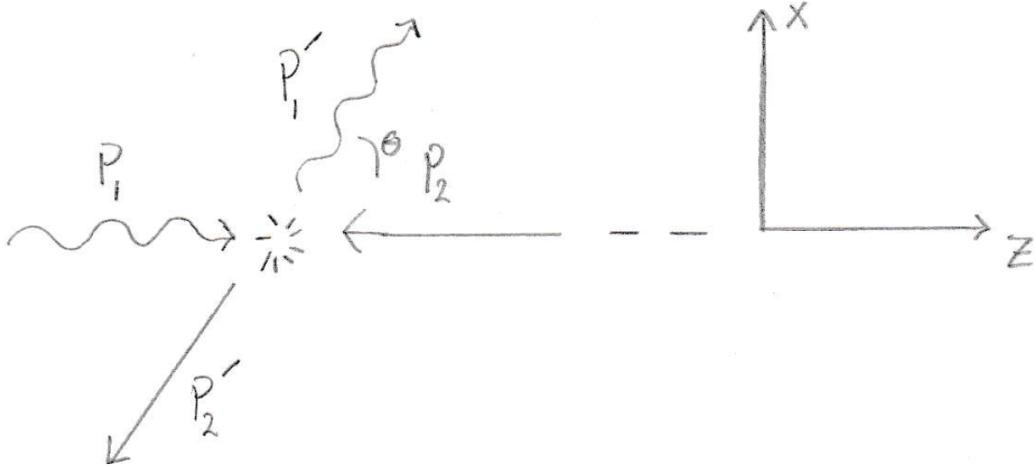


Figure 13.2: The kinematics of massless scattering.

$$p_1 = \begin{bmatrix} E_1 \\ 0 \\ 0 \\ |\vec{p}_1| \end{bmatrix}, \quad p_2 = \begin{bmatrix} E_2 \\ -\vec{p}_1 \end{bmatrix}, \quad p'_1 = \begin{bmatrix} E'_1 \\ |\vec{p}'_1| \sin(\theta) \\ 0 \\ |\vec{p}'_1| \cos(\theta) \end{bmatrix} \quad \text{and} \quad p'_2 = \begin{bmatrix} E'_2 \\ -\vec{p}'_1 \end{bmatrix}, \quad (13.13)$$

where $E_1 = |\vec{p}_1|$ and $E'_1 = |\vec{p}'_1|$. As compared to the massive case, the massless case differs by the kinematics and the flux factor. The flux factor in the massless case is $F = \sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2} = E_1(E_2 + E_1) = E_1 E_{CM} = |\vec{p}_1| E_{CM}$, where $E_{CM} \equiv E_1 + E_2 = E'_1 + E'_2$. By using this in equation (13.10)

$$\begin{aligned}\frac{d\sigma}{d\Omega} &= \frac{S}{4|\vec{p}_1|E_{CM}} \frac{1}{16\pi^2} \frac{|\vec{p}'_1|}{E_{CM}} \langle |\mathcal{M}|^2 \rangle \\ &= \frac{S}{64\pi^2 E_{CM}^2} \frac{|\vec{p}'_1|}{|\vec{p}_1|} \langle |\mathcal{M}|^2 \rangle.\end{aligned}\quad (13.14)$$

13.3 $1 \Rightarrow 2$ decay

The phase space factor for $1 \Rightarrow 2$ decay is given by

$$\int d\Phi^{(2)} = \int \frac{d^3 p'_1}{(2\pi)^3} \int \frac{d^3 p'_2}{(2\pi)^3} \frac{(2\pi)^4 \delta^{(4)}(p - p'_1 - p'_2)}{4E'_1 E'_2}. \quad (13.15)$$

Considering the scattering event from the frame in which the initial particle is at rest, the respective coordinates can, cf. figure 13.3, be described as

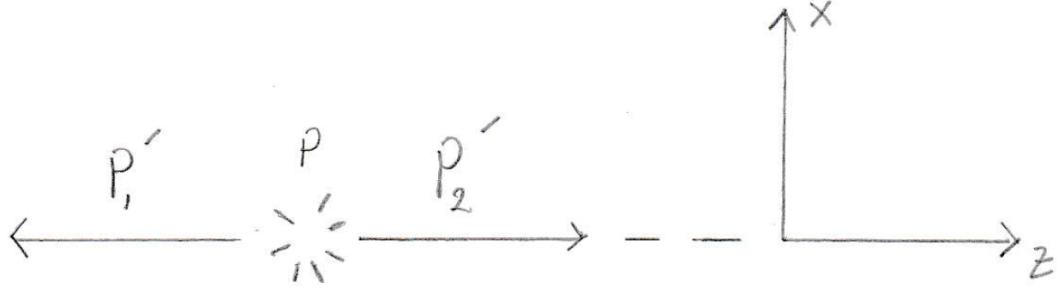


Figure 13.3: The kinematics of the decay.

$$p = \begin{bmatrix} m \\ \vec{0} \end{bmatrix}, \quad p'_1 = \begin{bmatrix} E'_1 \\ \vec{p}'_1 \end{bmatrix} \quad \text{and} \quad p'_2 = \begin{bmatrix} E'_2 \\ \vec{p}'_2 \end{bmatrix}, \quad (13.16)$$

where from conservation of momentum, $k = q + p$

$$\begin{bmatrix} m \\ \vec{0} \end{bmatrix} = \begin{bmatrix} E'_1 + E'_2 \\ \vec{p}'_1 + \vec{p}'_2 \end{bmatrix}. \quad (13.17)$$

Splitting the delta function

$$\delta^{(4)}(p - p'_1 - p'_2) = \delta(m - E'_1 - E'_2) \delta^{(3)}(\vec{p}'_1 + \vec{p}'_2). \quad (13.18)$$

From this the integration over $d^3 p'_2$ can be eliminated, setting $\vec{p}'_1 = -\vec{p}'_2$

$$\begin{aligned}\int d\Phi^{(2)} &= \int \frac{d^3 p'_1}{(2\pi)^3} \frac{(2\pi) \delta(m - E'_1 - E'_2)}{4E'_1 E'_2} \\ &= \int \frac{d\Omega}{16\pi^2} \int \frac{|\vec{p}'_1|^2 d|\vec{p}'_1|}{E'_1 E'_2} \delta(m - E'_1 - E'_2).\end{aligned}\quad (13.19)$$

To evaluate the integral the delta function must be rewritten in terms of the integration variable. Using that, because of the first integral, $E'_1 = E'_1(|\vec{p}'_1|)$, $E'_2 = E'_2(|\vec{p}'_1|)$. By applying the same procedure as in the previous cases

$$\delta(m - E'_1 - E'_2) = \frac{E'_1}{|\vec{p}'_1|} \delta(|\vec{p}'_1| - p_0). \quad (13.20)$$

By using this in equation 13.19

$$\begin{aligned}\int d\Phi^{(2)} &= \int \frac{d\Omega}{16\pi^2} \int \frac{|\vec{p}'_1| d|\vec{p}'_1|}{E'_2} \delta(|\vec{p}'_1| - p_0) \\ &= \int \frac{d\Omega}{16\pi^2} \frac{p_0}{E'_2}.\end{aligned}\tag{13.21}$$

By using equation (13.21) alongside equation (13.1)

$$\frac{d\Gamma}{d\Omega} = \frac{S}{32\pi^2 m} \frac{|\vec{p}'_1|}{E'_2} \langle |\mathcal{M}|^2 \rangle,\tag{13.22}$$

where it once more has been used that p_0 is the on-shell value of $|\vec{p}'_1|$ and so $p_0 = |\vec{p}'_1|$ has been used in the end.

13.4 The Invariant Matrix Element

The invariant matrix element is defined in terms of the transition amplitude between two states. The amplitude is given by

$$\mathcal{A} \equiv \langle f | S | i \rangle,\tag{13.23}$$

where S is the operator, called the scattering operator, evolving the initial state to the final state and

$$\begin{aligned}|i\rangle &= |\vec{p}_1\rangle \otimes |\vec{p}_2\rangle \otimes \cdots = \bigotimes_i |\vec{p}_i\rangle \\ \langle f | &= \cdots \otimes \langle \vec{p}'_2 | \otimes \langle \vec{p}'_1 | = \langle \vec{p}'_j | \bigotimes_j.\end{aligned}\tag{13.24}$$

For future reference I will use p as initial momenta and p' as final momenta. If $S = 1$ there are no scattering between the particles and \mathcal{A} is only non-vanishing if the initial and final state are equal. Motivated by this the S -operator is defined in terms of a non-interacting (I) and interacting (T) part cf.

$$S \equiv 1 + iT.\tag{13.25}$$

Using this definition in \mathcal{A} leads to the definition of the invariant matrix element. \mathcal{A} in terms of T is given by

$$\mathcal{A} = \langle f | i \rangle + \langle f | iT | i \rangle.\tag{13.26}$$

The T -part always carry a delta function securing momentum conservation at the vertexes. This is by convention factored from the last term alongside a factor of $(2\pi)^4$. Hereby

$$\begin{aligned}\mathcal{A} &= \langle f | i \rangle + (2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p) \underbrace{\frac{\langle f | iT | i \rangle}{(2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p)}}_{\equiv i\mathcal{M}} \\ &= \langle f | i \rangle + (2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p) i\mathcal{M}.\end{aligned}\tag{13.27}$$

So, the invariant matrix element is defined from the interactive part of a transition amplitude. The invariant matrix element is most often calculated from Feynman diagrams and Feynman rules. However, it can also be done by a more rigorous procedure - the one used before Feynman diagrams; by relating \mathcal{M} to the n-point functions of the theory. The relationship between the invariant matrix element and the n-point functions depend on the theory, and so in order to illustrate the relationship I have to pick a theory. To keep it simple I will consider an interactive scalar theory, eg. ϕ^4 -theory. The formula I am looking for, relation the invariant matrix element to the n-point functions of the theory, is called the LSZ reduction formula.

The LSZ Reduction Formula

Consider the transition amplitude

$$\mathcal{A} = \underbrace{\langle f | S | i \rangle}_{\text{Schroedinger scheme}} = \underbrace{\langle f, t_f | S | i, t_i \rangle}_{\text{Heisenberg scheme}} . \quad (13.28)$$

Since the goal is to relate the invariant matrix element to the n-point functions (which are on the form $\langle \Omega | T\{ \dots \} | \Omega \rangle$), it is desired to somehow to "remove" the particles from the states and represent them by fields in between the states. Particles can be "pulled" from states by using that $|\vec{p}\rangle = \sqrt{2E_{\vec{p}}} a_{\vec{p}}^\dagger |0\rangle$. The ladder operators are then related to the fields, so intuitively the idea seems reasonable. For a *free* theory the field can be expressed in terms of the mode expansion. This expression can be inverted to isolate the ladder operators to find

$$\sqrt{2E_{\vec{p}}} a_{\vec{p}} = i \int d^3x e^{ip \cdot x} \overleftrightarrow{\partial}_0 \phi_{free}, \quad \sqrt{2E_{\vec{p}}} a_{\vec{p}}^\dagger = -i \int d^3x e^{-ip \cdot x} \overleftrightarrow{\partial}_0 \phi_{free}, \quad (13.29)$$

where

$$e^{ip \cdot x} \overleftrightarrow{\partial}_0 \phi_{free} = e^{ip \cdot x} \partial_0 \phi_{free} - \phi_{free} \partial_0 e^{ip \cdot x}. \quad (13.30)$$

However, the fields under consideration are not free - they are interactive (recall the ϕ^4 -statement). However, interactive fields cannot be expanded in the mode expansion, so it is desired to relate the free fields to the interactive fields. To this end assume that infinitely far away in time, before and after the scattering, the fields are free. This assumption is reasonable for weakly interacting theories, but not in general - for example it is not reasonable for QCD. So, taking the theory under consideration to be weakly interacting

$$\phi \xrightarrow{t \Rightarrow \pm\infty} \sqrt{Z} \phi_{out/in}, \quad (13.31)$$

where Z is the wave function renormalization factor and $\phi_{out/in}$ are free fields infinitely before and after in time. From these definitions

$$\sqrt{2E_{\vec{p}}} a_{\vec{p}}^{in\dagger} = -i Z^{-\frac{1}{2}} \lim_{t \Rightarrow -\infty} \left(\int d^3x e^{-ip \cdot x} \overleftrightarrow{\partial}_0 \phi \right), \quad (13.32)$$

where the field on the RHS is the field of the full theory, i.e. not the free field. From this definition one of the particles of the initial state can be "pulled" from the initial state via.

$$\begin{aligned} \mathcal{A} &= \sqrt{2E_{\vec{p}_1}} \langle \Sigma \vec{p}', t_f | a_{\vec{p}_1}^{in\dagger} | \Sigma \vec{p} - \vec{p}_1, t_i \rangle \\ &= -i Z^{-\frac{1}{2}} \lim_{t \Rightarrow -\infty} \left(\int d^3x e^{-ip_1 \cdot x} \langle \Sigma \vec{p}', t_f | \overleftrightarrow{\partial}_0 \phi | \Sigma \vec{p} - \vec{p}_1, t_i \rangle \right). \end{aligned} \quad (13.33)$$

It is desired to get rid of the limit in the above expression. In order to do this use the identity

$$\left(\lim_{t \Rightarrow \infty} - \lim_{t \Rightarrow -\infty} \right) \int d^3x f(\vec{x}, t) = \int_{-\infty}^{\infty} dt \frac{d}{dt} \left(\int d^3x f(\vec{x}, t) \right). \quad (13.34)$$

To use this identity, the "subtracted limit" is needed. To get this, assume that there are no spectator particles - i.e. no particles that does not interact. If this is the case, then $\langle \Sigma \vec{p}' | a_{\vec{p}}^{out\dagger} = 0$. Since one can always add zero, the transition amplitude can be written as

$$\mathcal{A} = \sqrt{2E_{\vec{p}_1}} \langle \Sigma \vec{p}', t_f | (a_{\vec{p}_1}^{in\dagger} - a_{\vec{p}_1}^{out\dagger}) | \Sigma \vec{p} - \vec{p}_1, t_i \rangle. \quad (13.35)$$

By expressing $\sqrt{2E_{\vec{p}_1}} a_{\vec{p}_1}^{out\dagger}$ in terms of the field the identity can be used. Take $f(\vec{x}, t) \Rightarrow -i Z^{-\frac{1}{2}} e^{-ip_1 \cdot x} \overleftrightarrow{\partial}_0 \phi$. By using the identity, and swapping the time-derivative and integration over d^3x

$$\sqrt{2E_{\vec{p}_1}} (a_{\vec{p}_1}^{in\dagger} - a_{\vec{p}_1}^{out\dagger}) = i Z^{-\frac{1}{2}} \int d^4x \partial_0 (e^{-ip_1 \cdot x} \overleftrightarrow{\partial}_0 \phi), \quad (13.36)$$

where the positive sign comes from the fact that $a_{\vec{p}_1}^{in\dagger} - a_{\vec{p}_1}^{out\dagger} \propto \lim_{t \rightarrow -\infty} - \lim_{t \rightarrow \infty}$. Applying the time derivatives and writing it out

$$\begin{aligned}
\sqrt{2E_{\vec{p}_1}}(a_{\vec{p}_1}^{in\dagger} - a_{\vec{p}_1}^{out\dagger}) &= iZ^{-\frac{1}{2}} \int d^4x \partial_0(e^{-ip_1 \cdot x} \overleftrightarrow{\partial}_0 \phi) \\
&= iZ^{-\frac{1}{2}} \int d^4x \partial_0(e^{-ip_1 \cdot x} \partial_0 \phi + iE_{\vec{p}_1} e^{-ip_1 \cdot x} \phi) \\
&= iZ^{-\frac{1}{2}} \int d^4x (-iE_{\vec{p}_1} e^{-ip_1 \cdot x} \phi + e^{-ip_1 \cdot x} \partial_0^2 \phi + E_{\vec{p}_1}^2 e^{-ip_1 \cdot x} \phi + iE_{\vec{p}_1} e^{-ip_1 \cdot x} \phi) \\
&= iZ^{-\frac{1}{2}} \int d^4x e^{-ip_1 \cdot x} (\partial_0^2 + E_{\vec{p}_1}^2) \phi \\
&= iZ^{-\frac{1}{2}} \int d^4x e^{-ip_1 \cdot x} (\partial_0^2 + m_{\vec{p}_1}^2 + (\vec{p}_1)^2) \phi \\
&= iZ^{-\frac{1}{2}} \int d^4x (e^{-ip_1 \cdot x} \partial_0^2 \phi + \phi(m_{\vec{p}_1}^2 - \nabla^2) e^{-ip_1 \cdot x}) \\
&= iZ^{-\frac{1}{2}} \int d^4x (e^{-ip_1 \cdot x} (\square_x + m_{\vec{p}_1}^2) \phi),
\end{aligned} \tag{13.37}$$

where integration by parts has been used twice to shift ∇^2 from the exponential to ϕ . From using the above the transition amplitude is given by

$$\mathcal{A} = iZ^{-\frac{1}{2}} \int d^4x e^{-ip_1 \cdot x} (\square_x + m_{\vec{p}_1}^2) \langle \Sigma \vec{p}, t_f | \phi(x) | \Sigma \vec{p} - \vec{p}_1, t_i \rangle. \tag{13.38}$$

By continuing this way all particles from the initial and final states can be removed and represented by fields instead. Denoting the "position belonging to the final-state particles" by y and the "position belonging to the initial-state particles" by x the transition amplitude can therefore be written as

$$\begin{aligned}
\mathcal{A} &= \left(i \frac{1}{\sqrt{Z}}\right)^{n+m} \int \prod_{i=1}^m d^4x_i \int \prod_{j=1}^n d^4y_j e^{i(p'_j \cdot y_j - p_i \cdot x_i)} (\square_{x_i} + m_{\vec{p}_i}^2) (\square_{y_j} + m_{\vec{p}'_j}^2) \\
&\quad \cdot \langle \Omega | T\{\phi(x_1) \dots \phi(x_m) \phi(y_1) \dots \phi(y_n)\} | \Omega \rangle.
\end{aligned} \tag{13.39}$$

However, since no spectator particles has been assumed, equation (13.39) only denotes the interactive part of the transition amplitude. Therefore

$$\begin{aligned}
\langle f | iT | i \rangle &= \left(i \frac{1}{\sqrt{Z}}\right)^{n+m} \int \prod_{i=1}^m d^4x_i \int \prod_{j=1}^n d^4y_j e^{i(p'_j \cdot y_j - p_i \cdot x_i)} (\square_{x_i} + m_{\vec{p}_i}^2) (\square_{y_j} + m_{\vec{p}'_j}^2) \\
&\quad \cdot \langle \Omega | T\{\phi(x_1) \dots \phi(x_m) \phi(y_1) \dots \phi(y_n)\} | \Omega \rangle.
\end{aligned} \tag{13.40}$$

Equation (13.40) is a version of the so-called LSZ formula which relates the interactive part of the transition amplitude to the n-point functions of the interactive theory. Similar expressions can be obtained for fermion or boson fields by repeating a similar procedure. The result will be a similar expression with different fields, a possible minus sign (due to anti-commutation of fermion fields) and the Dirac operator instead of the KG operator. Note that $\phi(x)$ is not a free field, but a field of the entire field. This is the reason why $(\square + m^2)\phi \neq 0$, i.e. the KG equation is only valid for free fields.

Perturbation theory in QFT

After having related the n-point functions for the interactive theory to the invariant matrix element it is time to determine the n-point functions themselves. Recall that an interactive theory is considered, so for example the two-point function is not the propagator. This is only the case for a free theory - i.e. a non-interactive theory. What is desired is to relate the n-point functions of the interactive theory to the n-point functions of the free theories, i.e. relate the interactive fields to the free fields and the interactive vacuum state to the

free vacuum state. To this end use what is called the interaction scheme of QM. In this scheme both the states and operators are time-dependent cf.

$$|\psi_I(t)\rangle \equiv e^{iH_0t} |\psi_S(t)\rangle, \quad A_I(t) \equiv e^{iH_0t} A_S e^{-iH_0t}, \quad (13.41)$$

where I denotes the interaction scheme, S denotes the Schrödinger scheme and H_0 is the free part of an interactive Hamiltonian. The interactive theory is characterized by a Lagrangian density on the form

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{int}, \quad (13.42)$$

where $\mathcal{L}_0 = \mathcal{L}_{free}$. Equation (13.42) is true for all theories, not only scalar field theories. The Lagrangian density corresponds to a Hamiltonian density on the form

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{int}. \quad (13.43)$$

The solution to the EOM's for the interactive theory is in general very difficult (at best) to obtain. Since the theory is not free it is however clear that the field will not be given in terms of a simple mode expansions. However, the interactive field can be related to the free field by using the interaction scheme. Since the free field in the mode expansion is represented in the Heisenberg scheme, and $A_H(t) = e^{iHt} A_S e^{-iHt}$, the interactive field can be related to the free field in the interactive scheme via.

$$\phi(x) = U^\dagger(t, t_0) \phi_I(x) U(t, t_0), \quad (13.44)$$

where⁴ $U(t, t_0) = e^{iH_0(t-t_0)} e^{-iH(t-t_0)}$ is the time-evolution operator and $\phi_I(x)$, the field in the interactive scheme, is a free field so it can be written in terms of the mode expansion, i.e.

$$\phi_I(x) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{p}}}} \left(a_{\vec{p}} e^{-ip \cdot x} + a_{\vec{p}}^* e^{ip \cdot x} \right) = \text{free field.} \quad (13.45)$$

The time-evolution operator, U , obeys the Schrödinger equation in the interaction picture. This can be seen by

$$i \frac{\partial U(t, t_0)}{\partial t} = e^{iH_0(t-t_0)} (H - H_0) e^{-H(t-t_0)} = \underbrace{e^{iH_0(t-t_0)} H_{int} e^{-iH_0(t-t_0)}}_{\equiv H_I} U(t, t_0), \quad (13.46)$$

where $H_I = H_I(t)$ is the interaction Hamiltonian in the interaction picture. The Schrödinger equation can be solved to find

$$U(t, t_0) = T\{e^{-i \int d^4x \mathcal{H}_I}\} \Rightarrow S \equiv T\{e^{-i \int_{-\infty}^{\infty} d^4x \mathcal{H}_I}\}. \quad (13.47)$$

With the time-evolution operator in the interaction scheme expressed, and the interactive field expressed in terms of free fields and the time-evolution operator in the interaction scheme, I have related the interactive fields to the free fields - as I wanted. Next up is the relationship between the vacuum state of the interactive theory and the vacuum state of the free theory. In order to relate the two I consider

$$\begin{aligned} e^{-iHt} |0\rangle &= \sum_n e^{-iHt} |n\rangle \langle n|0\rangle \\ &= e^{-iE_0t} |\Omega\rangle \langle \Omega|0\rangle + \sum_{n \neq \Omega} e^{-iHt} |n\rangle \langle n|0\rangle, \end{aligned} \quad (13.48)$$

where $|n\rangle$ are eigenstates of the full theory and $E_0 < e_{n \neq 0}$. To get "rid" of the second term, employ a mathematical trick; let $t \Rightarrow T \Rightarrow \infty(1 - i\varepsilon)$ - where T just denotes complex time. Because $E_0 < e_{n \neq 0}$ the E_0 term will die out last. Therefore, in the limit of $T \Rightarrow \infty(1 - i\varepsilon)$

$$|\Omega\rangle = \lim_{T \Rightarrow \infty(1 - i\varepsilon)} \left(\frac{e^{iHT} |0\rangle}{e^{-iE_0T} \langle \Omega|0\rangle} \right). \quad (13.49)$$

⁴Note that $[H_0, H] \neq 0$ in general so the exponentials cannot be "collected".

Since T is taken in the limit of going to infinity, a small shift, $T \Rightarrow T + t_0$ makes no difference. Hereby

$$|\Omega\rangle = \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{e^{iH(T+t_0)} |0\rangle}{e^{-iE_0(T+t_0)} \langle \Omega | 0 \rangle} \right). \quad (13.50)$$

Use now that, since $H_0 |0\rangle = 0$

$$e^{iH_0(T+t_0)} |0\rangle = (1 - iH_0(T + t_0)) |0\rangle = |0\rangle. \quad (13.51)$$

Hence, the vacuum state in the interactive theory can be written as

$$\begin{aligned} |\Omega\rangle &= \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{e^{iH(T+t_0)} e^{iH_0(T+t_0)} |0\rangle}{e^{-iE_0(T+t_0)} \langle \Omega | 0 \rangle} \right) \\ &= \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{U(t_0, T) |0\rangle}{e^{-iE_0(T+t_0)} \langle \Omega | 0 \rangle} \right). \end{aligned} \quad (13.52)$$

By using equation (13.52) and the expression for ϕ in terms of ϕ_I

$$\langle \Omega | T\{\phi(x_1) \dots \phi(x_N)\} | \Omega \rangle = \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{\langle 0 | U^\dagger(t_0, T) [U(t_0, t) \phi_I(x_1) U(t, t_0)] \dots U(t_0, -T) | 0 \rangle}{e^{-iE_0(2T)} |\langle \Omega | 0 \rangle|^2} \right). \quad (13.53)$$

To get rid of the denominator, I divide by

$$1 = \langle \Omega | \Omega \rangle = \frac{\langle 0 | U(T, t_0) U(t_0, -T) | 0 \rangle}{e^{-iE_0(2T)} |\langle \Omega | 0 \rangle|^2}. \quad (13.54)$$

Next collect the fields in the nominator under a time-ordering. By doing so the time-evolution operators can be collected to the exponential. Hereby the n-point function can be written as

$$\langle \Omega | T\{\phi(x_1) \dots \phi(x_N)\} | \Omega \rangle = \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{\langle 0 | T\{\phi_I(x_1) \dots \phi_I(x_N) e^{-i \int d^4x \mathcal{H}_I}\} | 0 \rangle}{\langle 0 | T\{e^{-i \int d^4x \mathcal{H}_I}\} | 0 \rangle} \right). \quad (13.55)$$

The last step is then to expand the exponential in a Taylor series and use \mathcal{H}_I from the interactive theory. \mathcal{H}_I can be found from

$$\mathcal{H}_I = e^{iH_0(t-t_0)} \mathcal{H}_{int} e^{-iH_0(t-t_0)} = \mathcal{H}_{int}(\phi_I). \quad (13.56)$$

Hence, \mathcal{H}_I is the interactive part of the Hamiltonian density with $\phi \Rightarrow \phi_I$ = free field. Collecting the results of equation (13.27), (13.40) and (13.55)

$$\begin{aligned} i\mathcal{M} &= \frac{\langle f | iT | i \rangle}{(2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p)} \\ &= \frac{(i \frac{1}{\sqrt{Z}})^{n+m}}{(2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p)} \int \prod_{i=1}^m d^4 x_i \int \prod_{j=1}^n d^4 y_j e^{i(p'_j \cdot y_j - p_i \cdot x_i)} (\square_{x_i} + m_{p_i}^2) (\square_{y_j} + m_{p'_j}^2) \\ &\quad \cdot \langle \Omega | T\{\phi(x_1) \dots \phi(x_m) \phi(y_1) \dots \phi(y_n)\} | \Omega \rangle \\ &= \frac{(i \frac{1}{\sqrt{Z}})^{n+m}}{(2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p)} \int \prod_{i=1}^m d^4 x_i \int \prod_{j=1}^n d^4 y_j e^{i(p'_j \cdot y_j - p_i \cdot x_i)} (\square_{x_i} + m_{p_i}^2) (\square_{y_j} + m_{p'_j}^2) \\ &\quad \cdot \lim_{T \rightarrow \infty(1-i\varepsilon)} \left(\frac{\langle 0 | T\{\phi_I(x_1) \dots \phi_I(y_n) e^{-i \int d^4x \mathcal{H}_I}\} | 0 \rangle}{\langle 0 | T\{e^{-i \int d^4x \mathcal{H}_I}\} | 0 \rangle} \right). \end{aligned} \quad (13.57)$$

Equation (13.57) is valid for a generic scalar theory.

Example 13.1.

Consider a scattering amplitude for a process with two initial particles, with momenta \vec{p}_1 and \vec{p}_2 , into two final particles, with momenta \vec{p}'_1 and \vec{p}'_2 , in ϕ^4 -theory. Take the interactive part of the Hamiltonian density to be on the form

$$\mathcal{H}_I = \frac{\lambda}{4!} \phi^4, \quad (13.58)$$

where $\phi = \phi_I$ in this example. The coupling, λ , defines the strength of the interaction.

1. Draw the Feynman diagrams in position space. Compute the amplitude in the case in which there is no interactions, i.e. for $\lambda = 0$, and motivate why there is no contribution as expected.

I begin by writing explicitly the LSZ formula, with the KG operators taken to the LHS and represented in momentum space, for ϕ^4 -theory

$$\left(\prod_{i=1}^2 \frac{i\sqrt{Z}}{p'_i - m^2} \right) \left(\prod_{j=1}^2 \frac{i\sqrt{Z}}{p_j - m^2} \right) \langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = \int d^4 x_1 \int d^4 x_2 \int d^4 x_3 \int d^4 x_4 \\ \cdot e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} \frac{\langle 0 | T \{ \phi(x_1) \phi(x_2) \phi(x_3) \phi(x_4) e^{-i \frac{\lambda}{4!} \int d^4 x \phi^4} \} | 0 \rangle}{\langle 0 | T \{ e^{-i \frac{\lambda}{4!} \int d^4 x \phi^4} \} | 0 \rangle}. \quad (13.59)$$

I take the lowest non-trivial order in perturbation theory - Ie. I will work to $\mathcal{O}(\lambda)$. In this theory

$$Z = 1 + \mathcal{O}(\lambda^2). \quad (13.60)$$

Therefore, I take $Z = 1$. I also expand the exponential cf.

$$e^{-i \frac{\lambda}{4!} \int d^4 x \phi^4} = 1 - i \frac{\lambda}{4!} \int d^4 x \phi^4 + \mathcal{O}(\lambda^2). \quad (13.61)$$

To $\mathcal{O}(\lambda^0)$, with $Z = 1$, I find from the LSZ formula

$$\frac{i}{p'_1 - m^2} \frac{i}{p'_2 - m^2} \frac{i}{p_1 - m^2} \frac{i}{p_2 - m^2} \langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = \int d^4 x_1 \int d^4 x_2 \int d^4 x_3 \int d^4 x_4 \\ \cdot e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} \langle 0 | T \{ \phi(x_1) \phi(x_2) \phi(x_3) \phi(x_4) \} | 0 \rangle. \quad (13.62)$$

The VEV is computed by using Wicks theorem. Wicks theorem states that a time-ordered series of field-operators is related to the normal-ordered series of field-operators, via.

$$T \{ \phi_1 \phi_2 \dots \phi_n \} = N \{ \phi_1 \phi_2 \dots \phi_n + \text{All contractions} \}. \quad (13.63)$$

Wicks theorem is extremely useful when considering a VEV, because the normal ordered product of free field operators vanish⁵. The only terms that does not vanish are the ones which are fully contracted. For the VEV under consideration

$$\begin{aligned} VEV &= \langle 0 | T \{ \phi_1 \phi_2 \phi_3 \phi_4 \} | 0 \rangle \\ &= \langle 0 | N \{ \phi_1 \phi_2 \phi_3 \phi_4 + \overline{\phi_1} \phi_2 \phi_3 \phi_4 + \phi_1 \overline{\phi_2} \phi_3 \phi_4 + \phi_1 \phi_2 \overline{\phi_3} \phi_4 + \phi_1 \phi_2 \phi_3 \overline{\phi_4} \\ &\quad + \phi_1 \overline{\phi_2} \phi_3 \phi_4 + \phi_1 \phi_2 \overline{\phi_3} \phi_4 + \phi_1 \phi_2 \phi_3 \overline{\phi_4} + \phi_1 \phi_2 \phi_3 \phi_4 + \phi_1 \phi_2 \phi_3 \phi_4 \} | 0 \rangle. \end{aligned} \quad (13.64)$$

Now use that the contractions are complex numbers and as such they can be moved outside the Normal-ordering operator and bracket. Hence, what remains, besides the contraction, for fully contracted terms is $\langle 0 | N \{ 1 \} | 0 \rangle = 1$. Hereby

$$\langle 0 | T \{ \phi_1 \phi_2 \phi_3 \phi_4 \} | 0 \rangle = \overline{\phi_1} \phi_2 \phi_3 \phi_4 + \phi_1 \overline{\phi_2} \phi_3 \phi_4 + \phi_1 \phi_2 \overline{\phi_3} \phi_4. \quad (13.65)$$

Defining now

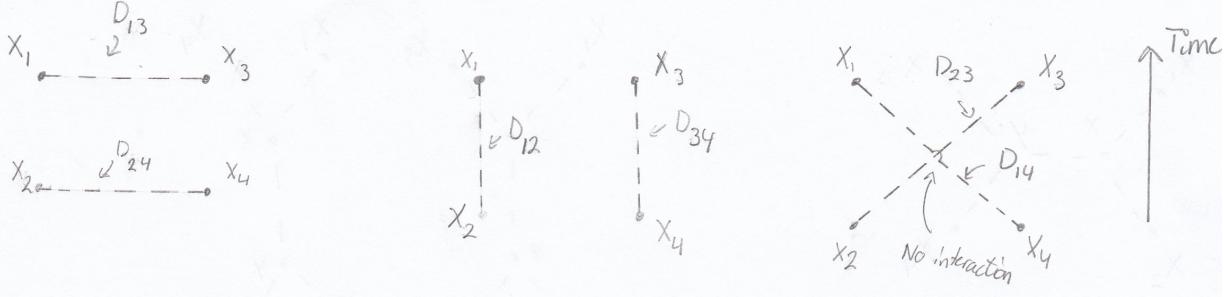
$$\overline{\phi(x_i)} \phi(x_j) \equiv D(x_i - x_j) = D_{ij}. \quad (13.66)$$

The VEV becomes

$$\langle 0 | T \{ \phi_1 \phi_2 \phi_3 \phi_4 \} | 0 \rangle = D_{12} D_{34} + D_{13} D_{24} + D_{14} D_{23}. \quad (13.67)$$

The VEV can be represented by Feynman diagrams

⁵Because the normal-ordering operator acts to put the annihilation operators to the right, they will act on the vacuum and make the term vanish.



The diagrams show two particles traveling between two separate space-time-points without interacting. By using the expression for the VEV in the LSZ reduction formula to $\mathcal{O}(\lambda^0)$, with $z = 1$, reveals

$$\frac{i}{p'_1 - m^2} \frac{i}{p'_2 - m^2} \frac{i}{p_1 - m^2} \frac{i}{p_2 - m^2} \langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = \int d^4 x_1 \int d^4 x_2 \int d^4 x_3 \int d^4 x_4 \\ \cdot e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} (D_{12} D_{34} + D_{13} D_{24} + D_{14} D_{23}). \quad (13.68)$$

There are three integrals, of the propagators, to be evaluated. The first

$$I_1 \equiv \int d^4 x_1 \int d^4 x_2 \int d^4 x_3 \int d^4 x_4 e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} D(x_1 - x_2) D(x_3 - x_4). \quad (13.69)$$

To evaluate I_1 I take $x' = x_1 - x_2$, $x'' = \frac{1}{2}(x_1 + x_2)$, $x''' = x_3 - x_4$, $x'''' = \frac{1}{2}(x_3 + x_4)$. Hereby

$$I_1 \equiv \left(\underbrace{\int d^4 x' \int d^4 x'' e^{i(p'_1 + p'_2)x''} e^{\frac{i}{2}(p'_1 - p'_2)x'} D(x')}_{\equiv I_4} \right) \\ \cdot \left(\underbrace{\int d^4 x''' \int d^4 x'''' e^{-i(p_1 + p_2)x''''} e^{-\frac{i}{2}(p_1 - p_2)x'''} D(x''')}_{\equiv I_5} \right). \quad (13.70)$$

Now

$$I_4 = \int d^4 x' \int d^4 x'' e^{i(p'_1 + p'_2)x''} e^{\frac{i}{2}(p'_1 - p'_2)x'} D(x') \\ = (2\pi)^4 \delta^{(4)}(p'_1 + p'_2) \int \frac{d^4 p'}{(2\pi)^4} \frac{i}{p'^2 - m^2} \underbrace{\int d^4 x' e^{-i(p' - \frac{p'_1}{2} + \frac{p'_2}{2}) \cdot x'}}_{= (2\pi)^4 \delta^{(4)}(p' - \frac{p'_1}{2} + \frac{p'_2}{2})}, \quad (13.71)$$

where the propagator has been expressed in the momentum representation⁶. Now, use $\delta^{(4)}(p'_1 + p'_2) \Rightarrow p'_1 = -p'_2$ in $\delta^{(4)}(p' - \frac{p'_1}{2} + \frac{p'_2}{2})$. Hereby $\delta^{(4)}(p' - p'_1)$. Hence, the integral can be killed off to reveal

$$I_4 = \frac{(2\pi)^4}{p'^2 - m^2}. \quad (13.72)$$

Likewise

$$I_5 = (2\pi)^4 \delta^{(4)}(p_1 + p_2) \int \frac{d^4 p}{(2\pi)^4} \frac{i}{p^2 - m^2} \underbrace{\int d^4 x''' e^{-i(p + \frac{p_1}{2} - \frac{p_2}{2}) \cdot x'''}}_{= (2\pi)^4 \delta^{(4)}(p + \frac{p_1}{2} - \frac{p_2}{2})}. \quad (13.73)$$

⁶Using $D(x_1 - x_2) = \int \frac{d^4 p'}{(2\pi)^4} \frac{i}{p'^2 - m^2} e^{-ip' \cdot (x_1 - x_2)}$.

Repeating the procedure results in $k_1 = -k_2 \Rightarrow \delta^{(4)}(p + \frac{p_1}{2} - \frac{p_2}{2}) = \delta^{(4)}(p + p_1)$. Hence, the integral will let $p = -p_1$. However, only p^2 is present in the integral, so $(-p_1)^2 = p_1^2$. Hence

$$I_5 = \frac{(2\pi)^4}{p_1^2 - m^2}. \quad (13.74)$$

Putting I_4 and I_5 together

$$I_1 = (2\pi)^8 \frac{i}{p_1'^2 - m^2} \frac{i}{p_1^2 - m^2}. \quad (13.75)$$

For I_2 I define $x' = x_1 - x_3, x'' = \frac{1}{2}(x_1 + x_3), x''' = x_2 - x_4, x'''' = \frac{1}{2}(x_2 + x_4)$. From these definitions $i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4) = \frac{i}{2}(p'_1 + p_1) \cdot x' + i(p'_1 - p_1) \cdot x'' + \frac{i}{2}(p'_2 + p_2) \cdot x''' + i(p'_2 - p_2) \cdot x''''$. This enables me to write the integral as

$$\begin{aligned} I_2 &\equiv \int d^4x_1 \int d^4x_2 \int d^4x_3 \int d^4x_4 e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} D(x_1 - x_3) D(x_2 - x_4) \\ &= \underbrace{\left(\int d^4x' \int d^4x'' e^{i(p'_1 - p_1)x''} e^{\frac{i}{2}(p'_1 + p_1)x'} D(x') \right)}_{\equiv I'_4} \\ &\quad \cdot \underbrace{\left(\int d^4x''' \int d^4x'''' e^{i(p'_2 - p_2)x''''} e^{\frac{i}{2}(p'_2 + p_2)x'''} D(x''') \right)}_{\equiv I'_5}. \end{aligned} \quad (13.76)$$

Now

$$\begin{aligned} I'_4 &= \underbrace{\int d^4x'' e^{i(p'_1 - p_1)x''}}_{=(2\pi)^4 \delta^{(4)}(p'_1 - p_1)} \int d^4x' e^{\frac{i}{2}(p'_1 + p_1)x'} D(x') \\ &= (2\pi)^4 \delta^{(4)}(p'_1 - p_1) \int \frac{d^4p'}{(2\pi)^4} \frac{i}{p'^2 - m^2} \underbrace{\int d^4x' e^{-i(p' - \frac{p'_1}{2} - \frac{p_1}{2}) \cdot x'}}_{(2\pi)^4 \delta^{(4)}(p' - \frac{p'_1}{2} - \frac{p_1}{2})}. \end{aligned} \quad (13.77)$$

Let again $\delta^{(4)}(p'_1 - p_1) \Rightarrow p'_1 = p_1$ and $\delta^{(4)}(p' - \frac{p'_1}{2} - \frac{p_1}{2}) = \delta^{(4)}(p' - p'_1)$. So

$$I'_4 = (2\pi)^4 \frac{i}{p_1'^2 - m^2}. \quad (13.78)$$

Likewise

$$I'_5 = (2\pi)^4 \frac{i}{p_2'^2 - m^2}. \quad (13.79)$$

Hence

$$I_2 = (2\pi)^8 \frac{i}{p_1'^2 - m^2} \frac{i}{p_2'^2 - m^2}. \quad (13.80)$$

I_3 can be found in the same way. The result is that $I_3 = I_2$. Hence, the LSZ formula

$$\begin{aligned} I_1 + I_2 + I_3 &= \frac{i}{p_1' - m^2} \frac{i}{p_2' - m^2} \frac{i}{p_1 - m^2} \frac{i}{p_2 - m^2} \langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle \\ &= (2\pi)^8 \frac{i}{p_1'^2 - m^2} \frac{i}{p_1^2 - m^2} + (2\pi)^8 \frac{i}{p_1'^2 - m^2} \frac{i}{p_2'^2 - m^2} \\ &\quad + (2\pi)^8 \frac{i}{p_1'^2 - m^2} \frac{i}{p_2^2 - m^2}. \end{aligned} \quad (13.81)$$

From the above I can isolate $\langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle$. The result is that

$$\langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = -(2\pi)^8 (p_2'^2 - m^2)(p_2^2 - m^2) - 2(2\pi)^8 (p_1^2 - m^2)(p_2^2 - m^2). \quad (13.82)$$

Now I go on mass shell, i.e. I enforce $p_i'^2 = m^2$ and $p_i^2 = m^2$, and so it is clear that $\langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = 0$. This will be the case whenever the poles on both sides do not match/cancel each other out! I could have guessed this result since $\langle \dots | iT | \dots \rangle$ describes the interaction part of the S-matrix and to zeroth order the particles do no interact - As illustrated by the Feynman diagrams in position space. In terms of Feynman diagrams; all disconnected diagrams will result in zero contribution.

2. Compute the amplitude to $\mathcal{O}(\lambda)$.

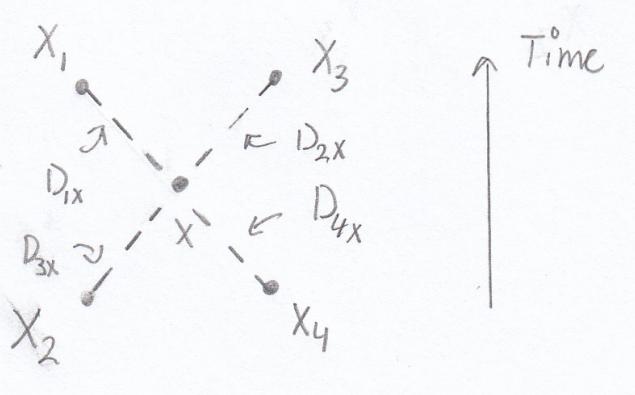
The first non-vanishing contribution to $\langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle$ comes from $\mathcal{O}(\lambda)$. A fast way to see this is the following; the first non-vanishing order is given as the first order above $\mathcal{O}(\lambda^0)$ in which the number of fields from \mathcal{H}_I and the scattering process is even! The reason for this is Wicks theorem. If there is an uneven number of fields they cannot all be contracted on the same time, and so the time-ordered series vanish. At $\mathcal{O}(\lambda)$, and $Z = 1$, the LSZ formula is given by

$$\left(\prod_{i=1}^2 \frac{i\sqrt{Z}}{p'_i - m^2} \right) \left(\prod_{j=1}^2 \frac{i\sqrt{Z}}{p_j - m^2} \right) \langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = \int d^4x_1 \int d^4x_2 \int d^4x_3 \int d^4x_4 \\ \cdot e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} \left(-\frac{i\lambda}{4!} \right) \int d^4x \frac{\langle 0 | T\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\phi^4(x)\} | 0 \rangle}{\langle 0 | T\{1 - i\frac{\lambda}{4!} \int d^4x \phi^4\} | 0 \rangle}. \quad (13.83)$$

To evaluate the VEV I once again use Wicks theorem. The number of contractions fill a full page⁷, so I will not do it here. Instead I will jump to the result

$$\langle 0 | T\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\phi^2(x)\} | 0 \rangle = 4! D(x_1 - x) D(x_2 - x) D(x_3 - x) D(x_4 - x). \quad (13.84)$$

The Feynman diagram in position space is given by



The factor of $4!$ cancels the factor of $\frac{1}{4!}$ from the Hamiltonian. Hereby, the I can write the LSZ reduction formula

$$\frac{i}{p'_1 - m^2} \frac{i}{p'_2 - m^2} \frac{i}{p_1 - m^2} \frac{i}{p_2 - m^2} \langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = \int d^4x_1 \int d^4x_2 \int d^4x_3 \int d^4x_4 \\ \cdot e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} (-i\lambda) \int d^4x \langle 0 | T\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\phi^4(x)\} | 0 \rangle \\ \cdot \frac{1}{\langle 0 | T\{1 - i\frac{\lambda}{4!} \int d^4x \phi^4\} | 0 \rangle}. \quad (13.85)$$

I consider

$$I = \int d^4x_1 \int d^4x_2 \int d^4x_3 \int d^4x_4 e^{i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4)} \\ \cdot (-i\lambda) \int d^4x \langle 0 | T\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\phi^4(x)\} | 0 \rangle. \quad (13.86)$$

⁷See notes from 8. semester.

To evaluate the integral I define $y_i = x_i - x$. Hereby $i(p'_1 \cdot x_1 + p'_2 \cdot x_2 - p_1 \cdot x_3 - p_2 \cdot x_4) = i(p'_1 p'_2 - p_1 \cdot p_2) \cdot x + i(p'_1 \cdot y_1 p'_2 \cdot y_2 - p_1 \cdot y_3 - p_2 \cdot y_4)$. Using this

$$I = -i\lambda \int d^4x e^{i(p'_1 + p'_2 - p_1 - p_2) \cdot x} \int d^4y_1 e^{ip'_1 \cdot y_1} D(y_1) \int d^4y_2 e^{ip'_2 \cdot y_2} D(y_2) \\ \cdot \int d^4y_3 e^{-ip_1 \cdot y_3} D(y_3) \int d^4y_4 e^{-ip_2 \cdot y_4} D(y_4). \quad (13.87)$$

Use now that $\int d^4y e^{op \cdot y} f(y) = \tilde{f}(p)$. Hereby

$$I = -i\lambda \tilde{D}(p'_1) \tilde{D}(p'_2) \tilde{D}(p_1) \tilde{D}(p_2) \int d^4x e^{i(p'_1 + p'_2 - p_1 - p_2) \cdot x} \\ = -i\lambda \tilde{D}(p'_1) \tilde{D}(p'_2) \tilde{D}(p_1) \tilde{D}(p_2) (2\pi)^4 \delta^{(4)}(p'_1 + p'_2 - p_1 - p_2) \quad (13.88) \\ = -i\lambda \frac{i}{p'_1 - m^2} \frac{i}{p'_2 - m^2} \frac{i}{p_1 - m^2} \frac{i}{p_2 - m^2} (2\pi)^4 \delta^{(4)}(p'_1 + p'_2 - p_1 - p_2).$$

using I in the LSZ formula the propagators now cancel, and so I find

$$\langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = -i\lambda (2\pi)^4 \delta^{(4)}(p'_1 + p'_2 - p_1 - p_2) \frac{1}{\langle 0 | T \{ 1 - i\frac{\lambda}{4!} \int d^4x \phi^4 \} | 0 \rangle}. \quad (13.89)$$

The denominator, ie $\langle 0 | T \{ 1 - i\frac{\lambda}{4!} \int d^4x \phi^4 \} | 0 \rangle$, only contribute with vacuum diagrams. this can be seen because no external fields enter. Hence, no particles will come out of this term, and only diagrams on the form shown below will enter - these are called vacuum diagrams, or bubble diagrams.



These diagrams do however also enter in the nominator, since what I have considered is only what can be measured. In practice all these vacuum diagrams can be added to the diagram describing the interaction cf.



Hence, the fraction in the LSZ formula will be on the form

$$\frac{\text{Diagram with 4 external lines}}{\text{Diagram with 4 external lines} + \text{Diagram with 5 external lines}} = \text{Diagram with 4 external lines}$$

So, I can omit the vacuum diagrams in the denominator if I only consider diagrams with external particles. Hereby, I find

$$\langle \vec{p}'_1 \vec{p}'_2 | iT | \vec{p}_1 \vec{p}_2 \rangle = -i\lambda(2\pi)^4 \delta^{(4)}(p'_1 + p'_2 - p_1 - p_2). \quad (13.90)$$

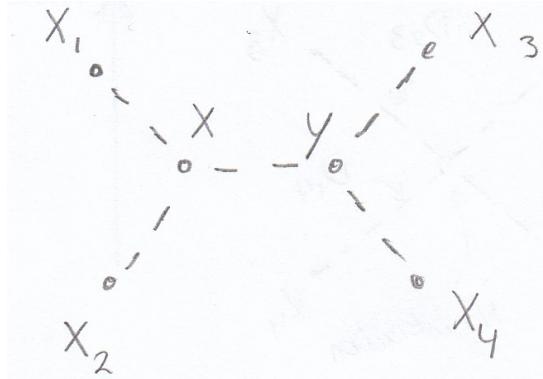
From which it is clear that $i\mathcal{M} = -i\lambda$.

3. Consider now a theory with $\mathcal{H}_I = \frac{\lambda}{3!}\phi^3$. What is the first non-vanishing term for this theory? Draw the relevant Feynman diagram.

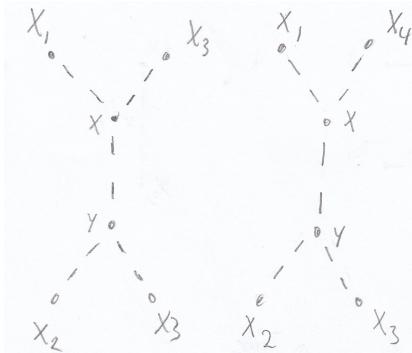
The $\mathcal{O}(\lambda^0)$ term vanishes for the same reason that the $\mathcal{O}(\lambda^0)$ term for ϕ^4 -theory vanished; there is no interaction to zeroth order. The $\mathcal{O}(\lambda^1)$ -term will however also vanish because the VEV will be on the form

$$VEV = \langle 0 | T\{\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\phi^3(x)\} | 0 \rangle. \quad (13.91)$$

Since the number of fields is uneven, not all fields can be contracted at the same time and so the VEV will be zero. Recall that this is because Wicks theorem is used to go from a time-ordered sequence of operators to a normal-ordered series. The normal-ordered terms will be zero if not all fields are contracted. Contracted fields make up a complex number that can go outside the VEV, but if there are uncontracted terms they will stay inside the VEV, and since the field is normal-ordered, the annihilation operator will be all the way to the right and annihilate onto the vacuum to make the terms vanish. Because of the previous arguments, the first non-vanishing order will be $\mathcal{O}(\lambda^2)$. Here, there will be connected terms and the number of fields will be 10 (even), so all terms can be contracted at the same time. Regarding the diagram; the previous procedure can be repeated to find the diagram, or alternatively the Feynman rules can be used. For each factor of $(-i\lambda)$ there is a vertex, where there, in this case, will be three joining lines. Hence



Again the vacuum diagrams will cancel. There are however other diagrams possible with two vertices:



These are all the possible diagrams to leading order - Since x_i has been connected to all x'_j s. From the Feynman rules the matrix element can be found rather quickly.

Example 13.2.

Determine the lowest non-vanishing order of perturbation theory in Compton scattering.

Compton scattering is an inelastic scattering between a photon and an electron, i.e.

$$e^-(p_1) + \gamma(p_2) \Rightarrow e^-(p'_1) + \gamma(p'_2). \quad (13.92)$$

From equation (13.27) the amplitude is given by

$$\begin{aligned} \mathcal{A} &= \langle \vec{p}'_2, \vec{p}'_1 | S | \vec{p}_2, \vec{p}_1 \rangle \\ &= \langle \vec{p}'_2, \vec{p}'_1 | T \left\{ 1 - i \int d^4x \mathcal{H}_I(x) + \frac{(-i)^2}{2!} \int d^4x_1 \int d^4x_2 \mathcal{H}_I(x_2) \mathcal{H}_I(x_1) + \dots \right\} | \vec{p}_2, \vec{p}_1 \rangle. \end{aligned} \quad (13.93)$$

The Hamiltonian density, \mathcal{H}_I , is given by

$$\mathcal{H}_I(x) = -|e| \bar{\psi}(x) \gamma^\mu \psi(x) A_\mu(x). \quad (13.94)$$

The initial and final states are taken to be different and so $\langle \vec{p}'_2, \vec{p}'_1 | \vec{p}_2, \vec{p}_1 \rangle = 0$. To evaluate the time-ordered product use Wick's theorem. Wick's theorem relates the time-ordered product ($T\{\dots\}$) of a sequence of field operators to the normal-ordered product ($N\{\dots\}$). In mathematical form Wick's theorem is given by

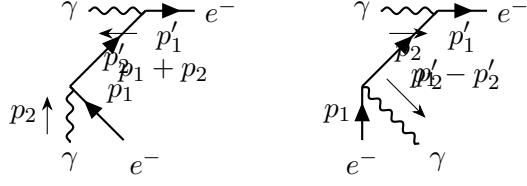
$$T\{\dots\} = N\{\dots + \text{All contractions}\}. \quad (13.95)$$

where "all contractions" refers to all contractions between the operators that the time ordering operator acts upon. The time ordering operator lists the contents with the latest entry farthest to the left whereas the normal ordering operator lists all the annihilation operators, belonging to the fields, to the right. Hence, the vacuum expectation value (VEV) of a normal ordered product is zero - since the annihilation operator acting on the vacuum is zero. Because the contraction between two field-operators is just a complex number, it can be moved outside the normal ordering operator. For this reason, only terms in which all field operators are contracted to another field operator contributes to the VEV. This leads to the important conclusion, that the number of fields must be an even number if there is to be any contribution from a given order of the perturbation theory. If the number of fields are uneven, it is not possible to contract all fields in a term, and the VEV will be zero. Because the number of fields vary with the order of the perturbation series, the leading order is the first order above zero which does not have an uneven number of fields. An important point here is that the number of fields is not the number of fields given in the interaction Hamiltonian density! There are extra fields stemming from the particle in the states involved (i.e. the initial and final states). These fields can be "pulled" out of the states (final and initial) by inserting the appropriate ladder operator, the vacuum state and a normalization factor. There are several different ways to explain, and do, this, but the point to remark here is that there appears an extra field for each final and initial particle involved in the process. Since there, for Compton scattering, are two particles, in both the initial and final state, the total number of fields is 4 plus what comes from the interaction Hamiltonian density. At $\mathcal{O}(i)$ there are three fields from the interaction Hamiltonian density, which results in seven fields in total, and so the contribution from this order vanishes. The leading order for Compton scattering is therefore $\mathcal{O}(i^2)$ (where there are 10 fields).

Example 13.3.

Determine the invariant matrix element for Compton scattering via the Feynman rules.

From considering the interaction Hamiltonian density of equation (13.94) it is clear that the vertices of QED will be of three particles; two fermions and one photon. As seen in the previous example the leading order for Compton scattering, and QED scattering in general, is $\mathcal{O}(i^2)$. Since the number of vertices follow the order of perturbation theory there will be two vertices. Hence, the relevant diagrams will have two vertices to each of which two fermions and one photon will be attached. There are two diagrams fulfilling the requirements:



where the direction of time is upwards. In order to write down the invariant matrix element from the Feynman diagrams, use the following procedure; at each side of the propagator is go towards the fermion current and use the feynman rules, then insert the propagator and go towards the fermion current on the other side of the propagator. Naming the vertices in the process means

$$\begin{aligned} i\mathcal{M}_{vertical} &= \bar{u}^{s'}(\vec{p}_1)(-iQ\gamma^\mu)\epsilon_\mu^{(\lambda')*}(\vec{p}_2)\frac{i(\not{p}_1 + \not{p}_2 + m)}{(p_1 + p_2)^2 - m^2}\epsilon_\rho^{(\lambda)}(\vec{p}_2)(-iQ\gamma^\rho)u^s(\vec{p}_1) \\ &= -i|e|^2\epsilon_\mu^{(\lambda')*}(\vec{p}_2)\epsilon_\rho^{(\lambda)}(\vec{p}_2)\frac{\bar{u}^{s'}(\vec{p}_1)\gamma^\mu(\not{p}_1 + \not{p}_2 + m)\gamma^\rho u^s(\vec{p}_1)}{(p_1 + p_2)^2 - m^2}, \end{aligned} \quad (13.96)$$

$$\begin{aligned} i\mathcal{M}_{horizontal} &= \bar{u}^{s'}(\vec{p}_1)(-iQ\gamma^\mu)\epsilon_\mu^{(\lambda)}(\vec{p}_2)\frac{i(\not{p}_1 - \not{p}_2 + m)}{(p_1 - p_2)^2 - m^2}\epsilon_\rho^{(\lambda')*}(\vec{p}_2)(-iQ\gamma^\rho)u^s(\vec{p}_1) \\ &= -i|e|^2\epsilon_\mu^{(\lambda)}(\vec{p}_2)\epsilon_\rho^{(\lambda')*}(\vec{p}_2)\frac{\bar{u}^{s'}(\vec{p}_1)\gamma^\mu(\not{p}_1 - \not{p}_2 + m)\gamma^\rho u^s(\vec{p}_1)}{(p_1 - p_2)^2 - m^2}. \end{aligned} \quad (13.97)$$

Example 13.4.

Determine the invariant matrix element for Compton scattering in position space.

Position space refers to the analog of the LSZ formula. However, instead of re-deriving the LSZ formula for QED it is convenient to use a slightly different approach. Using that the lowest non-vanishing order for QED is $\mathcal{O}(i^2)$, alongside the fact that if the zeroth order is neglected in the expansion of the S -matrix; $\mathcal{A} \Rightarrow \langle \dots' | iT | \dots \rangle$, equation (13.93) can be written viz

$$\langle \vec{p}'_2, \vec{p}'_1 | iT | \vec{p}_2, \vec{p}_1 \rangle = \frac{(i)^2 |e|^2}{2!} \int d^4x_1 \int d^4x_2 \langle \vec{p}'_2, \vec{p}'_1 | T\{\mathcal{H}_I(x_2)\mathcal{H}_I(x_1)\} | \vec{p}_2, \vec{p}_1 \rangle, \quad (13.98)$$

where

$$\mathcal{H}_I(x_2)\mathcal{H}_I(x_1) = \bar{\psi}(x_1)\gamma^\mu\psi(x_1)A_\mu(x_1)\bar{\psi}(x_2)\gamma^\nu\psi(x_2). \quad (13.99)$$

To evaluate equation 13.98 use Wick's theorem. From Wick's theorem the time-ordering of a product of field operators, sandwiched between two states, can be written as the normal ordered product of all possible contractions between the field operators. To determine which terms should enter the sum, use the definitions of the different contractions

$$\begin{aligned} \overline{\psi(x_1)\psi(x_2)} &= 0 \\ \overline{\bar{\psi}(x_1)\bar{\psi}(x_2)} &= 0 \\ \psi(x_1)\bar{\psi}(x_2) &= iS_F(x_1 - x_2) = \int \frac{d^4q}{(2\pi)^4} \frac{i(\not{q} + m)e^{-iq \cdot (x_1 - x_2)}}{q^2 - m^2} \\ \psi(x_i)|\vec{p}_1\rangle &= \int \frac{d^3q}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{q}}}} \sum_{s'} (u^{s'}(\vec{q})e^{-iq \cdot x_i}a_{\vec{q}}^{s'} + v^{s'}(\vec{q})e^{iq \cdot x_i}b_{\vec{q}}^{s'\dagger}) \sqrt{2E_{\vec{p}_1}} a_{\vec{p}_1}^{s\dagger} |0\rangle \\ &= \int \frac{d^3q}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{q}}}} \sum_{s'} u^{s'}(\vec{q})e^{-iq \cdot x_i}a_{\vec{q}}^{s'} \sqrt{2E_{\vec{p}_1}} a_{\vec{p}_1}^{s\dagger} |0\rangle \\ &= \int \frac{d^3q}{(2\pi)^3} \frac{\sqrt{2E_{\vec{p}_1}}}{\sqrt{2E_{\vec{q}}}} \sum_{s'} u^{s'}(\vec{q})e^{-iq \cdot x_i} (a_{\vec{p}_1}^{s\dagger}a_{\vec{q}}^{s'} + (2\pi)^3\delta^{(3)}(\vec{q} - \vec{p}_1)\delta_{s,s'}) |0\rangle \\ &= u^s(\vec{p}_1)e^{-ip_1 \cdot x_i} |0\rangle = \text{fermion, initial state.} \end{aligned} \quad (13.100)$$

From similar calculations

$$\bar{\psi}(x_i) \overline{|\vec{p}_1\rangle} = \bar{v}^s(p) e^{-ip \cdot x} |0\rangle = \text{anti fermion, initial state.} \quad (13.101)$$

In the case of the bra-contractions the other field operator is used

$$\begin{aligned} \langle \overline{\vec{p}'_1} | \psi(x_i) &= \langle 0 | b_{\vec{p}'_1}^s \sqrt{2E_{\vec{p}'_1}} \int \frac{d^3 q}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{q}}}} \sum_{s'} (u^{s'}(\vec{q}) e^{-iq \cdot x_i} a_{\vec{q}}^{s'} + v^{s'}(\vec{q}) e^{iq \cdot x_i} b_{\vec{q}}^{s'\dagger}) \\ &= \int \frac{d^3 q}{(2\pi)^3} \frac{\sqrt{2E_{\vec{p}'_1}}}{\sqrt{2E_{\vec{q}}}} \sum_{s'} \langle 0 | b_{\vec{p}'}^s b_{\vec{q}}^{s'\dagger} v^{s'}(\vec{q}) e^{iq \cdot x_i} \\ &= \int \frac{d^3 q}{(2\pi)^3} \frac{\sqrt{2E_{\vec{p}'_1}}}{\sqrt{2E_{\vec{q}}}} \sum_{s'} v^{s'}(\vec{q}) e^{iq \cdot x_i} \langle 0 | (b_{\vec{q}}^{s'\dagger} b_{\vec{p}'}^s + (2\pi)^3 \delta^{(3)}(\vec{q} - \vec{p}'_1) \delta_{s,s'}) \\ &= \langle 0 | v^{s'}(\vec{p}'_1) e^{ip'_1 \cdot x_i} = \text{anti fermion, final state.} \end{aligned} \quad (13.102)$$

Similarly

$$\langle \overline{\vec{p}'_1} | \bar{\psi}(x_i) = \langle 0 | \bar{u}^{s'}(\vec{p}'_1) e^{ip'_1 \cdot x_i} = \text{fermion, final state.} \quad (13.103)$$

Lastly, for the photons the possible contractions are

$$\begin{aligned} A_\mu \overline{(x_i) |\vec{p}_2\rangle} &= \int \frac{d^3 q}{(2\pi)^3} \frac{1}{\sqrt{2E_{\vec{q}}}} \sum_r (\varepsilon_\mu^r(\vec{q}) e^{-iq \cdot x_i} a_{\vec{q}}^r + \varepsilon_\mu^{r*}(\vec{q}) e^{iq \cdot x_i} a_{\vec{q}}^{r\dagger}) \sqrt{2E_{\vec{p}_2}} a_{\vec{p}_2}^{s\dagger} |0\rangle \\ &= \int \frac{d^3 q}{(2\pi)^3} \frac{\sqrt{2E_{\vec{k}}}}{\sqrt{2E_{\vec{q}}}} \sum_r \varepsilon_\mu^r(\vec{q}) e^{-iq \cdot x_i} a_{\vec{q}}^r a_{\vec{p}_2}^{s\dagger} |0\rangle \\ &= \int \frac{d^3 q}{(2\pi)^3} \frac{\sqrt{2E_{\vec{p}_2}}}{\sqrt{2E_{\vec{q}}}} \sum_r \varepsilon_\mu^r(\vec{q}) e^{-iq \cdot x_i} (a_{\vec{k}}^{s\dagger} a_{\vec{q}}^r + (2\pi)^3 \delta^{(3)}(\vec{q} - \vec{p}_2) \delta_{s,r}) |0\rangle \\ &= \varepsilon_\mu^r(\vec{p}_2) e^{-ip_2 \cdot x_i} |0\rangle. \end{aligned} \quad (13.104)$$

Likewise

$$\langle \overline{\vec{p}_2} | A_\mu(x_i) = \langle 0 | \varepsilon_\mu^{r'*}(\vec{p}_2) e^{ip_2 \cdot x_i}. \quad (13.105)$$

With the contractions defined as above Wick's theorem can be used to significantly simplify equation 13.98. From Wick's theorem all terms must be contracted to contribute. Further more, contractions between A and the other fields are zero, just as contractions among ψ 's and $\bar{\psi}$'s. Hence, to contribute there must be terms on the form

$$\psi \overline{(x_1) \bar{\psi}(x_2)}, \quad \psi \overline{(x_i) |\vec{p}\rangle}, \quad \langle \overline{\vec{p}'} | \bar{\psi}(x_i), \quad \langle \overline{\vec{k}'} | A_\mu(x_i), \quad A_\mu \overline{(x_i) |\vec{k}\rangle}. \quad (13.106)$$

There are four terms like this. Defining $I \equiv \langle \overline{\vec{p}_2}, \overline{\vec{p}_1} | T\{\mathcal{H}_I(x_2)\mathcal{H}_I(x_1)\} | \vec{p}_2, \vec{p}_1 \rangle$

$$\begin{aligned} I &= N \left\{ \langle \overline{\vec{p}'_1}, \overline{\vec{p}'_2} | \bar{\psi}(x_1) \gamma^\mu \psi(x_1) A_\mu(x_1) \bar{\psi}(x_2) \gamma^\nu \psi(x_2) A_\nu(x_2) | \vec{p}_1, \vec{p}_2 \rangle \right. \\ &\quad + \langle \overline{\vec{p}'_1}, \overline{\vec{p}'_2} | \bar{\psi}(x_1) \gamma^\mu \psi(x_1) A_\mu(x_1) \bar{\psi}(x_2) \gamma^\nu \psi(x_2) A_\nu(x_2) | \vec{p}_1, \vec{p}_2 \rangle \\ &\quad + \langle \overline{\vec{p}'_1}, \overline{\vec{p}'_2} | \bar{\psi}(x_1) \gamma^\mu \psi(x_1) A_\mu(x_1) \bar{\psi}(x_2) \gamma^\nu \psi(x_2) A_\nu(x_2) | \vec{p}_1, \vec{p}_2 \rangle \\ &\quad \left. + \langle \overline{\vec{p}'_1}, \overline{\vec{p}'_2} | \bar{\psi}(x_1) \gamma^\mu \psi(x_1) A_\mu(x_1) \bar{\psi}(x_2) \gamma^\nu \psi(x_2) A_\nu(x_2) | \vec{p}_1, \vec{p}_2 \rangle \right\}. \end{aligned} \quad (13.107)$$

There are two different fermion terms and two different photon terms. The two different fermion terms are in fact identical. This can be seen as follows

$$\begin{aligned} N\{\overline{[\psi(x_1)\mathcal{A}(x_1)\psi(x_1)]}\overline{[\psi(x_2)\mathcal{A}(x_2)\psi(x_2)]}\} &= (-1)^4 N\{\overline{[\bar{\psi}(x_2)\mathcal{A}(x_2)\psi(x_2)]}\overline{[\bar{\psi}(x_1)\mathcal{A}(x_1)\psi(x_1)]}\} \\ &= N\{\overline{[\bar{\psi}(x_1)\mathcal{A}(x_1)\psi(x_1)]}\overline{[\bar{\psi}(x_2)\mathcal{A}(x_2)\psi(x_2)]}\}. \end{aligned} \quad (13.108)$$

Hereby

$$I = 2\bar{u}^{s'}(\vec{p}'_1)\gamma^\mu iS_F(x_1 - x_2)\gamma^\nu u^s(\vec{p}_1)e^{ip'_1 \cdot x_1}e^{-ip_1 \cdot x_2}N\{\langle \vec{p}'_2 | A_\mu(x_1)A_\nu(x_2) | \vec{p}_2 \rangle\}. \quad (13.109)$$

The A' 's can each couple to either the initial or final state, so

$$N\{\langle \vec{p}'_2 | A_\mu(x_1)A_\nu(x_2) | \vec{p}_2 \rangle\} = \varepsilon_\mu^{r'*}(\vec{p}'_2)\varepsilon_\nu^r(\vec{p}_2)e^{ip'_2 \cdot x_1}e^{-ip_2 \cdot x_2} + \varepsilon_\nu^{r'*}(\vec{p}'_2)\varepsilon_\mu^r(\vec{p}_2)e^{ip'_2 \cdot x_2}e^{-ip_2 \cdot x_1}. \quad (13.110)$$

Hereby

$$\begin{aligned} \langle \vec{p}'_1, \vec{p}'_2 | i\hat{T} | \vec{p}_1, \vec{p}_2 \rangle &= i|e|^2 \sum_{i=1}^2 T_i. \end{aligned} \quad (13.111)$$

$\cdot \left(\int d^4x_1 \int d^4x_2 \bar{u}^{s'}(\vec{p}'_1)\gamma^\mu S_F(x_1 - x_2)\gamma^\nu u^s(\vec{p}_1)e^{ip'_1 \cdot x_1}e^{-ip_1 \cdot x_2}\varepsilon_\mu^{r'*}(\vec{p}'_2)\varepsilon_\nu^r(\vec{p}_2)e^{ip'_2 \cdot x_1}e^{-ip_2 \cdot x_2} \right. \\ \left. + \int d^4x_1 \int d^4x_2 \bar{u}^{s'}(\vec{p}'_1)\gamma^\mu S_F(x_1 - x_2)\gamma^\nu u^s(\vec{p}_1)e^{ip'_1 \cdot x_1}e^{-ip_1 \cdot x_2}\varepsilon_\nu^{r'*}(\vec{p}'_2)\varepsilon_\mu^r(\vec{p}_2)e^{ip'_2 \cdot x_2}e^{-ip_2 \cdot x_1} \right)$

Considering each of the terms

$$\begin{aligned} T_1 &= \int d^4x_1 \int d^4\bar{u}^{s'}(\vec{p}'_1)\gamma^\mu S_F(x_1 - x_2)\gamma^\nu u^s(\vec{p}_1)e^{ip'_1 \cdot x_1}e^{-ip_1 \cdot x_2}\varepsilon_\mu^{r'*}(\vec{p}'_2)\varepsilon_\nu^r(\vec{p}_2)e^{ip'_2 \cdot x_1}e^{-ip_2 \cdot x_2} \\ &= \int \frac{d^4q}{(2\pi)^4} \bar{u}^{s'}(\vec{p}'_1)\gamma^\mu \frac{\not{q} + m}{q^2 - m^2} \gamma^\nu u^s(\vec{p}_1)\varepsilon_\mu^{r'*}(\vec{p}'_2)\varepsilon_\nu^r(\vec{p}_2) \underbrace{\int d^4x_1 e^{ip'_2 \cdot x_1}e^{ip'_1 \cdot x_1}e^{-iqx_1}}_{(2\pi)^4 \delta^{(4)}(p'_2 + p'_1 - q)} \underbrace{\int d^4x_2 e^{-ip_2 \cdot x_2}e^{-ip_1 \cdot x_2}e^{iqx_2}}_{(2\pi)^4 \delta^{(4)}(p_2 + p_1 - q)} \\ &= (2\pi)^4 \delta^{(4)}(p'_2 + p'_1 - p_1 - p_2) \bar{u}^{s'}(\vec{p}'_1)\gamma^\mu \frac{\not{p}_1 + \not{p}_2 + m}{(p_1 + p_2)^2 - m^2} \gamma^\nu u^s(\vec{p}_1)\varepsilon_\mu^{r'*}(\vec{p}'_2)\varepsilon_\nu^r(\vec{p}_2), \\ T_2 &= \int d^4x_1 \int d^4x_2 \bar{u}^{s'}(\vec{p}'_1)\gamma^\mu S_F(x_1 - x_2)\gamma^\nu u^s(\vec{p}_1)e^{ip'_1 \cdot x_1}e^{-ip_1 \cdot x_2}\varepsilon_\nu^{r'*}(\vec{p}'_2)\varepsilon_\mu^r(\vec{p}_2)e^{ip'_2 \cdot x_2}e^{-ip_2 \cdot x_1} \\ &= \int \frac{d^4q}{(2\pi)^4} \bar{u}^{s'}(\vec{p}'_1)\gamma^\mu \frac{\not{q} + m}{q^2 - m^2} \gamma^\nu u^s(\vec{p}_1)\varepsilon_\nu^{r'*}(\vec{p}'_2)\varepsilon_\mu^r(\vec{p}_2) \underbrace{\int d^4x_1 e^{ip'_1 \cdot x_1}e^{-ip_2 \cdot x_1}e^{-iqx_1}}_{(2\pi)^4 \delta^{(4)}(p'_1 - p_2 - q)} \underbrace{\int d^4x_2 e^{-ip_1 \cdot x_2}e^{ip'_2 \cdot x_2}e^{iqx_2}}_{(2\pi)^4 \delta^{(4)}(p_1 - p'_2 - q)} \\ &= (2\pi)^4 \delta^{(4)}(p'_2 + p'_1 - p_1 - p_2) \bar{u}^{s'}(\vec{p}'_1)\gamma^\mu \frac{\not{p}_1 - \not{p}'_2 + m}{(p'_1 - p'_2)^2 - m^2} \gamma^\nu u^s(\vec{p}_1)\varepsilon_\nu^{r'*}(\vec{p}'_2)\varepsilon_\mu^r(\vec{p}_2). \end{aligned} \quad (13.112)$$

By combining equations (13.111) and (13.112)

$$\begin{aligned} \langle \vec{p}'_1, \vec{p}'_2 | i\hat{T} | \vec{p}_1, \vec{p}_2 \rangle &= -i(2\pi)^4 \delta^{(4)}(p'_2 + p'_1 - p_1 - p_2) |e|^2 \varepsilon_\nu^{r'*}(\vec{p}'_2)\varepsilon_\mu^r(\vec{p}_2) \\ &\quad \cdot \left(\bar{u}^{s'}(\vec{p}'_1)\gamma^\mu \frac{\not{p}_1 + \not{p}_2 + m}{(p_1 + p_2)^2 - m^2} \gamma^\nu u^s(\vec{p}_1) + \bar{u}^{s'}(\vec{p}'_1)\gamma^\nu \frac{\not{p}_1 - \not{p}'_2 + m}{(p_1 - p'_2)^2 - m^2} \gamma^\mu u^s(\vec{p}_1) \right), \end{aligned} \quad (13.113)$$

where the dummy indices (μ, ν) have been relabeled on the γ' 's on the second term in order to factorize the ϵ' 's. From equation (13.113)

$$\begin{aligned} i\mathcal{M} &= -i|e|^2 \varepsilon_\nu^{r'*}(\vec{p}'_2)\varepsilon_\mu^r(\vec{p}_2) \\ &\quad \cdot \left(\bar{u}^{s'}(\vec{p}'_1)\gamma^\mu \frac{\not{p}_1 + \not{p}_2 + m}{(p_1 + p_2)^2 - m^2} \gamma^\nu u^s(\vec{p}_1) + \bar{u}^{s'}(\vec{p}'_1)\gamma^\nu \frac{\not{p}_1 - \not{p}'_2 + m}{(p_1 - p'_2)^2 - m^2} \gamma^\mu u^s(\vec{p}_1) \right). \end{aligned} \quad (13.114)$$

By comparison between equation (13.114), (13.97) and (13.96) it is clear that the same result is obtained as in the previous example where the invariant matrix element was written directly from the Feynman rules of QED.

13.5 Effective field theory

Particle physics often consider scattering events between hadrons and mesons. The issue with such composite particles is that the amplitude written in terms of the Feynman rules does not apply unless the particles considered are free. Quarks confined in hadrons and mesons are not free (the majority of the hadron mass is made up of the gluon field produced by the quark) and further more; the underlying theory describing the behavior of the quarks (QCD) is non-perturbative. In such cases the Feynman rules cannot be used and as such the position space approach is used. It goes as follows; the interaction (the states evolve via the interaction Hamiltonian density, \mathcal{H}_I , which in turn is read off from the interacting part of the interaction Lagrangian density, \mathcal{L}_I) Lagrangian density is related to the amplitude for a scattering by using the definition of S from equation (13.47) alongside equation (13.23) through (13.27). Taking the states to be more general, eg. $|p_1, A\rangle$ for a particle scattering of a nucleus. p_1 denotes the momentum of the scattered particle whereas A denotes the more complex state of the nucleus (eg. with momentum, spin,...) and using equation (13.47) in equation (13.23) (and expanding the exponential)

$$\begin{aligned}\mathcal{A} &= \langle f|S|i\rangle \\ &= \langle f|i\rangle - i \int d^4x \langle f|\mathcal{H}_I(x)|i\rangle + \mathcal{O}(i^2),\end{aligned}\tag{13.115}$$

where i denotes the initial state, f denotes the final state and $\mathcal{H}_I(x)$ is the interaction Hamiltonian in the interaction picture. From equation (13.27)

$$\mathcal{A} = \langle f|i\rangle + (2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p) i \mathcal{M}.\tag{13.116}$$

By equating equation (13.115) and (13.116)

$$(2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p) i \mathcal{M} = -i \int d^4x \langle f|\mathcal{H}_I(x)|i\rangle + \mathcal{O}(i^2).\tag{13.117}$$

Taking p^μ to be the space-time translation operator, the interaction Hamiltonian density can be written as follows

$$\mathcal{H}_I(x) = e^{ipx} \mathcal{H}_I(0) e^{-ipx}.\tag{13.118}$$

By applying equation (13.118) in equation (13.117)

$$\begin{aligned}(2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p) \mathcal{M} &= - \int d^4x \langle f|e^{ipx} \mathcal{H}_I(0) e^{-ipx}|i\rangle + \mathcal{O}(i^2) \\ &= - \int d^4x e^{-i(\Sigma p - \Sigma p')x} \langle f|\mathcal{H}_I(0)|i\rangle + \mathcal{O}(i^2) \\ &= -(2\pi)^4 \delta^{(4)}(\Sigma p' - \Sigma p) \langle f|\mathcal{H}_I(0)|i\rangle + \mathcal{O}(i^2).\end{aligned}\tag{13.119}$$

From equation (13.119) it is clear that

$$\mathcal{M} = -\langle f|\mathcal{H}_I(0)|i\rangle + \mathcal{O}(i^2).\tag{13.120}$$

If the interaction Lagrangian does not contain derivatives, then $-\mathcal{L}_I(0) = \mathcal{H}_I(0)$. Hereby

$$\mathcal{M} = \langle f|\mathcal{L}_I(0)|i\rangle + \mathcal{O}(i^2).\tag{13.121}$$

$\mathcal{L}_I(0)$ can then be described by an effective Lagrangian density. An effective Lagrangian density is a Lagrangian density that is proportional to product of the two currents of the interaction. The effective Lagrangian density is valid only in the limit in which the momentum of the mediator squared goes to zero, i.e. $q^2 \Rightarrow 0$. In the case of a massless mediator, eg. a photon, this would make the propagator diverge, so this approach is only valid in the case of a massive propagator, eg. the low energy limit of the electroweak theory. The usefulness of this approach is when scatterings between constituent particles is considered.

Example 13.5.

For example WIMP-nucleus scattering. In this case the WIMP will scatter with the quarks inside the nucleons in the nucleus. Hence, the particle are not free and the usual Feynman rules do not apply. However, an effective Lagrangian density can be written viz

$$\mathcal{L}_I(0) \simeq \sum_{x,y} \sum_q \frac{C_q^{(xy)}}{\Lambda^2} [\bar{\chi} \Gamma_x \chi] [\bar{q} \Gamma_y q], \quad (13.122)$$

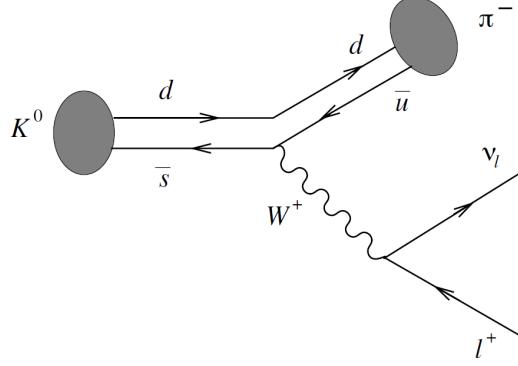
where q denotes quarks, Λ^2 is the mediator mass squared, Γ_i denote the possible objects that makes the term a Lorentz scalar and $C_q^{(xy)} \sim g^2$ is related to the square of the coupling. From this effective Lagrangian density, the invariant matrix element is given by

$$\mathcal{M} \simeq \sum_{x,y} \sum_q \frac{C_q^{(xy)}}{\Lambda^2} \langle \chi_f | \bar{\chi} \Gamma_x \chi | \chi_i \rangle \langle A_f | \bar{q} \Gamma_y q | A_i \rangle, \quad (13.123)$$

where A denotes the nucleus states. From this expression the matrix elements can, through a considerable amount of work, be determined.

Example 13.6.

As an illustrative example, consider the decay of a neutral kaon, K^0 , into negative pion, π^- , a neutrino, ν_l , and its corresponding lepton, l^+ , i.e. $K^0 \Rightarrow \pi^- + l^+ + \nu_l$ (see the Feynman diagram below).



In this case the states read $|i\rangle = |\nu_l = 0\rangle \otimes |l^+ = 0\rangle \otimes |K^0\rangle$ and $\langle f| = \langle \pi^- | \otimes \langle l^+ | \otimes \langle \nu_l |$. The effective Lagrangian density for the decay is given by

$$\mathcal{L}_I(0) = -\frac{G_f \sin(\theta_C)}{\sqrt{2}} \bar{\nu}_l \gamma_\mu (1 - \gamma^5) l \bar{s} \gamma^\mu (1 - \gamma^5) u. \quad (13.124)$$

Equation (13.124) is a Fermi Lagrangian density since the mediator is integrated out. It is however not an effective Lagrangian density since the coupling is specified. From equation (13.124)

$$\begin{aligned} \mathcal{M} &= -\frac{G_f \sin(\theta_C)}{\sqrt{2}} \langle f | \bar{\nu}_l \gamma_\mu (1 - \gamma^5) l \bar{s} \gamma^\mu (1 - \gamma^5) u | i \rangle + \mathcal{O}(i^2) \\ &= -\frac{G_f \sin(\theta_C)}{\sqrt{2}} \langle \nu_l | \bar{\nu}_l | 0 \rangle \gamma_\mu (1 - \gamma^5) \langle l^+ | l | 0 \rangle \langle \pi^- | \bar{s} \gamma^\mu (1 - \gamma^5) u | K^0 \rangle + \mathcal{O}(i^2) \\ &= -\frac{G_f \sin(\theta_C)}{\sqrt{2}} \bar{u}(\nu_l) \gamma_\mu (1 - \gamma^5) v(l^+) \langle \pi^- | \bar{s} \gamma^\mu (1 - \gamma^5) u | K^0 \rangle + \mathcal{O}(i^2), \end{aligned} \quad (13.125)$$

where it is understood that all matrix elements are evaluated at $x = 0$. The matrix element of the quark current, i.e. the second operator, cannot be evaluated directly like the matrix element of the leptonic matrix element, i.e. the first operator. Instead the matrix element of the second operator can be parameterized in terms of form factors. The parameterization is determined by Lorentz invariance and depends in general on

the specific process under consideration. The procedure to determine the amplitude in processes involving composite particles is however clear; use the Fermi Lagrangian and determine the quark-matrix elements via parameterization.

Appendices

A | Code for Computational Fluid Dynamics

2 D Poiseuille Flow

```
Clear[vx, vy, Fx, Fy, p, Nx, Ny, U, ix, iy, t, dt, T, s, A];
ρ = 1.2041; (*density of the fluid (air at 293.15K in this case)*)
L = 10; (*Physical length of the grid*)
a = 1; (*Physical height of the grid*)
Nx = 40; (*Number of grid points over the length of the grid*)
Ny = 20; (*Number of grid points over the height of the grid*)
U = 1; (*Entry velocity of the fluid*)
maxtime = 400; (*The number of time-iterations*)
Rey = 100; (*Reynolds number for the fluid*)
nu =  $\frac{U * a}{Rey}$ ; (*The dynamics viscosity of the fluid*)
dt = 0.02; (*The physical time between time stamps*)
Δx =  $\frac{L}{Nx}$ ; (*The physical distance between grid points in the length direction*)
Δy =  $\frac{a}{Ny}$ ; (*The physical distance between grid points in the height direction*)
(*The relevant series of matrices are created*)
vx = Table[Table[0, {Ny + 2}, {Nx + 1}], {maxtime}];
vy = Table[Table[0, {Ny + 1}, {Nx + 2}], {maxtime}];
Fx = Table[Table[0, {Ny + 2}, {Nx + 1}], {maxtime}];
Fy = Table[Table[0, {Ny + 1}, {Nx + 2}], {maxtime}];
p = Table[Table[0, {Ny + 2}, {Nx + 2}], {maxtime}];
s = Table[Table[0, {Ny + 2}, {Nx + 2}], {maxtime}];
A = Table[Table[0, {Ny + 2}, {Nx + 2}], {maxtime}];

(*The velocity in the length direction is initialized*)
For[i = Nx + 1, i ≥ 1, i--,
  For[j = Ny + 1, j > 1, j--,
    vx[[1, j, i]] = U;
  ];
];

(*The A-matrix used in the pressure-subroutine with incorporated BC's (A.p=s) .*)
Clear[n, A];
A = Table[0, {(Ny) * (Nx)}, {(Ny) * (Nx)}];
n = 0;
For[i = 1, i ≤ (Ny) * (Nx), i++,
  n++;
  If[n == Nx + 1, n = 1, 0];
  For[j = 1, j ≤ (Ny) * (Nx), j++,
    A[[i, j]] =
      -2  $\left( \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta x)^2} \right)$  + If[1 ≤ i ≤ Nx ∨ Ny * Nx - Nx + 1 ≤ i ≤ Ny * Nx,  $\frac{1}{(\Delta y)^2}$ , 0] +
      If[n == 1 ∨ n == Nx,  $\frac{1}{(\Delta x)^2}$ , 0] * KroneckerDelta[i, j] +  $\frac{1}{(\Delta x)^2}$  *
      (If[n ≠ Nx, KroneckerDelta[i, j - 1], 0] + If[n ≠ 1, KroneckerDelta[i, j + 1], 0]) +
       $\frac{1}{(\Delta y)^2}$  * (KroneckerDelta[i, j - (Nx)] + KroneckerDelta[i, j + Nx])
  ];
];
```

```

];
];
(*The correct solution is only obtained
when the rank of A is equal to the number of rows*)
Clear[n];
n = 0;
For[i = 1, i ≤ (Ny) * (Nx), i++,
n++;
If[Dimensions[A] [[1]] ≠ MatrixRank[A],
A = A[[1 ; (Ny) * (Nx) - n, 1 ; (Ny) * (Nx) - n]];
, Indeterminate];
];
(*Begin the time loop untill maxtime-1*)
For[t = 1, t < maxtime, t++,
(*calculate the next F_x-vlaue for internal points*)
For[iy = 2, iy ≤ Ny + 1, iy++,
(*iy runs over the vertical physical space + dummy boundary*)
For[ix = 2, ix ≤ Nx, ix++, (*ix runs over the horizontal physical space+
dummy boundary*)
Fx[[t, iy, ix]] = ρ * (-vx[[t, iy, ix]] *  $\frac{vx[[t, iy, ix+1]] - vx[[t, iy, ix-1]]}{2 * Δx}$  -
 $\frac{1}{4} (vy[[t, iy, ix]] + vy[[t, iy, ix+1]] + vy[[t, iy-1, ix]] +$ 
 $vy[[t, iy-1, ix+1]]) * \frac{vx[[t, iy+1, ix]] - vx[[t, iy-1, ix]]}{2 * Δy}$ ) +
nu * ( $\frac{1}{(Δx)^2} (vx[[t, iy, ix+1]] + vx[[t, iy, ix-1]] - 2 vx[[t, iy, ix]]) +$ 
 $\frac{1}{(Δy)^2} (vx[[t, iy+1, ix]] + vx[[t, iy-1, ix]] - 2 vx[[t, iy, ix]])$ );
];
];
(*calculate the next F_y-vlaue for internal points*)
For[iy = 2, iy ≤ Ny, iy++,
For[ix = 2, ix ≤ Nx + 1, ix++,
Fy[[t, iy, ix]] = ρ * (-vy[[t, iy, ix]] *  $\frac{vy[[t, iy+1, ix]] - vy[[t, iy-1, ix]]}{2 * Δy}$  -
 $\frac{1}{4} (vx[[t, iy, ix]] + vx[[t, iy, ix-1]] + vx[[t, iy+1, ix]] +$ 
 $vx[[t, iy+1, ix-1]]) * \frac{vy[[t, iy, ix+1]] - vy[[t, iy, ix-1]]}{2 * Δx}$ ) +
nu * ( $\frac{1}{(Δy)^2} (vy[[t, iy+1, ix]] + vy[[t, iy-1, ix]] - 2 vy[[t, iy, ix]]) +$ 
 $\frac{1}{(Δx)^2} (vy[[t, iy, ix+1]] + vy[[t, iy, ix-1]] - 2 vy[[t, iy, ix]])$ );
];
];
(*calculate the next source-vlaue for internal points*)
For[ix = 2, ix ≤ Nx + 1, ix++,
For[iy = 2, iy ≤ Ny + 1, iy++,
s[[t, iy, ix]] =

```

```


$$\frac{Fx[[t, iy, ix]] - Fx[[t, iy, ix - 1]]}{\Delta x} + \frac{Fy[[t, iy, ix]] - Fy[[t, iy - 1, ix]]}{\Delta y} + \frac{\rho}{dt} * \left( \frac{vx[[t, iy, ix]] - vx[[t, iy, ix - 1]]}{\Delta x} + \frac{vy[[t, iy, ix]] - vy[[t, iy - 1, ix]]}{\Delta y} \right);$$

];
];
(*Start the subroutine that determine the pressure corresponding to the source*)
(*First, arrange the s-matrix indices in a vector, S*)
Clear[S, n];
S = Table[0, {Nx * Ny}];
n = 0;
For[i = 2, i ≤ Ny + 1, i++,
  For[j = 2, j ≤ Nx + 1, j++,
    n++;
    S[[n]] = s[[t, i, j]];
  ];
];
(*Next, remove entries from the end of S corresponding
to the rows removed in A in order to obtain rank(A)=rows(A) *)
Clear[temp];
S = S[[1 ;; Dimensions[A][[1]]]];
(*Solve A.x=S. Inverting such large matrices is computationally too heavy,
so approximate methods must be applied. This can be done via SOR,
but the inbuilt function is much, much faster.*)
temp = LinearSolve[A, S];
(*"temp" is a vector corresponding to S. This is rearranged
into a matrix with pressure measurements, corresponding to s*)
n = 0;
For[i = 1, i ≤ Ny, i++,
  For[j = 1, j ≤ Nx, j++,
    n++;
    If[n ≤ Dimensions[temp][[1]] ∧ temp[[j + (i - 1) * Nx]] > 10^-5,
      p[[t, i + 1, j + 1]] = temp[[j + (i - 1) * Nx]];
      , Indeterminate];
  ];
];
(*The boundary terms are set equal to the outer interla points. Without
incorporating the BC's in the A-matrix this would eb wrong*)
For[ix = 2, ix ≤ Nx + 1, ix++,
  p[[t, 1, ix]] = p[[t, 2, ix]];
  p[[t, Ny + 2, ix]] = p[[t, Ny + 1, ix]];
];
For[iy = 2, iy ≤ Ny + 1, iy++,
  p[[t, iy, 1]] = p[[t, iy, 2]];
  p[[t, iy, Nx + 2]] = p[[t, iy, Nx + 1]];
];
(*calculate the next v_x-vlaue for internal points*)
For[ix = 1, ix ≤ Nx + 2, ix++,
  For[iy = 1, iy ≤ Ny + 2, iy++,
    If[2 ≤ iy ≤ Ny + 1,
      If[2 ≤ ix ≤ Nx,
        vx[[t + 1, iy, ix]] =

```

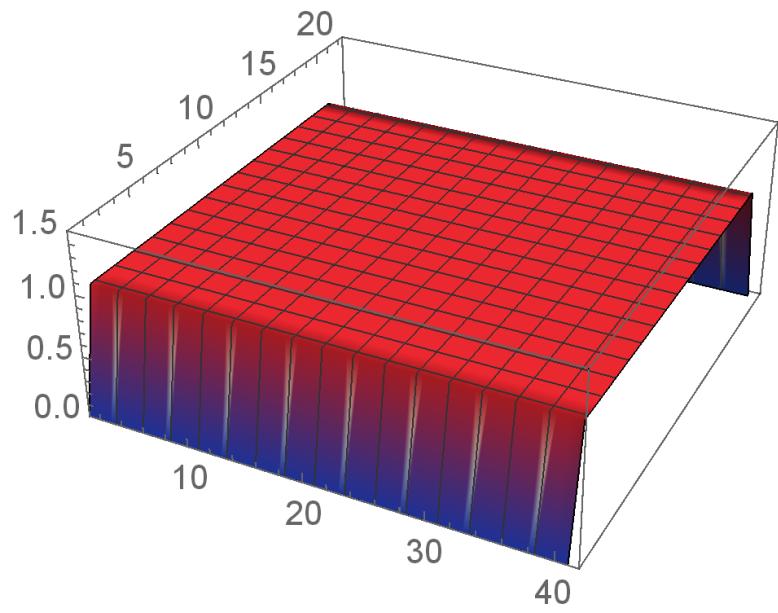
```

vx[[t, iy, ix]] + dt / ρ * (Fx[[t, iy, ix]] - p[[t, iy, ix+1]] - p[[t, iy, ix]]) / Δx;
, Indeterminate];
, Indeterminate];
(*calculate the next v_y-value for internal points*)
If[2 ≤ iy ≤ Ny,
If[2 ≤ ix ≤ Nx + 1,
vy[[t + 1, iy, ix]] =
vy[[t, iy, ix]] + dt / ρ * (Fy[[t, iy, ix]] - p[[t, iy+1, ix]] - p[[t, iy, ix]]) / Δy;
, Indeterminate];
, Indeterminate];
(*Apply BC's for v_x and v_y*)
If[2 ≤ iy ≤ Ny + 1,
vx[[t + 1, iy, 1]] = 0;
vx[[t + 1, iy, Nx + 1]] = vx[[t + 1, iy, Nx]];
, Indeterminate];

If[2 ≤ iy ≤ Ny,
vy[[t + 1, iy, 1]] = -vy[[t + 1, iy, 2]];
vy[[t + 1, iy, Nx + 2]] = vy[[t + 1, iy, Nx + 1]];
, Indeterminate];

If[1 ≤ ix ≤ Nx + 1,
vx[[t + 1, 1, ix]] = -vx[[t + 1, 2, ix]];
vx[[t + 1, Ny + 2, ix]] = -vx[[t + 1, Ny + 1, ix]];
, Indeterminate];
];
];
];
];

```



3 D Poiseuille Flow

```
Clear[vx, vy, vz, Fx, Fy, Fz, p, Nx, Ny, Nz, U, ix, iy, iz, t, dt, T, s, A];
ρ = 1.2041; (*density of the fluid (air at 293.15K in this case) *)
L = 10; (*Physical length of the grid*)
a = 1; (*Physical height of the grid*)
b = 1; (*Physical depth of the grid*)
Nx = 12; (*Number of grid points over the length of the grid*)
Ny = 10; (*Number of grid points over the height of the grid*)
Nz = 10; (*Number of grid points over the depth of the grid*)
U = 1; (*Entry velocity of the fluid*)
maxtime = 400; (*The number of time-iterations*)
Rey = 100; (*Reynolds number for the fluid*)
nu =  $\frac{U * a}{Rey}$ ; (*The dynamics viscosity of the fluid*)
dt = 0.02; (*The physical time between time stamps*)
Δx =  $\frac{L}{Nx}$ ; (*The physical distance between grid points in the length direction*)
Δy =  $\frac{a}{Ny}$ ; (*The physical distance between grid points in the height direction*)
Δz =  $\frac{b}{Nz}$ ; (*The physical distance between grid points in the height direction*)

(*The relevant series of matrices are created*)
vx = Table[Table[Table[0, {Ny + 2}, {Nx + 1}], {Nz + 2}], {maxtime}];
vy = Table[Table[Table[0, {Ny + 1}, {Nx + 2}], {Nz + 2}], {maxtime}];
vz = Table[Table[Table[0, {Ny + 2}, {Nx + 2}], {Nz + 1}], {maxtime}];
Fx = Table[Table[Table[0, {Ny + 2}, {Nx + 1}], {Nz + 2}], {maxtime}];
Fy = Table[Table[Table[0, {Ny + 1}, {Nx + 2}], {Nz + 2}], {maxtime}];
Fz = Table[Table[Table[0, {Ny + 2}, {Nx + 2}], {Nz + 1}], {maxtime}];
p = Table[Table[Table[0, {Ny + 2}, {Nx + 2}], {Nz + 2}], {maxtime}];
s = Table[Table[Table[0, {Ny + 2}, {Nx + 2}], {Nz + 2}], {maxtime}];
A = Table[Table[Table[0, {Ny + 2}, {Nx + 2}], {Nz + 2}], {maxtime}];

(*The velocity in the length direction is initialized*)
For[l = Nz + 1, l > 1, l--,
  For[i = Nx + 1, i ≥ 1, i--,
    For[j = Ny + 1, j > 1, j--,
      vx[[1, l, j, i]] = U;
    ];
  ];
];

(*The A-matrix used in the pressure-subroutine with incorporated BC's (A.p=s) .*)
Clear[a, n, k, A, m, w];
A = Table[0, {(Ny) * (Nx) * (Nz)}, {(Ny) * (Nx) * (Nz)}];
k = 0;
n = 0;
For[i = 1, i ≤ (Ny) * (Nx) * (Nz), i++,
  n++;
  If[n == Nx + 1, n = 1, 0];
  k++;
  If[k == Ny * Nx + 1, k = 1, 0];
```

```

For[j = 1, j ≤ (Ny) * (Nx) * (Nz), j++,
A[[i, j]] = -2  $\left( \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta x)^2} + \frac{1}{(\Delta z)^2} \right)$ 
+ If[n == 1,  $\frac{1}{(\Delta x)^2}$ , 0]
+ If[n == Nx,  $\frac{1}{(\Delta x)^2}$ , 0]
+ If[1 ≤ k ≤ (Ny - 1) * Nx, 0,  $\frac{1}{(\Delta y)^2}$ ]
+ If[Nx + 1 ≤ k ≤ Ny * Nx, 0,  $\frac{1}{(\Delta y)^2}$ ]
+ If[1 ≤ i ≤ Nx * Ny,  $\frac{1}{(\Delta z)^2}$ , 0]
+ If[Ny * Nx * Nz - Nx * Ny + 1 ≤ i ≤ Ny * Nx * Nz,  $\frac{1}{(\Delta z)^2}$ , 0] * KroneckerDelta[i, j] +
 $\frac{1}{(\Delta x)^2}$  *
(If[n ≠ Nx, KroneckerDelta[i, j - 1], 0] + If[n ≠ 1, KroneckerDelta[i, j + 1], 0]) +
 $\frac{1}{(\Delta y)^2}$  *
(If[1 ≤ k ≤ (Ny - 1) * Nx, KroneckerDelta[i, j - (Nx)], 0] +
If[Nx + 1 ≤ k ≤ Ny * Nx, KroneckerDelta[i, j + Nx], 0]) +
 $\frac{1}{(\Delta z)^2}$  *
(KroneckerDelta[i, j - (Nx * Ny)] + KroneckerDelta[i, j + Nx * Ny])
];
];
(*The correct solution is only obtained
when the rank of A is equal to the number of rows*)
Clear[n];
n = 0;
For[i = 1, i ≤ (Ny) * (Nx) * (Nz), i++,
n++;
If[Dimensions[A][[1]] ≠ MatrixRank[A],
A = A[[1 ;; (Ny) * (Nx) * (Nz) - n, 1 ;; (Ny) * (Nx) * (Nz) - n]];
, Indeterminate];
];
(*Begin the time loop untill maxtime-1*)
For[t = 1, t < maxtime, t++,
(*calculate the next F_x-vlaue for internal points*)
For[iz = 2, iz ≤ Nz + 1, iz++,
(*iz runs over the depth physical space + dummy boundary*)
For[iy = 2, iy ≤ Ny + 1, iy++,
(*iy runs over the vertical physical space + dummy boundary*)
For[ix = 2, ix ≤ Nx, ix++, (*ix runs over the horizontal
physical space+ dummy boundary*)

Fx[[t, iz, iy, ix]] =

```

```

- $\rho * \left( \frac{vx[[t, iz, iy, ix]] - vx[[t, iz, iy, ix-1]]}{2 * \Delta x} + \frac{1}{4} (vy[[t, iz, iy, ix]] + vy[[t, iz, iy, ix+1]] + vy[[t, iz, iy-1, ix]] + vy[[t, iz, iy-1, ix+1]]) * \frac{vx[[t, iz, iy+1, ix]] - vx[[t, iz, iy-1, ix]]}{2 * \Delta y} + \frac{1}{4} (vz[[t, iz, iy, ix]] + vz[[t, iz, iy+1, ix]] + vz[[t, iz-1, iy, ix]] + vz[[t, iz-1, iy+1, ix]]) * \frac{vx[[t, iz+1, iy, ix]] - vx[[t, iz-1, iy, ix]]}{2 * \Delta z} \right)$ 
+ nu *  $\left( \frac{1}{(\Delta x)^2} (vx[[t, iz, iy, ix+1]] + vx[[t, iz, iy, ix-1]] - 2 vx[[t, iz, iy, ix]]) + \frac{1}{(\Delta y)^2} (vx[[t, iz, iy+1, ix]] + vx[[t, iz, iy-1, ix]] - 2 vx[[t, iz, iy, ix]]) + \frac{1}{(\Delta z)^2} (vx[[t, iz+1, iy, ix]] + vx[[t, iz-1, iy, ix]] - 2 vx[[t, iz, iy, ix]]) \right)$ 
];
];
];
];

(*calculate the next F_y-value for internal points*)
For[iy = 2, iy < Ny, iy++,
  For[ix = 2, ix < Nx + 1, ix++,
    For[iz = 2, iz < Nz + 1, iz++,
      Fy[[t, iz, iy, ix]] = - $\rho * \left( \frac{1}{4} (vx[[t, iz, iy, ix]] + vx[[t, iz, iy, ix-1]] + vx[[t, iz, iy+1, ix]] + vx[[t, iz, iy-1, ix]]) * \frac{vy[[t, iz, iy, ix+1]] - vy[[t, iz, iy, ix-1]]}{2 * \Delta x} + vy[[t, iz, iy, ix]] * \frac{vy[[t, iz, iy+1, ix]] - vy[[t, iz, iy-1, ix]]}{2 * \Delta y} + \frac{1}{4} (vz[[t, iz, iy, ix]] + vz[[t, iz-1, iy, ix]] + vz[[t, iz, iy+1, ix]] + vz[[t, iz-1, iy+1, ix]]) * \frac{vy[[t, iz+1, iy, ix]] - vy[[t, iz-1, iy, ix]]}{2 * \Delta z} \right)$ 
+ nu *  $\left( \frac{1}{(\Delta x)^2} (vy[[t, iz, iy, ix+1]] + vy[[t, iz, iy, ix-1]] - 2 vy[[t, iz, iy, ix]]) + \frac{1}{(\Delta y)^2} (vy[[t, iz, iy+1, ix]] + vy[[t, iz, iy-1, ix]] - 2 vy[[t, iz, iy, ix]]) + \frac{1}{(\Delta z)^2} (vy[[t, iz+1, iy, ix]] + vy[[t, iz-1, iy, ix]] - 2 vy[[t, iz, iy, ix]]) \right)$ 

```



```


$$\frac{vz[[t, iz, iy, ix]] - vz[[t, iz - 1, iy, ix]]}{\Delta z} \Big) \Big);
];
];
];

(*Start the subroutine that determine the pressure corresponding to the source*)
(*First, arrange the s-matrix indices in a vector, S*)
Clear[S, n, i, j, k];
S = Table[0, {Nx * Ny * Nz}];
n = 0;
For[k = 2, k ≤ Nz + 1, k++,
  For[i = 2, i ≤ Ny + 1, i++,
    For[j = 2, j ≤ Nx + 1, j++,
      n++;
      S[[n]] = s[[t, k, i, j]];
    ];
  ];
];
(*Next, remove entries from the end of S corresponding
to the rows removed in A in order to obtain rank(A)=rows(A)*)
Clear[temp];
S = S[[1 ;; Dimensions[A][[1]]]];
(*Solve A.x=S. Inverting such large matrices is computationally too heavy,
so approximate methods must be applied. This can be done via SOR,
but the inbuilt function is much, much faster.*)
temp = LinearSolve[A, S];
(*"temp" is a vector corresponding to S. This is rearranged
into a matrix with pressure measurements, corresponding to s*)
n = 0;
For[k = 1, k ≤ Nz, k++,
  For[i = 1, i ≤ Ny, i++,
    For[j = 1, j ≤ Nx, j++,
      n++;
      If[n ≤ Dimensions[temp][[1]] ∧ temp[[j + (i - 1) * Nx + (k - 1) * Ny * Nx]] > 10-8,
        p[[t, k + 1, i + 1, j + 1]] = temp[[j + (i - 1) * Nx + (k - 1) * Ny * Nx]];
        , Indeterminate];
    ];
  ];
];
(*The boundary terms are set equal to the outer interla points. Without
incorporating the BC's in the A-matrix this would eb wrong*)
For[iz = 2, iz ≤ Nz + 1, iz++,
  For[ix = 2, ix ≤ Nx + 1, ix++,
    p[[t, iz, 1, ix]] = p[[t, iz, 2, ix]];
    p[[t, iz, Ny + 2, ix]] = p[[t, iz, Ny + 1, ix]];
  ];
];
For[iz = 2, iz ≤ Nz + 1, iz++,
  For[iy = 2, iy ≤ Ny + 1, iy++,
    p[[t, iz, iy, 1]] = p[[t, iz, iy, 2]];
    p[[t, iz, iy, Nx + 2]] = p[[t, iz, iy, Nx + 1]];
  ];
];
];$$

```

```

For[ix = 2, ix ≤ Nx + 1, ix++,
  For[iy = 2, iy ≤ Ny + 1, iy++,
    p[[t, 1, iy, ix]] = p[[t, 2, iy, ix]];
    p[[t, Nz + 2, iy, ix]] = p[[t, Nz + 1, iy, ix]];
  ];
];
(*calculate the next v_x-vlaue for internal points*)
For[iz = 1, iz ≤ Nz + 2, iz++,
  For[ix = 1, ix ≤ Nx + 2, ix++,
    For[iy = 1, iy ≤ Ny + 2, iy++,

      If[2 ≤ iz ≤ Nz + 1,
        If[2 ≤ iy ≤ Ny + 1,
          If[2 ≤ ix ≤ Nx,
            vx[[t + 1, iz, iy, ix]] = vx[[t, iz, iy, ix]] +  $\frac{dt}{\rho} * \left( Fx[[t, iz, iy, ix]] - \frac{p[[t, iz, iy, ix + 1]] - p[[t, iz, iy, ix]]}{\Delta x} \right)$ ;
            , Indeterminate];
          , Indeterminate];
        , Indeterminate];
      , Indeterminate];
      (*calculate the next v_y-vlaue for internal points*)
      If[2 ≤ iz ≤ Nz + 1,
        If[2 ≤ iy ≤ Ny,
          If[2 ≤ ix ≤ Nx + 1,
            vy[[t + 1, iz, iy, ix]] = vy[[t, iz, iy, ix]] +  $\frac{dt}{\rho} * \left( Fy[[t, iz, iy, ix]] - \frac{p[[t, iz, iy + 1, ix]] - p[[t, iz, iy, ix]]}{\Delta y} \right)$ ;
            , Indeterminate];
          , Indeterminate];
        , Indeterminate];
      , Indeterminate];
      (*calculate the next v_z-vlaue for internal points*)
      If[2 ≤ iz ≤ Nz,
        If[2 ≤ iy ≤ Ny + 1,
          If[2 ≤ ix ≤ Nx + 1,
            vz[[t + 1, iz, iy, ix]] = vz[[t, iz, iy, ix]] +  $\frac{dt}{\rho} * \left( Fz[[t, iz, iy, ix]] - \frac{p[[t, iz + 1, iy, ix]] - p[[t, iz, iy, ix]]}{\Delta z} \right)$ ;
            , Indeterminate];
          , Indeterminate];
        , Indeterminate];
      , Indeterminate];
      (*Applty BC's for v_x, v_y and v_z*)

      If[2 ≤ iz ≤ Nz + 1,
        If[2 ≤ iy ≤ Ny + 1,
          vx[[t + 1, iz, iy, 1]] = U;
          vx[[t + 1, iz, iy, Nx + 1]] = vx[[t + 1, iz, iy, Nx]];
          , Indeterminate];
        , Indeterminate];
      , Indeterminate];
    ];
];

```

```

If[2 <= iz <= Nz + 1,
  If[1 <= ix <= Nx + 1,
    vx[[t + 1, iz, 1, ix]] = -vx[[t + 1, iz, 2, ix]];
    vx[[t + 1, iz, Ny + 2, ix]] = -vx[[t + 1, iz, Ny + 1, ix]];
    , Indeterminate];
  , Indeterminate];

If[2 <= iy <= Ny + 1,
  If[1 <= ix <= Nx + 1,
    vx[[t + 1, 1, iy, ix]] = -vx[[t + 1, 2, iy, ix]];
    vx[[t + 1, Nz + 2, iy, ix]] = -vx[[t + 1, Nz + 1, iy, ix]];
    , Indeterminate];
  , Indeterminate];

If[2 <= iz <= Nz + 1,
  If[2 <= iy <= Ny,
    vy[[t + 1, iz, iy, 1]] = -vy[[t + 1, iz, iy, 2]];
    vy[[t + 1, iz, iy, Nx + 2]] = vy[[t + 1, iz, iy, Nx + 1]];
    , Indeterminate];
  , Indeterminate];

If[2 <= iz <= Nz + 1,
  If[1 <= ix <= Nx + 2,
    vy[[t + 1, iz, 1, ix]] = 0;
    vy[[t + 1, iz, Ny + 1, ix]] = 0;
    , Indeterminate];
  , Indeterminate];

If[2 <= iy <= Ny + 1,
  If[1 <= ix <= Nx + 2,
    vy[[t + 1, 1, iy, ix]] = -vy[[t + 1, 2, iy, ix]];
    vy[[t + 1, Nz + 2, iy, ix]] = -vy[[t + 1, Nz + 1, iy, ix]];
    , Indeterminate];
  , Indeterminate];

If[2 <= iz <= Nz + 1,
  If[2 <= iy <= Ny,
    vz[[t + 1, iz, iy, 1]] = -vz[[t + 1, iz, iy, 2]];
    vz[[t + 1, iz, iy, Nx + 2]] = vz[[t + 1, iz, iy, 1]];
    , Indeterminate];
  , Indeterminate];

If[2 <= iy <= Ny + 1,
  If[1 <= ix <= Nx + 2,
    vz[[t + 1, 1, iy, ix]] = 0;
    vz[[t + 1, Nz + 1, iy, ix]] = 0;
    , Indeterminate];
  , Indeterminate];

If[1 <= iz <= Ny + 1,
  If[1 <= ix <= Nx + 2,
    vz[[t + 1, iz, 1, ix]] = -vz[[t + 1, iz, 2, ix]];
    vz[[t + 1, iz, Ny + 2, ix]] = -vz[[t + 1, iz, Ny + 1, ix]];
    , Indeterminate];
  , Indeterminate];

```

```
, Indeterminate];  
];  
];  
];  
];
```

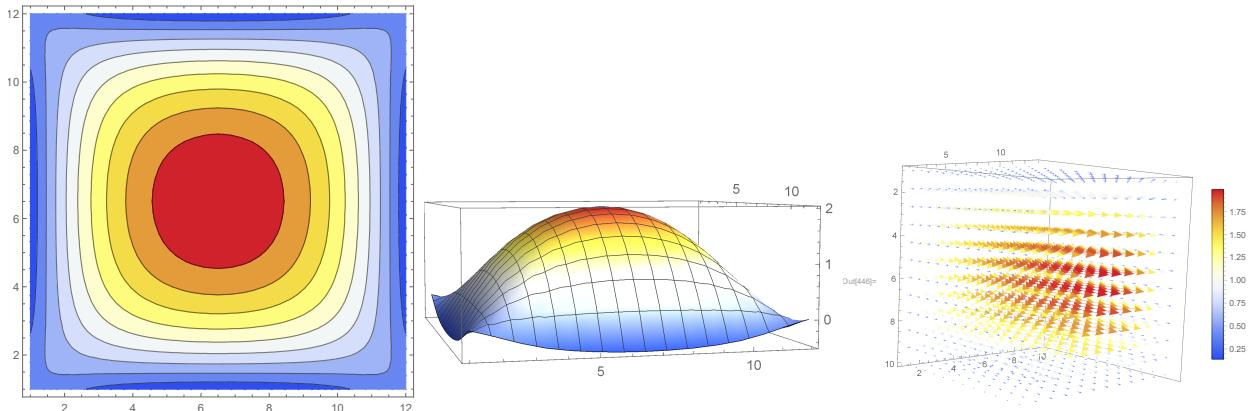
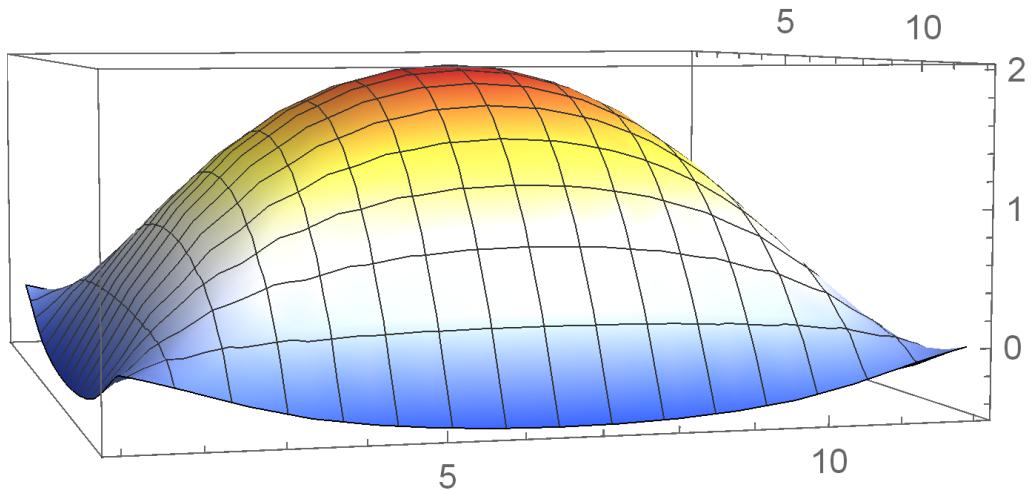


Figure A.1: Left: A contour plot of the 2D slice of the flow velocity in the length direction at about halfway in the flow direction (the x -direction). Middle: A 3D plot of the 2D slice of the flow velocity in the length direction at the end of the channel. The height signify the value of the velocity field in the length direction. Right: A 3D vector plot of the velocity field in the channel.



The Navier-Stokes equations (equation (2.100)) are not solveable in the general case, and so approximations obtained from the physical scenario under consideration must be used in order to apply Navier-Stokes equations.

B | Hamiltonian Monte Carlo

This appendix is taken from Petersen. The Hamiltonian Monte Carlo Algorithm (HMC algorithm) is a Markov Chain Monte Carlo (MCMC) algorithm used to evaluate integrals on the form

$$\begin{aligned}\mathbb{E}[f] &= \int f(\theta)g(\theta)d\theta \\ &\approx \frac{1}{N} \sum_{j \in g} f(\theta_j),\end{aligned}\tag{B.1}$$

with f being a generic function and N denoting the number of samples from the posterior distribution, g . The sample $\{j\}$ from g can be generated via a MCMC algorithm that has g as a stationary distribution. The Markov chain is defined by an initial distribution for the initial state of the chain, θ , and a set of transition probabilities, $p(\theta'|\theta)$, determining the sequential evolution of the chain. A distribution of points in the Markov Chain are said to comprise a stationary distribution if they are drawn from the same distribution and that this distribution persist once established. Hence, if g is the a stationary distribution of the Markov Chain defined by the initial point θ and the transition probability $p(\theta'|\theta)$, then (Neal, 1996)

$$g(\theta') = \int p(\theta'|\theta)g(\theta)d\theta.\tag{B.2}$$

Equation (B.2) is implied by the stronger condition of detailed balance, defined viz

$$p(\theta'|\theta)g(\theta) = p(\theta|\theta')g(\theta').\tag{B.3}$$

A Markov chain is ergodic if it has a unique stationary distribution, called the equilibrium distribution, to which it converge from any initial state. $\{i\}$ can be taken as a sequential subset (discarding the part of the chain before the equilibrium distribution) of a Markov chain that has $g(\theta)$ as its equilibrium distribution. The simplest MCMC algorithm is perhaps the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970). The MH algorithm works by randomly initiating all coefficients for the distribution wanting to be sampled. Then, a loop runs a subjective number of times in which one coefficient at a time is perturbed by a symmetric proposal distribution. A common choice of proposal distribution is the normal distribution with the coefficient value as the mean and a subjectively chosen variance. If $g(\theta') \geq g(\theta)$ the perturbation of the coefficient is accepted, otherwise the perturbation is accepted with probability $\frac{g(\theta')}{g(\theta)}$.

The greatest weaknesses of the MH algorithm is i) a slow approach to the equilibrium distribution, ii) relatively high correlation between samples from the equilibrium distribution and iii) a relatively high rejection rate of states. ii) can be rectified by only accepting every n 'th accepted state, with n being some subjective number. For $n \rightarrow \infty$ the correlation naturally disappears, so there is a trade off between efficiency and correlation. Hence, in the end the weaknesses of the MH algorithm can be boiled down to inefficiency. This weakness is remedied by the HCM algorithm (Duane et al., 1987) in which Hamiltonian dynamics are used to generate proposed states in the Markov chain and thus guide the journey in parameter space. Hamiltonian dynamics are useful for proposing states because (Neal, 2012) 1) the dynamics are reversible, implying that detailed balance is fulfilled and so there exist a stationary distribution, 2) the Hamiltonian (H) is conserved during the dynamics if there is no explicit time dependence in the Hamiltonian ($\frac{dH}{dt} = \frac{\partial H}{\partial t}$), resulting in all proposed states being accepted in the case the dynamics are exact and 3) Hamiltonian dynamics preserve

the volume in phase space (q_i, p_i -space), which means that the Jacobian is unity (relevant for Metropolis updates that succeeds the Hamiltonian dynamics in the algorithm). By making sure the algorithm travel (in parameter space) a longer distance between proposed states, the proposed states can be ensured to have very low correlation, hence alleviating issues 1) and 2) of the MH algorithm. The price to pay for using the HMC algorithm relative to the MH algorithm is a) the HMC algorithm is gradient based meaning it requires the Hamiltonian to be continuous and b) the computation time can be long depending on the distribution being sampled (e.g. some recurrent ANNs are computationally heavy due to extensive gradient calculations).

As previously stated, the HMC algorithm works by drawing a physical analogy and using Hamiltonian dynamics to generate proposed states and thus guide the journey in parameter space. The analogy consists in viewing g as the canonical probability distribution describing the probability of a given configuration of parameters. In doing so, g is related to the Hamiltonian, H , viz

$$g = e^{\frac{F-H}{k_B T}} \Rightarrow H = F - k_B T \ln[g], \quad (\text{B.4})$$

where $F = -k_B T \ln[Z]$ denotes Helmholtz free energy of the (fictitious in this case) physical system and Z is the partition function. $\ln[g(\theta)]$ contain the position (by analogy) variables of the Hamiltonian and so Z must contain the momentum variables. Almost exclusively (Betancourt and Girolami, 2013) $Z \sim \mathcal{N}(0, \sqrt{m_i})$ is taken yielding the Hamiltonian

$$H = -k_B T \left[\ln[g] - \sum_i \frac{p_i^2}{2m_i} \right] + \text{const}, \quad (\text{B.5})$$

where i run over the number of variables and "const" is an additive constant (up to which the Hamiltonian is always defined). $T = k_b^{-1}$ is most often taken (Neal, 2012), however, the temperature can be used to manipulate the range of states which can be accepted e.g. via simulated annealing (MacKay, 2002). Here $T = k_b^{-1}$ will be adopted in accordance with (Neal, 1996, 2012) and as such

$$H = \sum_i \frac{p_i^2}{2m_i} - \ln[g]. \quad (\text{B.6})$$

The dynamics in parameter space are determined by Hamiltons equations

$$\dot{\theta}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial \theta_i}, \quad (\text{B.7})$$

with θ_i denoting the different variables (coefficients). In order to implement Hamiltons equations, they are discretized via the leap frog method (Neal, 1996, 2012) viz

$$\begin{aligned} p_i \left(t + \frac{\epsilon}{2} \right) &= p_i(t) - \frac{\epsilon}{2} \frac{\partial H(\theta_i(t), p_i(t))}{\partial \theta_i}, \\ \theta_i(t + \epsilon) &= \theta_i(t) + \frac{\epsilon}{m_i} p_i \left(t + \frac{\epsilon}{2} \right), \\ p_i(t + \epsilon) &= p_i \left(t + \frac{\epsilon}{2} \right) - \frac{\epsilon}{2} \frac{\partial H(\theta_i(t + \frac{\epsilon}{2}), p_i(t + \frac{\epsilon}{2}))}{\partial \theta_i}, \end{aligned} \quad (\text{B.8})$$

with ϵ being an infinitesimal parameter. In the algorithm the initial state is defined by a random initialization of coordinates and momenta, yielding H_{initial} . Subsequently Hamiltonian dynamics are simulated a subjective (L loops) amount of time resulting in a final state, H_{final} , the coordinates of which take the role of proposal state. The loop that performs L steps of ϵ in time is here referred to as the dive. During the dive, the Hamiltonian remains constant, so ideally $H_{\text{initial}} = H_{\text{final}}$, however, imperfections in the discretization procedure of the dynamics can result in deviations from this equality (for larger values of ϵ , as will be discussed further later on). For this reason, the proposed state is accepted as the next state in the Markov chain with probability

$$\mathbb{P}(\text{transition}) = \min [1, e^{H_{\text{initial}} - H_{\text{final}}}] . \quad (\text{B.9})$$

Whether or not the proposed state is accepted, a new proposed state is next generated via Hamiltonian dynamics and so the loop goes on for a subjective amount of time.

Most often, the HMC algorithm will be ergodic, meaning it will converge to its unique stationary distribution from any given initialization (i.e. the algorithm will not be trapped in some subspace of parameter space), however, this may not be so for a periodic Hamiltonian if $L\epsilon$ equal the periodicity. This potential problem can however be avoided by randomly choosing L and ϵ from small intervals for each iteration. The intervals are in the end subjective, however, with some constraints and rules of thumb; the leap frog method has an error of $\mathcal{O}(\epsilon^2)$ (Neal, 1996) and so the error can be controlled by ensuring that $\epsilon \ll 1$. A too small value of ϵ will waste computation time as a correspondingly larger number of iterations in the dive (L) must be used to obtain a large enough trajectory length $L\epsilon$. If the trajectory length is too short the parameter space will be slowly explored by a random walk instead of the otherwise approximately independent sampling (the advantage of non-random walks in HMC is a more uncorrelated Markov chain and better sampling of the parameter space). A rule of thumb for the choice of ϵ can be derived from a one dimensional Gaussian Hamiltonian

$$H = \frac{q^2}{2\sigma^2} + \frac{p^2}{2}. \quad (\text{B.10})$$

The leap frog step for this system is a linear map from $t \rightarrow t + \epsilon$. The mapping can be written

$$\begin{bmatrix} q(t + \epsilon) \\ p(t + \epsilon) \end{bmatrix} = \begin{bmatrix} 1 - \frac{\epsilon^2}{2\sigma^2} & \epsilon \\ \epsilon(\frac{1}{4}\epsilon^2\sigma^{-4} - \sigma^{-2}) & 1 - \frac{1}{2}\epsilon^2\sigma^{-2} \end{bmatrix} \begin{bmatrix} q(t) \\ p(t) \end{bmatrix} \quad (\text{B.11})$$

The eigenvalues of the coefficient matrix represent the powers of the exponentials that are the solutions to the differential equation. They are given by

$$\text{Eigenvalues} = 1 - \frac{1}{2}\epsilon^2\sigma^{-2} \pm \epsilon\sigma^{-1}\sqrt{\frac{1}{4}\epsilon^2\sigma^{-2} - 1}. \quad (\text{B.12})$$

In order for the solutions to be bounded, the eigenvalues must be imaginary, meaning that

$$\epsilon < 2\sigma. \quad (\text{B.13})$$

In higher dimensions a rule of thumb is to take $\epsilon \lesssim 2\sigma_x$, where σ_x is the standard deviation in the most constrained direction, i.e. the square root of the smallest eigenvalue of the covariance matrix. In general (Betancourt and Girolami, 2013) a stable solution with $\frac{1}{2}p^T\Sigma^{-1}p$ as the kinetic term in the Hamiltonian require

$$\epsilon_i < 2\lambda_i^{-\frac{1}{2}}, \quad (\text{B.14})$$

for each eigenvalue λ_i of the matrix

$$M_{ij} = (\Sigma^{-1})_{ij} \frac{\partial^2 H}{\partial q_i \partial q_j}, \quad (\text{B.15})$$

which means that in the case of $\Sigma^{-1} = \text{diag}(m_i^{-1})$;

$$\epsilon_i < 2 \sqrt{\frac{m_i}{\frac{\partial^2 H}{\partial q_i^2}}}. \quad (\text{B.16})$$

Setting ϵ according to equation (B.14) can however introduce issues for hierarchical models (models including hyper parameters) since the reversibility property of Hamiltonian dynamics is broken if ϵ depend on any parameters. This issue can be alleviated by using the MH algorithm on a subgroup of parameters (Neal, 1996, 2012) (which are then allowed in the expression for ϵ) that is to be included in ϵ . However, unless the MH algorithm is used for all parameters, some degree of approximation is required. The HMC algorithm is shown in algorithm ??.

```

Save:  $q$  and  $V(q)$ , with  $q$  randomly initialized;
for  $i < N+1$  do
     $p =$  Sampled from standard normal distribution;
     $H_{old} = H(q, p);$ 
     $p = p - \frac{\epsilon}{2} \frac{\partial H(q, p)}{\partial q};$ 
     $L = RandomInteger(L_{lower}, L_{upper});$ 
    for  $j < L+1$  do
         $q = q + \epsilon \frac{p}{mass};$ 
        if  $j \neq L$  then
             $p = p - \epsilon \frac{\partial H(q, p)}{\partial q};$ 
        end
    end
     $p = p - \frac{\epsilon}{2} \frac{\partial H(q, p)}{\partial q};$ 
     $H_{new} = H(q, p);$ 
     $u =$  Sample from uniform distribution;
    if  $u < \min(1, e^{-(H_{new} - H_{old})})$  then
         $H_{old} = H_{new};$ 
        Save:  $q$  and  $V(q);$ 
    end
end

```

Algorithm 1: Hamiltonian Monte Carlo Algorithm in pseudo code

C | Nested Sampling

A major challenge in estimating the evidence via conventional Monte Carlo Methods is that generally the prior is a very broad and regular distribution whereas the likelihood is a very narrow and irregular distribution. This poses a challenge when the evidence is estimated conventionally, i.e. as the mean of the likelihood evaluated at points in parameter space corresponding to samples from the prior distribution. For a reasonable number of samples, the conventional procedure has a relatively high likelihood of relatively poor sampling in regions near the peaks in the likelihood distribution. This means a conventional estimate of the evidence via Monte Carlo Methods has a high variance. Nested Sampling (Skilling, 2004) (NS) address this challenge by accounting for the likelihood distribution when sampling the prior distribution. Consider the integral

$$Z = \int L(\theta)\pi(\theta)d\theta, \quad (\text{C.1})$$

with L being the likelihood distribution and π the prior distribution. Conventional Monte Carlo methods approximate this integral via importance sampling, meaning

$$\begin{aligned} Z &= \mathbb{E}_\pi[L] \\ &\approx \frac{1}{N} \sum_{i \in \pi} L(\theta_i), \end{aligned} \quad (\text{C.2})$$

where the second equality become exact for $N \rightarrow \infty$. NS project the integral down into one dimension viz¹

$$\begin{aligned} Z &= \int_0^1 L(\xi)d\xi \\ &\approx \sum_i L(\xi_i)\Delta\xi_i, \end{aligned} \quad (\text{C.3})$$

where

$$\xi(\lambda) = \int_{L>\lambda} \pi(\theta)d\theta, \quad (\text{C.4})$$

is the proportion of the prior with likelihood greater than λ and $\Delta\xi_i \equiv \xi_{i-1} - \xi_i$. Due to the constraint $L > \lambda$ on the integral bound of ξ , $L(\xi)$ is a decreasing function of ξ , meaning $L(\xi_1) > L(\xi_2)$ if $\xi_1 < \xi_2$. The sum in equation (C.3) can then be evaluated by generating a sequence

$$\{\{L(\xi_m), \xi_m\}, \{L(\xi_{m-1}), \xi_{m-1}\}, \dots, \{L(\xi_1), \xi_1\}\}, \quad (\text{C.5})$$

with $\xi_1 < \xi_2 < \dots < \xi_m$. The sorting operation eliminate coordinate dependent complications of geometry, topology and dimensionality (Skilling, 2006). A sequence upholding equation (C.5) can be generated as follows; consider n random draws from g with corresponding values of L and ξ . Let $L(\xi^*)$ denote the minimum value of L in the sample with ξ^* the corresponding value of ξ in the sample. $\{L(\xi^*), \xi^*\}$ is replaced by another set which is sampled from g with the constraint that $\xi_{new} < \xi^*$ and stored in a list of discarded states. Continuing this sequence again and again will fill the list of discarded states that uphold equation

¹ Attempting a higher accuracy via better numerical approximations of the integral is mute since the uncertainty in ξ dominate the approximation (Skilling, 2004).

(C.5). In practice $L(\xi)$ is not readily available, so instead L can be generated from values of θ . The value of ξ_k can be determined by using that (Skilling, 2004)

$$\xi_k = \xi_0 \prod_{i=1}^k t_i, \quad (\text{C.6})$$

with $t_i = \frac{\xi_k}{\xi_{k-1}}$, called the shrinkage ratio. The shrinkage ratio follow a beta distribution

$$p(t) = nt^{n-1}, \quad (\text{C.7})$$

with n being the number of initial samples from g (the number of live points), such that

$$\begin{aligned} \langle \ln(t) \rangle &= \mathbb{E}[\ln(t)] \pm \sqrt{V[\ln(t)]} \\ &= \int_0^1 nt^{n-1} \ln(t) dt \pm I_2, \\ &= \frac{1}{n}(-1 \pm 1) \end{aligned} \quad (\text{C.8})$$

with

$$I_2 = \sqrt{\int_0^1 nt^{n-1} \ln(t)^2 dt - \left(\int_0^1 nt^{n-1} \ln(t) dt \right)^2}. \quad (\text{C.9})$$

Using $\ln(\xi_k) = \sum_{i=1}^k \ln(t_i)$ and taking t_i to be i.i.d. yield

$$\begin{aligned} \langle \ln(\xi_k) \rangle &= k \mathbb{E}[\ln(t)] \pm \sqrt{k V[\ln(t)]} \\ &= \frac{1}{n}(-k \pm \sqrt{k}) \end{aligned} \quad (\text{C.10})$$

Ignoring uncertainty ξ_k can be approximated by the mean viz

$$\xi_k \approx e^{-\frac{k}{n}}, \quad (\text{C.11})$$

meaning

$$\Delta \xi_i \approx e^{-\frac{i}{n}} (e^{\frac{1}{n}} - 1). \quad (\text{C.12})$$

A heuristic measure for terminating the collection of samples is to require that the maximum likelihood collected make up only a small fraction, B , of the evidence, meaning

$$\max(\{L\}) \xi_j < BZ, \quad (\text{C.13})$$

for iteration j . Another approach to terminating the collection of samples is to use that most of the area in the $L\xi$ -plane is usually found in the region (Skilling, 2004, 2006) $\xi \sim e^{-\mathcal{H}} \sim e^{-\frac{i}{n}}$, meaning the collection of samples can be terminated when

$$i \gg n\mathcal{H}, \quad (\text{C.14})$$

with \mathcal{H} being the information (Skilling, 2004)

$$\begin{aligned} \mathcal{H} &= \int \frac{L(\xi)}{Z} \ln \left(\frac{L(\xi)}{Z} \right) d\xi \\ &\approx \sum_i \frac{L(\xi_i)}{Z} \ln \left(\frac{L(\xi_i)}{Z} \right) \Delta \xi_i. \end{aligned} \quad (\text{C.15})$$

Terminating at $i \sim n\mathcal{H}$ yield (equation (C.10)) an uncertainty $\delta(\langle \ln(\xi_i) \rangle) = \sqrt{\frac{\mathcal{H}}{n}}$ meaning

$$\ln(Z) \approx \ln \left(\sum_i L(\xi_i) \Delta \xi_i \right) \pm \sqrt{\frac{\mathcal{H}}{n}}. \quad (\text{C.16})$$

The NS algorithm with equation (C.14) as termination criterion is shown in algorithm ???. A and B are parameters of the algorithm. The "Remainder" in the second to last line in algorithm ?? fills in the missing band $0 < \xi < e^{-\frac{k+1}{n}}$ with the average value of the remaining values of L . Due to the chosen stopping criterion, the "Remainder" will be construction be small.

Import: $S = 'n$ samples $\theta_1, \theta_2, \dots, \theta_n$ from the prior distribution' with L being the corresponding likelihoods;

Initialize: $k = 0, a = 0, B = 1, Z = \text{Empty list};$

while $f > B$ **do**

Let $L^* \equiv \min(L)$ and $S^* \hat{=} L^*$;

$S2 = S \setminus S^* \wedge L2 = L \setminus L^*$;

Define $\Delta\xi_k = e^{\frac{k+1}{n}}(e^{\frac{1}{n}} - 1)$;

Store $L^* \Delta\xi_k$ in Z ;

$S_{new}, L_{new} = \text{proposer}(\text{random}(S2), L^*)$;

$S = S2 \cup S_{new} \wedge L = L2 \cup L_{new}$;

$f = \frac{\max(L)e^{-\frac{k+1}{n}}}{\sum_{s=0}^k Z_s}$;

if $a == A$ **then**

Display status, e.g. $f, n\mathcal{H} - k, k, \sum_{s=0}^k Z_s, \dots$;

$a = 0$;

end

$k = k + 1$;

$a = a + 1$;

end

$\text{Remainder} = \frac{1}{n} \sum_i L_i e^{-\frac{k+1}{n}}$;

$Z \approx \sum_{s=0}^k Z_s + \text{Remainder}$;

Algorithm 2: Nested Sampling Algorithm in pseudo code

Bibliography

- R.A. Adams and C. Essex. *Calculus: A Complete Course*. Pearson Canada, 2009. ISBN 9780321549280.
URL <https://books.google.dk/books?id=vFdKPgAACAAJ>.
- M. Betancourt and Mark Girolami. Hamiltonian monte carlo for hierarchical models. 12 2013. doi: 10.1201/b18502-5.
- S. Blundell, S.J. Blundell, and K.M. Blundell. *Concepts in Thermal Physics*. Comprehensive Assessment of Water Management in Agriculture. Oxford University Press, 2006. ISBN 9780198567691. URL <https://books.google.dk/books?id=oaDdsxEACAAJ>.
- Ta-Pei Cheng. *Relativity, Gravitation and Cosmology: A Basic Introduction*. Oxford University Press, 2010.
- P. A. M. Dirac. *Die Prinzipien der Quantenmechanik. (German) [The Principles of Quantum Mechanics]*. Verlag von S. Hirzel, Leipzig, Germany, 1930. Translated by Werner Bloch from the English original.
- S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195:216–222, 1987. doi: 10.1016/0370-2693(87)91197-X.
- H. Goldstein, C.P. Poole, and J.L. Safko. *Classical Mechanics*. Addison Wesley, 2002. ISBN 9780201657029.
URL <https://books.google.dk/books?id=tJCuQgAACAAJ>.
- David J. Griffiths. *Introduction to Quantum Mechanics (2nd Edition)*. Pearson Prentice Hall, 2nd edition, April 2004. ISBN 0131118927.
- W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi: 10.1093/biomet/57.1.97.
- J. M. Knudsen and Poul G. Hjorth. *Elements of Newtonian Mechanics, 3rd edition*. Springer, 2000.
- Tom Lancaster and Stephen J Blundell. *Quantum field theory for the gifted amateur*. Oxford University Press, Oxford, Apr 2014. doi: 0199699321. URL <https://cds.cern.ch/record/1629337>.
- B. Lautrup. *Physics of Continuous Matter: Exotic and Everyday Phenomena in the Macroscopic World*. CRC Press, 2011. ISBN 9781439894200. URL <https://books.google.dk/books?id=ohbSBQAAQBAJ>.
- David J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2002.
- B.R. Martin. *Nuclear and Particle Physics: An Introduction*. Wiley, 2006. ISBN 9780470025321. URL <https://books.google.it/books?id=WXHG1wSgQDMC>.
- Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi: 10.1063/1.1699114. URL <http://link.aip.org/link/?JCP/21/1087/1>.
- Radford M. Neal. Springer-Verlag, Berlin, Heidelberg, 1996. ISBN 0387947248.

- Radford M. Neal. Mcmc using hamiltonian dynamics. 2012. URL <http://arxiv.org/abs/1206.1901>. cite arxiv:1206.1901.
- Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Westview, Boulder, CO, 1995. URL <https://cds.cern.ch/record/257493>. Includes exercises.
- J. Petersen. *The Missing Mass Problem on Galactic Scales*. PhD thesis.
- Jun John Sakurai. *Modern quantum mechanics; rev. ed.* Addison-Wesley, Reading, MA, 1994. URL <https://cds.cern.ch/record/1167961>.
- F. Sannino. Lagrangian and hamiltonian mechanics.
- J. Schwichtenberg. *Physics from Symmetry*. Springer International Publishing, 1 edition, 2015.
- John Skilling. Nested Sampling. In Rainer Fischer, Roland Preuss, and Udo Von Toussaint, editors, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 735 of *American Institute of Physics Conference Series*, pages 395–405, November 2004. doi: 10.1063/1.1835238.
- John Skilling. Skilling, j.: Nested sampling for general bayesian computation. *bayesian anal.* 1(4), 833-860. *Bayesian Analysis*, 1:833–860, 12 2006. doi: 10.1214/06-BA127.
- Mark Srednicki. *Quantum Field Theory*. Cambridge Univ. Press, Cambridge, 2007. URL <https://cds.cern.ch/record/1019751>.
- Steven Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley, New York, NY, 1972. URL <https://cds.cern.ch/record/100595>.

Index

- 4-vector, 129
Action, classical, 31
Add. of angular mom., ladder method, 81
Angular momentum theorem, 16

Canonical quantization, 64
Canonical transformation, 37
Center of mass theorem, 13
Central force fields, 13
Clebsch-Gordon coefficients, 83
Commutator identities, 100
Commutator, definition, 65
Completeness relation, 55
Configuration space, 30
Conjugate momentum, 33
Conservation of angular momentum, 16
Conservation of linear momentum, 13
Conservative force fields, 13
Conserved charge, 147
Conserved current, 147
Constants of motion, 36
Current density, 129

Dirac equation, 142, 146
Dirac field, 138
Dirac matrices, 1

Eigenstates, simultaneous, 65
Eigenvalue, 54
Eigenvalue prob., intuitive understanding, 60
Eigenvalue problem, matrix notation, 59, 68
Eigenvector, 54
Einstein equation, 156
Energy-momentum tensor, 146
Equivalence principle, 149
Euler-Lagrange equation, 32
Expectation value in TDPT, 107

Fermi Lagrangian, 203
First order TDPT, 113
Four current, 129
Fourier transform, 4

Fourier transform, delta function, 166
Gauge invariance, 139
Generating functional, 171
Generators of a continuous group, 124
Generators of a discrete group, 123
Geodesic equation, 150, 151, 154
Grassmann fields, 173
Grassmann numbers, 173

Hamilton-Jacobi equation, 39
Hamiltonian in SR, 140
Hamiltonian on covariant form, 140
Hamiltonian, definition, 36
Hamilton's principle, 31
Heisenberg's equation of motion, 70
Hermitian adjoint, 52, 58
Hermitian operator, 53
Hilbert space, 51
Hydrogen, 100, 102

invariant subspace, 123
Irreducible representation, 123

KG equation coupled to EM field, 146
Kinematics of $1 \Rightarrow 2$ decay, 187
Kinematics of massive scattering, 184
Kinematics of massless scattering, 186
Klein-Gordon (KG) equation, 146
Klein-Gordon equation, 139, 147

Lagrangian in SR, 140
Lagrangian on covariant form, 140
Lagrangian, definition, 31
Legendre transformation, 35
Lie group, 62, 121
Lorentz covariance, 133

Manifold, 121

Natural units, 1
Navier-Stokes equation, 24
Newtonian limit (GR), 151
Newton's laws, 9

Noether current, 146
Noethers theorem, 146
Noethers theorem, classical field theory, 144
Noethers theorem, classical mechanics, 34
Non-inertial reference frames, 11

Outcome of an experiment, QM, 53

Planetary physics in Newtonian limit, 16, 21
Poisson bracket, 36
Principle of general covariance, 151
Principle of least action, 31

Random variable, 5

Scattering, hadron, 203
Scattering, meson, 203
Selection rules, 113, 114
Similarity transformation, 123
Simultaneous eigenkets, 65
Singlet state, 76
Spherical tensor, 100
Spin half system, 73
Standard cosmology, 158
State vector, 51
State vector in TDPT, 107
Symmetry operators, 62

TDPT, 90
TDPT, hydrogen atom, 113
TDPT, SHO, 107–109
TDPT, two-state system, 110, 111
TDSE, 67
Tensor notation in SR, 140
Tensor operator, 100
The Lorentz group, 127
Time evolution of state vector, 68
TIPT, deg., 88, 94, 96, 98
TIPT, free particle in a box, 96, 97
TIPT, non-deg., 85, 96, 98
TIPT, SHO, 94, 95, 98
TIPT, two-state system, 103
TIPT, uniform electric field, 102
TISE, 67
Transformation matrix, 58
Transition probability, 108, 109, 113
Triplet state, 76
Two-electron system, 76, 100

Uniform magnetic field, 70

Vector operator, 100

Weak gravitational field limit, 151
Wicks theorem, 193, 199, 200