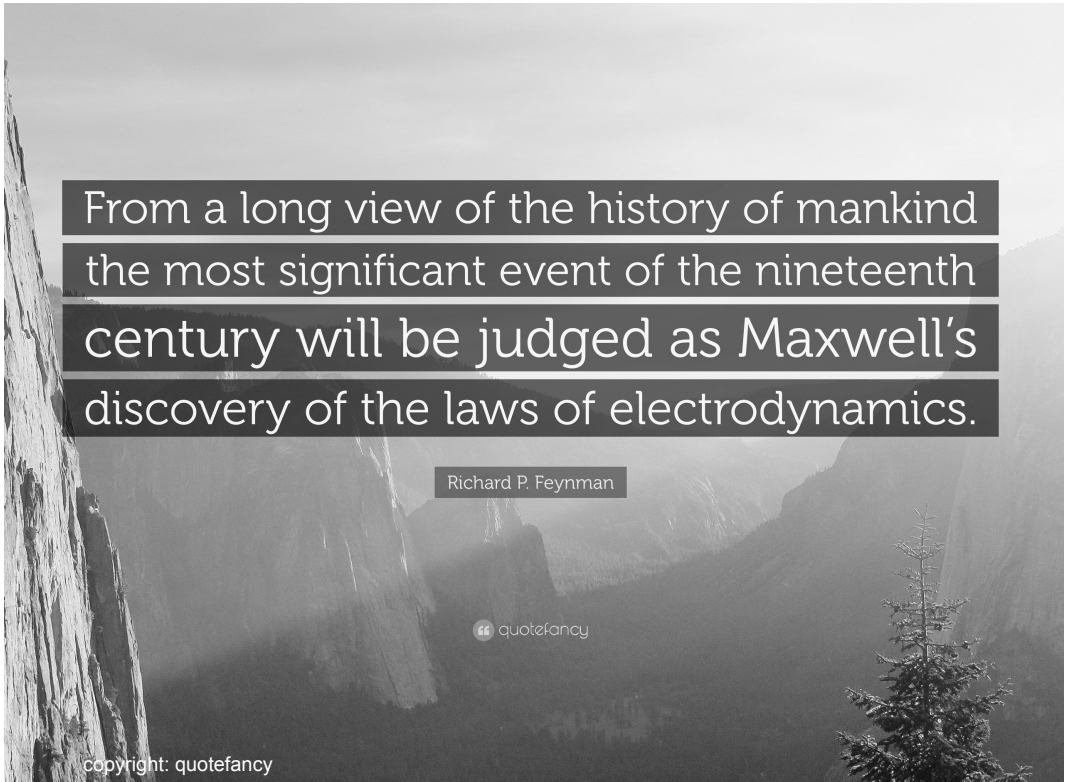


TP3: Classical Field Theory

A black and white photograph of a mountain valley, likely Yosemite, with a quote overlaid. The quote is: "From a long view of the history of mankind the most significant event of the nineteenth century will be judged as Maxwell's discovery of the laws of electrodynamics." The quote is presented in four lines, each within a dark rectangular background. The background image shows a steep rock face on the left, a valley with a river or stream in the center, and a large evergreen tree on the right. The sky is overcast.

From a long view of the history of mankind
the most significant event of the nineteenth
century will be judged as Maxwell's
discovery of the laws of electrodynamics.

Richard P. Feynman

quoteFancy

copyright: quoteFancy

Alexander Altland

Contents

1	Classical field theory	3
2	Introduction to Electrodynamics	5
2.1	History	5
2.1.1	Ancient world and Middle Ages	5
2.1.2	Modern times and Age of Enlightenment	5
2.1.3	Romantic era till late nineteenth century	6
2.1.4	Twentieth century	7
2.2	The apparatus of electrodynamics	7
2.2.1	Basic quantities	7
2.2.2	Interaction of matter and fields	8
2.2.3	Maxwell equations: generalities	10
2.2.4	Charges and currents	12
2.2.5	Integral form of the Maxwell equations	18
3	Electromagnetostatics	25
3.1	Decoupling of the Maxwell equations in the static limit	25
3.2	Electrostatics	26
3.2.1	Formulation of the problem	26
3.2.2	Boundary value problem I: infinite space	28
3.2.3	Poisson equation in non-cartesian coordinates	38
3.2.4	Spherical multipole Expansion	46
3.2.5	Boundary value problem II: finite space	48
3.3	Magnetostatics	53
3.3.1	Vector potential and solution of Maxwell equations	53
3.3.2	Magnetic multipole expansion	55
3.3.3	Magnetic forces	58
4	Electrodynamics	61
4.1	Magnetic field energy	61
4.2	Electromagnetic gauge field	63
4.3	Electromagnetic waves in vacuum	65
4.3.1	Solution of the homogeneous wave equations	65

4.3.2	Polarization	67
4.4	Green function of the wave equation	69
4.4.1	Computation of the Green function	69
4.4.2	Physical meaning of the Green function	72
4.4.3	Electromagnetic gauge field and Green functions	73
4.5	Field energy and momentum	74
4.5.1	Field energy	74
4.5.2	Field momentum	75
4.5.3	Energy and momentum of plane electromagnetic waves	77
4.6	Electromagnetic radiation	78
5	Macroscopic electrodynamics	83
5.1	Macroscopic Maxwell equations	83
5.1.1	Averaged sources	84
5.2	Applications of macroscopic electrodynamics	90
5.2.1	Electrostatics in the presence of matter	90
5.2.2	Magnetostatics in the presence of matter	93
5.3	Wave propagation in media	94
5.3.1	Model dielectric function	95
5.3.2	Plane waves in matter	97
5.3.3	Dispersion	98
5.3.4	Electric conductivity	101
6	Relativistic invariance	105
6.1	Introduction	105
6.1.1	Galilei invariance and its limitations	105
6.1.2	Einsteins postulates	107
6.1.3	First consequences	108
6.2	The mathematics of special relativity I: Lorentz group	110
6.2.1	Background	110
6.3	Aspects of relativistic dynamics	115
6.3.1	Proper time	115
6.3.2	Relativistic mechanics	116
6.4	Mathematics of special relativity II: Co- and Contravariance	119
6.4.1	Covariant and Contravariant vectors	120
6.5	Relativistic covariance of electrodynamics	122
6.5.1	Covariant formulation of fields and currents	122
6.5.2	Invariant formulation of Maxwells equations	124
6.5.3	Electrodynamics from a variational principle	124
6.5.4	Towards Quantum Electrodynamics	127
6.6	A glimpse of general relativity	130
6.6.1	Strong equivalence principle	130
6.6.2	The bending of light and gravitational redshift	131

6.6.3	The idea behind general relativity	132
6.6.4	Geodesic motion	134
6.6.5	Newtonian limit of gravitational dynamics	135
7	Hydrodynamics	137
7.1	The concept of a fluid particle	137
7.2	Evolution of integral quantities carried by a fluid	138
7.3	Conservation laws	139
7.3.1	Mass conservation	139
7.3.2	Stress tensor	141
7.3.3	Momentum conservation	142
7.4	Navier-Stokes equation: two simple examples	143
7.4.1	Flow of ideal fluid around a spherical obstacle	144
7.4.2	Steady viscous flow through cylindrical pipe	145
8	Appendix	147
8.1	Vector calculus in curvilinear coordinates	147
8.1.1	Curvilinear orthogonal systems	147
8.1.2	Vector differential operations in curvilinear coordinates	147

1.0 Classical field theory

Field theory addresses the physical properties of systems containing effectively infinitely many degrees of freedom. Think of the water molecules combined to form the flow of a river. From a fundamentalist's perspective, there is nothing special about such situations: individual degrees of freedom being described by classical point mechanics, or quantum mechanics, one is to solve the corresponding degrees of freedom to obtain the dynamics of the system at large. Of course, this view ignores the fact that the solution of equations of motion beyond $\mathcal{O}(10^{11})$ degrees of freedom is impossible, even on modern computers. What is more, we aren't even interested in the dynamics of individual microscopic system constituents. Rather, we wish to predict physically observable behavior, such as the current flow of a river.

Field theory, instead, approaches the problem in terms of a pragmatic three stage work program: first, identify *effective* degrees of freedom, subsuming large numbers of microscopic degrees of freedom in terms of collective variables. For example, in the case of our river these might be the local and time dependent velocities $\mathbf{v}(\mathbf{r}, t)$ of small volumes of water at a given coordinate and time. Second, we are to transcribe the fundamental equations of motion of classical or quantum mechanics into effective equations of motion for the collective variables. And, third, we need to solve them. Note that the number of degrees of freedom, e.g. the collection of all vectors $\{\mathbf{v}(\mathbf{r}, t)\}$, remains infinitely large.

There are cases where the nature of the microscopic degrees of freedom is not known, or where they may not even exist. A case in point is classical electrodynamics, which predicts forces exerted on test charges or currents through electric or magnetic *fields*. In this case, the degrees of freedom are vector valued variables such as $\mathbf{E}(\mathbf{r}, t)$, where $q\mathbf{E}(\mathbf{r}, t)$ is defined to be the space and time dependent force acting on a test particle of electric charge q . In this case $\mathbf{E}(\mathbf{r}, t)$ are the effective degrees of freedom, and no reference to underlying microscopic variables is made. (In fact, their identity cannot be determined within classical electrodynamics, one needs to 'quantize' the theory, to get closer to a microscopic level.)

All field theories have in common that they describe their effective degrees of freedom as mathematical *fields*, i.e. maps $U \rightarrow T, (\mathbf{r}, t) \mapsto \phi(\mathbf{r}, t)$, where U is a subset of space-time, i.e. a domain that can be parameterized in terms of spatial vectors \mathbf{r} and a time variable, t . Concerning the target space, T , there is huge freedom. From real numbers (think of a temperature field $T(\mathbf{r}, t)$) over vectors (a velocity field $\mathbf{v}(\mathbf{r}, t)$), to matrices (the rotation matrix $R(\mathbf{r}, t)$ specifying the orientation of a magnetization field in a ferromagnet).

There are a few large families of classical field theories every student of physics should at least have heard about. These are

- Classical electromagnetism.
- General relativity, i.e. the theory describing the geometric structure of space-time in terms of a space-time dependent metric $g(\mathbf{r}, t)$.
- Fluid dynamics, i.e. the theory predicting the flow of liquid matter.

Among these, electromagnetism is generally taught first: it is the simplest (by far), and it naturally introduces a plethora of concepts relevant to field theory in general. Following this trusted principle, about 80% of this course will focus on electromagnetism. The remaining 20% will be a short introduction to fluid dynamics, a teaser of general relativity, and an outlook on the quantization of electromagnetism, which serves as a gateway into *quantum* field theory.

2.0 Introduction to Electrodynamics

2.1 History

We begin with a brief chronological account of the development of electrodynamics:

2.1.1 Ancient world and Middle Ages

Ancient world: Unlike with mechanics/astronomy, there is no 'theory' of electromagnetism. Known phenomena include: the attraction of wool and other types of tissue by rubbed amber (greek → electron) and the attraction of metals such as iron by magnetite ('magnesia' → city in ancient Turkey).

Middle ages: Petrus Peregrinus (Pierre der Maricourt), 1269: applies experimental methods (!) to measure the forces exerted by a spherical magnet. Coins the term **poles** of a magnet.

2.1.2 Modern times and Age of Enlightenment

In his work 'de magnete', 1600, **William Gilbert**, 1544–1603 presents a thorough analysis of **magnetic phenomena**. Topics covered in his work include the magnetic nature of the earth, the character of the forces between magnets (→ torsion), a theoretical understanding of compasses, and the fundamental differences between electric and magnetic forces (attractive forces, 'tractio' vs. torsional forces 'verticitas').

In the 17th century, not much progress is made (Mechanics is considered more interesting.) However, electric devices (based on mechanisms of friction) become popular as tool of entertainment at European courts.

Early 18th century: the era of 'qualitative electrostatics': **Benjamin Franklin**, 1706–1790, invents the **lightning rod** and introduces the notion of **electric charge** (a term inspired by military vocabulary, the 'discharging' of a canon, similarly with 'batteries').¹ The existence of

¹Arguably Franklin owed much of his success to the fact that he lived in America and remained, therefore, largely uninfluenced by false concepts circulating in Europe.

positive and negative charges is understood.

Late 18th century: the era of 'quantitative electrostatics': **Charles Auguste de Coulomb**, 1736–1806, discovers the **Coulomb force** $\mathbf{F} \sim q_1 q_2 \mathbf{r}/r^3$ between electric charges. (Independently discovered by others.) Much underrated, **Henry Cavendish**, 1731–1810: discovers Coulomb law, understands essential aspects of the physics of dielectric materials and of electric conduction. (But is a lazy publisher.)

Alessandro Volta, 1745–1827, invents the Voltaic column (and presents it to Napoleon), emancipation of unreliable friction devices.

George Green (date of birth unknown) and **Karl Friedrich Gauss** (1777–1855) develop the 'final form' of electrostatics.

2.1.3 Romantic era till late nineteenth century

Early 19th century: the era of 'magnetostatics': the invention of the electric battery paves the way to the systematic analysis of electric currents and to conduction experiments.

Inspired by the romantic idea of a 'global spirit' and the understanding that all forces are mere manifestations of a unified force (Schelling), researchers such as Oersted and Faraday attempt to understand electricity and magnetism on a common basis. Nature philosophy actively influences science!

Hans Christian Oersted, 1777–1810, qualitative discovery of magnetic forces induced by electric currents. (Quantitative theory → Biot and Savart.) The advent of Electro–Magnetism. **Anrdré Marie Amperè**, 1775–1836, discovery of forces between conductors carrying currents. **Georg Simon Ohm**, 1789–1854, discovery of Ohm's law.

Michael Faraday, 1791–1867, seeks for reversal of Oersted's experiment ('convert magnetism into electricity') → the law of induction, principle of the generator. Faradays work leads to the emergence of a fundamentally new understanding of electromagnetic forces: the idea of a **field** (→ a conceptual revolution), i.e. the understanding that, independent of the existence of matter, fields are an intrinsic quality of space. Matter influences fields and vice versa. However, the existence of fields is not tied to the presence of matter. Essentially, this picture forms the basis of our contemporary understanding of all forces.²

James Clerk Maxwell, 1831–1879, Maxwells equations of electromagnetism (1862). Fields acquire own existence (independent of matter) and governed by their own fundamental laws.

Heinrich Hertz, 1857–1894, discovery of electromagnetic waves (1887) → first transatlantic radiowaves 1901. Unification of electrodynamics and optics.

²Faraday tried to include gravitation into the unified picture, a work that is still far from completion!

2.1.4 Twentieth century

Albert Einstein, 1879–1955, theory of special relativity (1905). (Work inspired by earlier insight of **Hendrik Lorentz**, 1853–1928.)

Realization that electrodynamics is not a closed theory (difficulties with point charges.)
→ embedding of electrodynamics in Quantum electrodynamics (developed as of 1929.)

2.2 The apparatus of electrodynamics

2.2.1 Basic quantities

Def.: **(classical)³ electrodynamics**: Classical theory of electric and magnetic phenomena.

Central to the theory of electrodynamics ED: four **electromagnetic fields**:

field	denotation (historical)
E	electric field strength
B	magnetic induction
D	electric displacement
H	magnetic field

Table 2.1: Electromagnetic fields

▷ Info. Mathematically, a **field** is a mapping $\phi : M \rightarrow T, x \mapsto \phi(x)$ from a base manifold M into a target manifold T .

Examples: Temperature distribution in the lecture hall, a mapping $(\text{lecture hall}) \subset \mathbb{R}^3 \rightarrow \mathbb{R}, \mathbf{r} \mapsto T(\mathbf{r})$ — a scalar field in three-dimensional space. Wind weather forecast, a mapping $(\text{region of forecast}) \subset \mathbb{R}^2 \rightarrow S^1, \mathbf{x} \mapsto \mathbf{n}(\mathbf{x})$, where the unit vector $\mathbf{n}(\mathbf{x})$ defines the direction of wind — a field on the unit circle, S^2 , etc.

³As with classical mechanics, classical electrodynamics (ED) is not a closed theory. It runs into problems at small length scales (→ infinite self energies) which find their solution within the larger framework of **quantum electrodynamics**.

The fields $\mathbf{X} \equiv \mathbf{E}, \dots, \mathbf{H}$ are vector fields in (3+1) dimensional space–time:⁴

$$\begin{aligned}\mathbf{X} : \mathbb{R}^3 \times \mathbb{R} &\rightarrow \mathbb{R}^3, \\ x \equiv (\mathbf{x}, t) &\mapsto \mathbf{X}(x).\end{aligned}$$

▷ Info. In fact, the fundamental fields of ED are *not* vector fields! The identification $\mathbf{E}, \dots, \mathbf{H} \leftrightarrow$ (vector fields) is an outgrowth of history. It makes sense only locally, and as long as we stick to a fixed inertial system. (cf. with the phase space of classical mechanics which, also, is not a vector space.) The reason why we stick to the vector field convention in the early parts of this course is its overwhelming prevalence in the physics community.

To complete the setup of the theory, we need to introduce two more fields, the **matter fields** of ED:

matter field	denotation
ρ	charge density
\mathbf{j}	current density

Table 2.2: Matter fields

The **charge density** is a scalar field $\rho : \mathbb{R}^4 \rightarrow \mathbb{R}, x \mapsto \rho(x)$ describing the distribution of charges in space and time (see section 2.2.4 below.) the **current density** $\mathbf{j} : \mathbb{R}^4 \rightarrow \mathbb{R}^3, x \mapsto \mathbf{j}(x)$ is a vector field describing the distribution of electric current (densities).

2.2.2 Interaction of matter and fields

Charge and current density act as ‘sources’ driving the generation of electromagnetic fields. The connection (matter fields \rightsquigarrow electromagnetic fields) is provided by the celebrated **Maxwell equations** (expressed in Gaussian units, see section 2.2.3 below)

$$\nabla \cdot \mathbf{D} = 4\pi\rho, \tag{2.1}$$

$$\nabla \times \mathbf{H} - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{D} = \frac{4\pi}{c} \mathbf{j}, \tag{2.2}$$

$$\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial}{\partial t} \mathbf{B} = 0, \tag{2.3}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{2.4}$$

⁴Notation: Most vectors in ED are defined in 3+1 space–time dimensions. The space like components will be denoted by boldface characters, e.g., $\mathbf{x} \in \mathbb{R}^3$. The unification with the time–like component (e.g., t) to a four–component object will be often be denoted by a non–boldface symbol, e.g., $x = (ct, \mathbf{x}) \in \mathbb{R}^4$, where the **vacuum speed of light** $c = 3 \cdot 10^8 \text{ m s}^{-1}$ is included to give all components the same dimension of physical length.

Eqs. (2.3) and (2.4) are the homogeneous (source-free) equations, Eqs.(2.1) and (2.2) the inhomogeneous Maxwell equations.

The Maxwell equations alone do not suffice to uniquely determine the dependence of the fields $\mathbf{E}, \dots, \mathbf{H}$ on the sources $j \equiv (\rho, \mathbf{j})$. (Too few differential equations for $4 \times 3 = 12$ field components!) Implicit to the Maxwell equations are relations $\mathbf{D} = \mathbf{D}[\mathbf{E}]$ ($\mathbf{H} = \mathbf{H}[\mathbf{B}]$) determining the electric displacement in terms of the electric field strength (the magnetic field in terms of the magnetic induction.)⁵ Particularly simple connections $\mathbf{D} \leftrightarrow \mathbf{E}$ ($\mathbf{H} \leftrightarrow \mathbf{B}$) apply in the case of vacuum electrodynamics (on which we will focus in the first chapters of this course):

$$\text{vacuum:} \quad \mathbf{D} = \epsilon_0 \mathbf{E}, \quad \mathbf{H} = \frac{1}{\mu_0} \mathbf{B}, \quad (2.5)$$

where the **dielectric constant** $\epsilon_0 = 1$ and the **magnetic permeability** $\mu_0 = 1$ are constants,⁶ i.e. up to constants of proportionality (whose significance will be discussed below), the fields \mathbf{D} and \mathbf{E} (\mathbf{H} and \mathbf{B}) can be identified with each other.

▷ Info. From a purist point of view the sources of Maxwells equation contain all charges and currents in the universe. In particular the abundance of microscopic charges and currents bound in extended matter are explicitly included in the source terms. Within this **microscopic or vacuum formulation of electrodynamics** one may identify $\mathbf{E} = \mathbf{D}$ and $\mathbf{H} = \mathbf{B}$. In general, however, the microscopic approach is met with insurmountable problems: *Conceptually*, the correct description of charges on microscopic scales requires a quantum mechanical formulation, i.e. a theoretical framework beyond the scope of classical electrodynamics. *Practically*, it is impossible to solve Maxwells equations for, say, the $\mathcal{O}(10^{23})$ charges embodied in a macroscopic piece of matter. This means that the microscopic form of Maxwells equations is applicable only to those situations where the system under consideration is effectively matter-free (except for a few isolated carriers of charges and currents that may be described in the idealized manner discussed in section 2.2.4 below.) The extension of the theory to the presence of extended matter will be discussed in chapter xxx below.

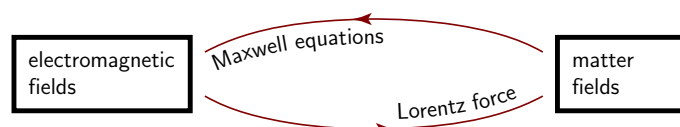


Figure 2.1: On the mutual influence (electromagnetic fields \leftrightarrow matter fields)

For a given dependence $\mathbf{D}[\mathbf{E}], \mathbf{H}[\mathbf{B}]$ (and given boundary conditions, see below) the

⁵For two arbitrary functions f and g , the notation $f[g]$ indicates that f is a **functional** of g , i.e. depends on the function g as a whole. E.g. the notation $\mathbf{D}[\mathbf{E}]$ indicates that, in general, $\mathbf{D}(\mathbf{x}, t)$ may depend on the values $\{\mathbf{E}(\mathbf{x}', t')\}$ takes at different instances of space and time.

⁶... assuming the value unity in the Gaussian system of units chosen here, cf. section 2.2.3 below.

Maxwell equations uniquely specify the fields in terms of the charges. Reciprocally, the fields exert influence on the charge distribution. Responsible for this feedback mechanism is the **Lorentz force (density)**

$$\mathbf{f} = \rho \mathbf{E} + \frac{1}{c} \mathbf{j} \times \mathbf{B}, \quad (2.6)$$

where $\mathbf{f}(\mathbf{x}, t)dV$ is the force a small element of matter concentrated in the volume element dV experiences due to the presence of the electromagnetic fields. Thus, *electromagnetic fields are influenced by matter fields and vice versa.*

▷ Info. Consequence of the mutual influence of fields on matter: a new understanding of **forces between particles**. Newtonian mechanics: the interaction between two (charged) particles A and B is instantaneous \rightsquigarrow contradiction with the principle of relativity, i.e. the existence of a maximal speed of signal propagation (the velocity of light), see below. In ED, the interaction between A and B is mediated by a field: the charge of particle A creates an electromagnetic field (via the Maxwell equations) which then influences B (via the Lorentz force.) This indirect interaction is retarded (by the speed of field propagation which will turn out to be the speed of light.) History: the experimentally observed forces between charges could not be understood in terms of an instantaneous Newtonian interaction and *required* the introduction of non-matter fields into our description of nature. Modern physics describes all four fundamental interactions as indirect processes matter \leftrightarrow (gauge) field \leftrightarrow matter.

Indeed, we will see that ED is compatible with the principles of special relativity while classical Newtonian mechanics is not. Precise formulation: under a general Lorentz transformation (see chapter xx), the Maxwell equation remain form invariant. However, this transformation is not manifest (a consequence of the interpretation of the EM fields as vector fields!)

2.2.3 Maxwell equations: generalities

Objectives of this lecture course:

- ▷ Understand connection matter fields \rightsquigarrow EM fields, i.e. explore how to solve the Maxwell equations. (Roughly two thirds of the course.)
- ▷ Understand the general structure of the theory \rightsquigarrow special relativity and its ramifications in ED.

Here: warmup to the first task.

Structure of the Maxwell equations

Maxwell equations: a system of *coupled first order linear partial differential equations*. (The good news: 'first order' and 'linear'). Apart from problems with exceptionally high symmetry no closed solutions can be found (cf. with situation in Newtonian mechanics.) \rightsquigarrow numerical

integration theory of Maxwell equations in engineering. However, a far reaching understanding of the general behaviour of fields can be developed, a consequence of the fact that

Electrodynamics is a linear theory.

Which means that if $\mathbf{E}, \dots, \mathbf{H}$ ($\mathbf{E}', \dots, \mathbf{H}'$) solves the Maxwell equations for sources (\mathbf{j}, ρ) ((\mathbf{j}', ρ')), the fields $\mathbf{E} + \mathbf{E}', \dots, \mathbf{H} + \mathbf{H}'$ will be a solution to $(\mathbf{j} + \mathbf{j}', \rho + \rho')$, the **superposition principle**. For example, to understand the electric field created by a system of point charges in space, it suffices to understand the field created by a *single* point charge. The field of the full charge distribution obtains by adding the constituent fields, etc.

Units

The detailed form of the equations of ED depends on the chosen system of units. Two principal approaches:

- ▷ The practitioner's point of view: Stick to the international system of units (**SI-system**, aka **MKSA-system**) wherein length (1 m), mass (1 kg), time (1 s), *and* current (1A) carry their own physical units. Express all equations so as to conform with this unit system.
- ▷ The purist's point of view: avoid the introduction of new units as much as possible and use the **Gauß system** aka **cgs-system** of units. For this, I am willing to pay the price of being at odds with most of the people (engineers & applied physicists) working with electromagnetic phenomena.

The SI-system is very lab-friendly — an invaluable advantage. Yet by introducing units which are only seemingly fundamental (1A), it tends to obscure the principal connection between physical quantities. For this reason *we will use the Gauß system of units in this course*.

▷ Info. The **definition of the Ampère** in the SI-system. Two infinite parallel wires kept at a distance of 1m carry a uni-directional electric current of 1A each if they exert an (attractive) force of $2 \cdot 10^{-7}$ N on each other.

In addition to the Ampère, the SI-system contains a number of secondary (non-fundamental) units. Specifically, the unit of electric charge, the **Coulomb**, C, is fixed as $1 \text{ C} \equiv 1 \text{ A}\cdot\text{s}$, i.e. a conductor carries 1C of charge if a current of strength 1A has flown into it for 1 second. The strength of the electric field is measured in volt/meter: A point in space is subject to an electric field of strength one **volt** per meter (1 V/m) if a charge of 1C placed at that point will experience a force of strength $F=1\text{N}=1 \text{ m kg/s}^2$. Since the electrostatic Lorentz force is $F = QE$, we find that $1\text{V} = 1 \text{ m}^2 \text{ kg}/(\text{A s}^3)$. The Maxwell equation $\nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0$ implies $[\mathbf{B}] = [\mathbf{E}] \times (\text{time/length})$. Magnetic fields are measured in units **tesla**, where $1\text{T} = 1 \text{ V s/m}^2$.

The **Maxwell equations in the SI-system** assume the form

$$\begin{aligned}\nabla \cdot \mathbf{D} &= \rho, \\ \nabla \times \mathbf{H} - \frac{\partial}{\partial t} \mathbf{D} &= \mathbf{j}, \\ \nabla \times \mathbf{E} + \frac{\partial}{\partial t} \mathbf{B} &= 0, \\ \nabla \cdot \mathbf{B} &= 0,\end{aligned}$$

where the value of the vacuum dielectric constant is $\epsilon_0 = 8.8544 \cdot 10^{-12}$ As/Vm and the magnetic permeability is given by $\mu_0 = 12.5664 \cdot 10^{-7}$ Vs/Am.

To elucidate the **difference between the two systems** consider two static charges q_1 and q_2 at distance r . Anticipating our discussion below we note that the charge q_1 (kept at $\mathbf{r} = 0$, say) will create a static **D**-field of strength $D(\mathbf{r}) = c_0 \cdot q_1/r$, where $r = |\mathbf{r}|$ and the constant $c_0 = 1$ (Gauß) $c_0 = 1/4\pi$ (SI) depends on the convention chosen in the first of the Maxwell equations. This corresponds to a field **E** of strength $E = \epsilon_0^{-1}D = c_0 q_1/(\epsilon_0 r^2)$. By virtue of the Lorentz law the second charge will then feel a (Coulomb) force $F = c_0 q_1 q_2/\epsilon_0 r^2$.

According to the philosophy of the Gauß system, this law — bridging between mechanic and electromagnetic quantities — *defines* the unit of charge. Specifically, we say that two point particles carry a charge of **1 esu** (one electrostatic charge unit) if they experience a mutual force of 1 dyn = 1 g cm/s² when kept at a distance of 1cm: 1 esu = 1 g^{1/2} cm^{3/2} s⁻¹. (Units for magnetic fields, currents, etc. are introduced in a similar spirit and will be discussed whenever we need them.) In contrast, the SI-system has already fixed the unit of charge by the redundant yet convenient introduction of the Ampère. Experimentally, one finds that two charges of 1C each kept at a distance of 1m exert a force of strength $F=8.987 \cdot 10^9$ N on each other, i.e. $8.987 \cdot 10^9$ N = $\frac{1}{4\pi\epsilon_0}$ 1 Q² m⁻². Solving for ϵ_0 , we obtain the value given above. Thus, the price to be paid for the introduction of handy units of charge is the appearance of inconvenient conversion factors between the fields.⁷ In the CGS system there is no need to distinguish between the electric field and the displacement field and we may set $\epsilon_0 = 1$. Relatedly, the SI-system obscures the connection between quantities which, fundamentally, carry the same physical dimension.

For an in-depth discussion of systems of units (including a few more than those discussed here), we refer to [1].

2.2.4 Charges and currents

General definition of charge and current distributions

We begin our analysis of the Maxwell equations with a discussion of its sources. First consider the concept of **charge**. Empirically, we know that

- ▷ The existence of charge is tied to matter.

⁷Similar reasoning can be applied to explain the value of the magnetic permeability.

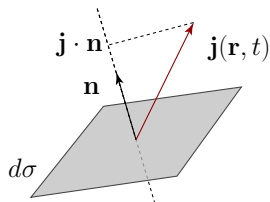
- ▷ charge is a scalar quantity that can be measured: Fixing a certain reference charge (for example 1 esu or 1C), we may assign a value to other charges by comparison⁸ with that reference charge.
- ▷ The value of the charge $Q_V \leftrightarrow V$ assigned to a volume V of matter can be positive and negative.
- ▷ Charge is an 'extensive' scalar quantity which means that if two disjoint regions of space, V_1 and V_2 , $V_1 \cap V_2 = \{ \}$, contain charges Q_{V_1} and Q_{V_2} , respectively, then the total charge contained in $V_1 \cup V_2$ is given by $Q_{V_1 \cup V_2} = Q_{V_1} + Q_{V_2}$.
- ▷ Charge is a 'conserved' quantity, i.e. as far as we know, nowhere in universe is charge created. (This, however, does not exclude the creation of positive and negative charges of equal and opposite magnitude.)

These features motivate the definition of a **charge density** distribution as follows: For a given point in space, \mathbf{x} , let $V_1 \supset V_2 \supset \dots$ be a sequence of volume elements converging onto the point \mathbf{x} . Let Q_n be the charge contained in V_n .⁹ We then define the charge density at \mathbf{x} as

$$\rho(\mathbf{x}, t) \equiv \lim_{n \rightarrow \infty} \frac{Q_n(t)}{V_n}. \quad (2.7)$$

With this definition, the total charge contained in a finite region of space, V , is given by

$$Q_V(t) = \int_V d^3x \rho(\mathbf{x}, t). \quad (2.8)$$



Since charges cannot be spontaneously created, the only way for Q_V to change is that charges enter or leave the volume V through its surface $\partial V = S$. We define the **(charge) current** through S by

$$I_S(t) + \dot{Q}_{V(S)}(t) = 0, \quad (2.9)$$

i.e. the total current flowing through the surface S equals the rate at which the charge contained in V changes. This is an example of a **global conservation law**.

We wish to generalize the local definition of current to a refined, local one: Define a **current density** vector field $\mathbf{j}(\mathbf{x}, t)$ by demanding that for any oriented surface element $d\sigma$ with normal vector $\mathbf{n}(\mathbf{r})$, $d\sigma(\mathbf{j} \cdot \mathbf{n})dt$ is the amount of charge passing through $d\sigma$ in the time interval $[t, t + dt]$ (why is this a genuine definition?) The amount of charge $dQ_{S,\text{out}}$ flowing during the interval $[t, t + dt]$ through a surface S of finite extent is obtained by integration $dQ_{S,\text{out}} = dt \int_S d\sigma (\mathbf{n} \cdot \mathbf{j})$. Defining the **current** through S , I_S by the rate of charge transfer, $I_S = dQ_{S,\text{out}}/dt$, we obtain

$$I_S(t) = \int_S d\sigma (\mathbf{j} \cdot \mathbf{n}).$$

⁸In practice, the notion 'comparison' implies an experiment, e.g., the measurement of the force acting on a charge.

⁹Notice that $Q_n = Q_n(t)$ may depend on time.

Specifically, for a closed surface, $-d_t Q_{V(S)} = d_t Q_{S,\text{out}} = I_S$, as stated by the global conservation law above. Further, recalling the definition of charge Eq. (2.8) and the conservation law (2.9), we obtain the $\int_V d^3x (\nabla \cdot \mathbf{j} + \partial_t \rho) = 0$. This equation must hold for any choice of the reference volume V . The only way to satisfy this condition is to require the **continuity equation**

$$\boxed{\partial_t \rho + \nabla \cdot \mathbf{j} = 0.} \quad (2.10)$$

Eq. (2.10) represents the ‘local formulation’ of the conservation of current. Advantage of the global formulation: It is arguably more intuitive. Advantage of the local formulation: It does not make reference to a test volume and it is ‘local’, i.e. makes a statement about charges and currents at one given instance of space–time.

A shorthand notation: Define the — at this stage purely formal — notation: $\partial_\mu \equiv (c^{-1}\partial_t, \nabla)$, $\mu = 0, 1, 2, 3$. Then the continuity equation assumes the form:

$$\partial_\mu j^\mu = 0, \quad (2.11)$$

where $j^{\mu=0} = c\rho$, $j^{\mu=1,2,3} = j^{i=1,2,3}$ are the components of the **four–vector current density**.¹⁰

Notice that current conservation is implied by the Maxwell equations: Taking the divergence of $c \times (2.2)$ and adding it to the time derivative of (2.1), we obtain

$$\underbrace{c\nabla \cdot (\nabla \times \mathbf{H})}_0 = 4\pi(\partial_t \rho + \nabla \cdot \mathbf{j}).$$

Point charges and current loops

To facilitate the analysis of the Maxwell equations, we will often work with charge and current distributions that are ‘singular’ in that they are concentrated on infinitely sharply concentrated regions in space (mathematically: sets of measure zero.) Specifically, we will consider point–like charges (the analog of point particles in classical mechanics) and currents flows concentrated on mathematical curves (the latter representing theorists cartoons of ‘infinitely thin wires’). To meaningfully represent such singular configurations in terms of the source functions ρ and \mathbf{j} , we need to introduce the notion of distributions.

▷ Math. How would one describe a ‘point charge’, i.e. an accumulation of a finite amount of charge, q , concentrated on a single point in space, in terms of a charge distribution ρ ? What we need is a ‘function’ that is zero almost everywhere, except for one single point. Such functions have been popularized in physics (long before they made their way to mathematics) as δ –functions. Our objective here is to give the ‘ δ –function’ a precise definition, and, equally important, to learn how to work with it.

To start with, consider the case of functions $f : \mathbb{R} \rightarrow \mathbb{R}$ in one–dimensional space. Define: $\mathcal{D}(\mathbb{R}) = \mathcal{C}_c^\infty(\mathbb{R})$ where $\mathcal{C}_c^n(\mathbb{R})$ is the space of n –fold differentiable functions with compact support.

¹⁰Following a widespread convention, we will denote the space like components of vectors by intermediate lowercase latin variables, $i, j, k, \dots = 1, 2, 3$. (3+1)–dimensional space–time vectors are indexed by intermediate lowercase Greek variables, $\lambda, \mu, \nu, \dots = 0, 1, 2, 3$.

(Elements of \mathcal{D} are very benign. For reasons to be come clear momentarily, they are sometimes called **test-functions**.) Notice that \mathcal{D} forms an (infinite-dimensional) real vector space: For two functions $\phi_1, \phi_2 \in \mathcal{D}$ and $c_1, c_2 \in \mathbb{R}$, $c_1\phi_1 + c_2\phi_2 \in \mathcal{D}$.¹¹

A continuous functional on \mathcal{D} is a continuous mapping $F : \mathcal{D} \rightarrow \mathbb{R}$, $\phi \mapsto F[\phi]$ assigning to every $\phi \in \mathcal{D}$ a number. A **continuous linear functional** is a functional that is linear: $F[c_1\phi_1 + c_2\phi_2] = c_1F[\phi_1] + c_2F[\phi_2]$. A **distribution** F on \mathbb{R} is a continuous linear functional subject to one additional condition: if the sequence of test functions $\phi_k \xrightarrow{k \rightarrow \infty} \phi$ converges to the test functions ϕ , then $\lim_{k \rightarrow \infty} F[\phi_k] \rightarrow F[\phi]$.

▷ Example: Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function. Define the functional

$$F_f[\phi] \equiv \int_{-\infty}^{\infty} dx f(x)\phi(x). \quad (2.12)$$

(Why does F_f meet the defining criteria above?) Distributions of this type are called **regular distributions**. In a way, the function ϕ probes the properties of f , hence the denotation ‘test function’.

Not every distribution is regular. As an important example we mention:

▷ Example: The **δ -distribution**: For $a \in \mathbb{R}$, define

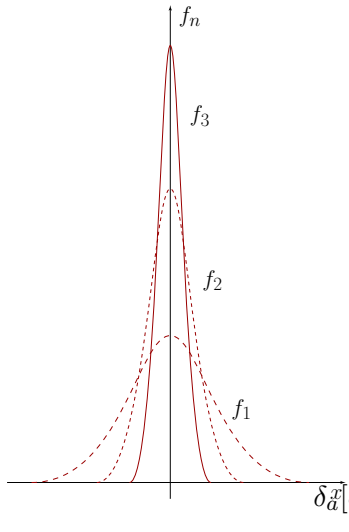
$$\delta_a[\phi] = \phi(a). \quad (2.13)$$

The δ -distribution meets all the criteria of the definition above. Evidently, however, there is no regular function such that $\delta_a[\phi] = \int dx f(x)\phi(x)$.

Although the δ -distribution as such is not regular, one may prove the following statement: every distribution F can be represented as the limit $F = \lim_{n \rightarrow \infty} F_{f_n}$ of a sequence of regular distributions. (As we shall see, the functions f_n ‘modeling’ a irregular distribution are of considerable applied relevance.) Specifically, let us construct a sequence of functions f_n^a such that $F_{f_n^a} \rightarrow \delta_a$ converges to the δ -distribution:

$$f_n^a(x) = \frac{n}{\epsilon} \exp\left(-\frac{\pi n^2(x-a)^2}{\epsilon^2}\right).$$

¹¹More generally, the set $\mathcal{C}^n(U, \mathbb{R}^r)$ containing the n -fold differentiable functions $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}^r$ forms a vector space, too. (Functions of this type can be added and multiplied by numbers.) This is why we often speak of a *space* of functions of a certain type. The interpretation of functions as elements of (infinite dimensional) spaces will play a pivotal role below.



The functions f_n^a are unit-normalized, $\int dx f_n^a(x) = 1$ and, for large enough n , sharply peaked around $x = a$. Taylor expanding the test-function ϕ around $x = a$, one verifies that

$$F_{f_n^a}[\phi] = \frac{n}{\epsilon} \int dx \exp\left(-\frac{\pi n^2(x-a)^2}{\epsilon^2}\right) \phi(x) = \phi(a) + \mathcal{O}(\epsilon^2/n^2),$$

i.e. $\lim_{n \rightarrow \infty} f_n^a[\phi] = \delta_a[\phi]$. In the physics community, it is customary to work with the (symbolic, because the limit does not exist) notation)

$$\delta(x-a) \equiv \lim_{n \rightarrow \infty} f_n^a(x), \quad (2.14)$$

and to think of $\delta(x-a)$ as a 'function', the **δ -function**, infinitely sharply peaked around a . Using this notation,

$$\delta_a^x[\phi] = \phi(a) = \int dx \delta(x-a)\phi(x) \equiv F_{\delta(x-a)}[\phi].$$

Remarks: (a) Of course, the δ -distribution may be generated by functions f_n different from the Gaussian functions above. (Invent examples.) (b) We will use the sloppy physics notation throughout.

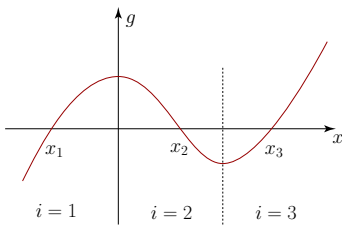
For a differentiable function, f , a straightforward integration by parts shows that $F_{f'}[\phi] = -F_f[\phi']$. Applying this result to the sequence of functions converging onto the δ -distribution, we obtain a formal definition of the 'derivative of the δ -function',

$$F_{\delta'(x-a)} = -F_{\delta(x-a)}[\phi'] = -\phi'(a).$$

Conversely, the δ -function may be obtained as the derivative of the Heaviside step function Θ . Of course, the step function $\Theta(x-a)$ is not differentiable in the conventional sense. (Its difference quotient is zero everywhere, except for $x = a$ where it diverges. Thus, the formal limit of the difference quotient, the 'derivative' has properties akin to a δ -function.). Formally, however, we may define

$$F_{\Theta'(x-a)}[\phi] = -F_{\Theta(x-a)}[\phi'] = -\int_{-\infty}^{\infty} dx \Theta(x-a)\phi'(x) = -\int_a^{\infty} dx \phi'(x) = \phi(a) = F_{\delta(x-a)}[\phi],$$

where in the second last equality we used that $\phi(\infty) = 0$ (compact support of test functions.) Since this equality holds for arbitrary functions testing the step function, we define $\frac{d}{dx}\Theta(x-a) = \delta(x-a)$.



For an arbitrary differentiable function g and a sequence of functions $f_n(x)$ converging onto $\delta(x)$, we define $\delta(g(x)) \equiv \lim_{n \rightarrow 0} f_n(g(x))$. Evaluating the test function integral,

$$\begin{aligned} F_{\delta(g)}[\phi] &= \int dx \delta(g(x))\phi(x) = \sum_i \int \frac{dg}{|g'(x)|} \delta(g)\phi(g^{-1}(x)) = \\ &= \sum_i \frac{1}{|g'(x_i)|} \phi(g^{-1}(x_i)), \end{aligned}$$

where the sum extends over all those regions where g is invertible (see the figure.) and x_i are the zeros of g in these domains. In the notation of the δ -function:

$$\delta(g(x)) = \sum_i \frac{1}{|g'(x_i)|} \delta(x-x_i). \quad (2.15)$$

To summarize our so far results (and a few more that are trivially proven):

$$\triangleright \delta(x - a)f(x) = \delta(x - a)f(a),$$

$$\triangleright \delta(x) = \delta(-x),$$

$$\triangleright \Theta'(x - a) = \delta(x - a),$$

$$\triangleright \delta(g(x)) = \sum_i \frac{1}{|g'(x_i)|} \delta(x - x_i),$$

$$\triangleright \delta(cx) = |c|^{-1} \delta(x).$$

The **generalization to higher dimensions** is straightforward: Assuming $d = 3$ for definiteness, we define

$$\delta(\mathbf{x} - \mathbf{a}) = \prod_{i=1}^3 \delta(x_i - a_i).$$

With this definition $\int d^3x f(\mathbf{x})\delta(\mathbf{x} - \mathbf{a}) = f(\mathbf{a})$. The generalization of Eq. (2.15) becomes (exercise)

$$\delta(\mathbf{g}(\mathbf{x})) = \sum_i \left| \frac{\partial \mathbf{g}}{\partial \mathbf{x}_i} \right|^{-1} \delta(\mathbf{x} - \mathbf{x}_i),$$

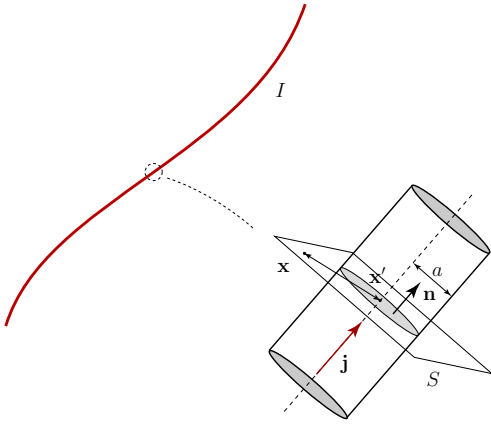
where $\mathbf{g} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is a function and $|\partial \mathbf{g} / \partial \mathbf{x}|$ a shorthand notation for the determinant of the Jacobi matrix. This equation can be used, e.g., to compute the form of the δ -function in non-cartesian coordinates. For example, in polar coordinates,

$$\delta(\mathbf{x}) = \frac{1}{r^2 \sin \theta} \delta(r) \delta(\theta) \delta(\phi).$$

(Check that this function obeys the defining properties of the δ -function.)

As with the ‘point particles’ of classical mechanics, it is often useful to consider **point charges**, i.e. infinitely small regions $V \searrow 0$ in space centered around a reference point \mathbf{x}_0 and carrying a fixed amount of charge $Q_{V \searrow 0} = q$. For example, we may chose $V(a)$ to be a ball of radius a carrying a charge density $\rho_a(\mathbf{x}) = \frac{3q}{4\pi a^3} 1_V(\mathbf{x})$, where $1_V(\mathbf{x})$ is the characteristic function of V (unity if $\mathbf{x} \in V$, zero otherwise.) Evidently, $Q_{V(a)} = q$, i.e. the shrinkage of the volume for $a \searrow 0$ is compensated by a diverging charge density. At the same time, the functions ρ_a converge onto a three-dimensional representation of the δ -function, i.e.

$$\rho(\mathbf{x}) = \lim_{a \searrow 0} \rho_a(\mathbf{x}) \equiv q\delta(\mathbf{x} - \mathbf{x}_0), \quad (2.16)$$

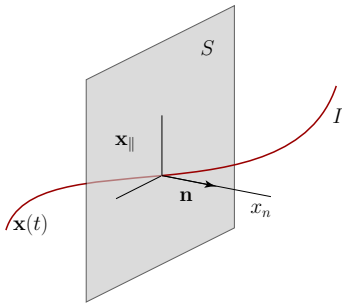


The current–analog of a point charge is a **current loops**, i.e. a curve in space carrying a uniform current I . To obtain an operational definition of a current loop, consider a curve γ in space. For most (all modulo sets of measure zero) points \mathbf{x} in space, we may identify a point $\mathbf{x}' \in \gamma$ that is closest to \mathbf{x} . (Just minimize the distance function $|\mathbf{x} - \mathbf{x}'(t)|$, where $\mathbf{x}'(t)$ is an explicit parameterization of γ .) Define $\mathbf{j}_a(\mathbf{x}) = \mathbf{n}_{\mathbf{x}'}(I/\pi a^2)\Theta(a - |\mathbf{x} - \mathbf{x}'|)$, where $\mathbf{n}_{\mathbf{x}'} = d_t\mathbf{x}'(t)/|d_t\mathbf{x}'(t)|$ is the unit vector along γ at \mathbf{x}' (cf. the figure.) To see that this definition

does the job, compute the current through any surface S cutting through γ . Without loss of generality, we may assume that S is planar and perpendicular to γ (why does this follow from charge conservation?). Then, $\mathbf{n}_{\mathbf{x}'} = \mathbf{n}$ equals the vector normal to S and $I = \int_S d\sigma \mathbf{j} \cdot \mathbf{n} = 2\pi \int_0^a \rho d\rho (I/\pi a^2) = I$, independent of a we define the current density of a loop as $\mathbf{j} = \lim_{a \searrow 0} \mathbf{j}_a$.

We next show that an alternative representation of the limiting distribution is given by

$$\mathbf{j}(\mathbf{x}) = I \int dt \dot{\mathbf{x}}(t) \delta(\mathbf{x}(t) - \mathbf{x}). \quad (2.17)$$



To prove this assertion, let us again demonstrate that the total current through any plane S cutting through the curve is given by I . Choosing coordinates so that at time t_0 the curve pierces S at $\mathbf{x} = 0$ and denoting the unit vector normal to S by \mathbf{n} , we obtain

$$\begin{aligned} I_S &= \int_S d\sigma \mathbf{n} \cdot \mathbf{j} = I \int dt \int d\sigma \dot{x}_n(t) \delta(\mathbf{x}_{\parallel}(t)) \delta(x_n(t)) = \\ &= I \int dt \int d\sigma \dot{x}_n(t) \delta(\mathbf{x}_{\parallel} - \mathbf{x}_{\parallel}(t)) \frac{\delta(t - t_0)}{d_t x_n(t)} = I \int d\sigma \delta(\mathbf{x}_{\parallel}(t_0)) = I, \end{aligned}$$

where \mathbf{x}_{\parallel} is the two–dimensional in–plane component of \mathbf{x} and we used (2.15).

2.2.5 Integral form of the Maxwell equations

▷ Math. To prepare our discussion below, we recapitulate **the laws of Stokes and Gauß**, and introduce a few more integral theorems of vector analysis.

Gauß' law: Let S be a smooth orientable and closed surface embedded in \mathbb{R}^3 . Let $\mathbf{v} : U \rightarrow \mathbb{R}^3$ be a vector field where the support $U \supset V(S)$ and $V(S)$ denotes the volume enclosed by S . Then

$$\boxed{\int_S d\sigma \mathbf{v} \cdot \mathbf{n} = \int_{V(S)} d^3x \nabla \cdot \mathbf{v},} \quad (2.18)$$

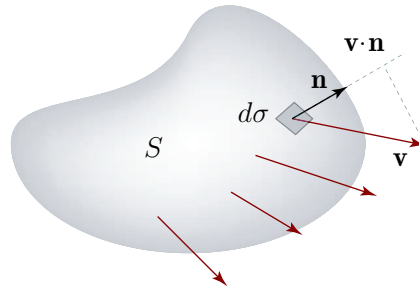


Figure 2.2: Schematic on Gauß's law

where $\mathbf{n}(\mathbf{x} \in S)$ is a unit ($|\mathbf{n}| = 1$) and outwardly directed vector at the point of the surface element of S .

▷ Example: Consider a radially symmetric vector field $\mathbf{v}(\mathbf{r}) = f(r)\mathbf{e}_r$, where \mathbf{e}_r is the radial unit vector and f an arbitrary function. Let S be a sphere of radius R centered at the origin $r = 0$. Using polar coordinates, $d\sigma = R^2 \sin\theta d\theta d\phi$ and $\mathbf{n} \cdot \mathbf{e}_r = 1$, the integral of \mathbf{v} over S is evaluated as

$$\int_S d\sigma \mathbf{v} \cdot \mathbf{n} = f(R)R^2 \int_0^\pi d\theta \sin\theta \int_0^{2\pi} d\phi = 4\pi R^2 f(R).$$

On the other hand (cf. section 8.1 for a discussion of vector differential operators in curvilinear coordinate systems)

$$\nabla \cdot (f\mathbf{e}_r) = \nabla f \cdot \mathbf{e}_r + f\nabla \cdot \mathbf{e}_r = f' + 2f/r.$$

Evaluating the r.h.s. of Gauß' theorem

$$\int_{V(S)} d^3x \nabla \cdot \mathbf{v} = \int_0^R r^2 dr \int_0^\pi \sin\theta d\theta \int_0^{2\pi} d\phi (f'(r) + 2f(r)/r) = 4\pi \int_0^R dr d_r(r^2 f) = 4\pi R^2 f(R),$$

we obtain the same answer as above.

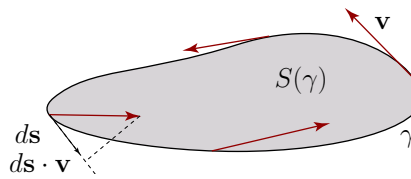


Figure 2.3: Schematic on Stokes law

Stokes law: Let γ be a smooth closed path in \mathbb{R}^3 and $S(\gamma)$ any surface bounded by γ . Further, let \mathbf{v} be a vector field defined on all of S (including its boundary $\partial S = \gamma$). Then,

$$\oint_\gamma ds \cdot \mathbf{v} = \int_{S(\gamma)} d\sigma \mathbf{n} \cdot (\nabla \times \mathbf{v}). \quad (2.19)$$

▷ Example: Consider a vector \mathbf{v} field with cylindrical symmetry, i.e. $\mathbf{v} = f(\rho)\mathbf{e}_\phi$, where cylindrical coordinates (ρ, ϕ, z) are implied. Integrating \mathbf{v} over a circular contour γ of radius R around the symmetry axis $\rho = 0$ of the problem, we obtain (exercise)

$$\oint_{\gamma} d\mathbf{s} \cdot \mathbf{v} = R \int_0^{2\pi} d\phi f(R) = 2\pi R f(R).$$

To obtain the same result by Stokes theorem, we compute

$$\nabla \times \mathbf{v} = \frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho f(\rho)) \mathbf{e}_z.$$

Thus, choosing for $S(\gamma)$ a circular disc with radius R ($d\sigma = \rho d\rho d\phi$; $\mathbf{n} = \mathbf{e}_z$),

$$\int_{S(\gamma)} d\sigma \mathbf{n} \cdot (\nabla \times \mathbf{v}) = \int_0^{2\pi} d\phi \int_0^R \rho d\rho \frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho f(\rho)) = 2\pi R f(R),$$

In agreement with the results of the line integration.

Notice that both Stokes and Gauß' law relate the integral of the derivative of a vector field over a certain manifold to the integral of the same vector field over the boundary of the manifold. The laws above are but two representatives of a family of identities of this type.¹² All these laws take the symbolic form

$$\int_M d\omega = \int_{\partial M} \omega, \quad (2.20)$$

where M is a manifold with boundary ∂M and 'd' stands for a derivative operation acting on some differentiable entity ω . (In fact, the notation above is not quite as symbolic as it may seem. It is shown in the theory of differential form that both (2.18) and (2.19) law derive from a general variant of Stokes theorem which takes exactly the form of (2.20).)

From Gauß's law one may derive a number of secondary integral identities which will be useful below: Applying (2.18) to a vector field of the form $\mathbf{v} = \theta \nabla \psi$, where θ and ψ are functions, and using the identity $\nabla \cdot (\theta \nabla \psi) = \nabla \theta \cdot \nabla \psi + \theta \Delta \psi$ we obtain **Green's first identity**

$$\int_S d\sigma \theta \partial_n \psi = \int_{V(S)} d^3r (\nabla \theta \cdot \nabla \psi + \theta \Delta \psi), \quad (2.21)$$

where $\partial_n \psi$ is the derivative of the function ψ in the direction normal to the surface S at the point of the surface element $d\sigma$.

Writing down Green's first identity with the role of θ and ψ interchanged and subtracting it from (2.21), we obtain **Green's second identity**

$$\int_S d\sigma (\theta \partial_n \psi - \psi \partial_n \theta) = \int_{V(S)} d^3r (\theta \Delta \psi - \psi \Delta \theta). \quad (2.22)$$

¹²... the simplest example of which is the one-dimensional identity $\int_a^b dx \partial_x f = f(b) - f(a)$. (Think of the function f as a one-component vector field and the interval $[a, b]$ as an integration domain with 'boundary' $\partial[a, b] = \{a\} \cup \{b\}$.)

Following the logics of the previous section in reverse order, we derive ‘global’ representations of the Maxwell equations. Historically, the – experimental – discovery of these global laws preceded the formulation of the laws of ED as differential equations by Maxwell.

▷ Math. Consider a vector field $\mathbf{v} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. We denote the scalar field $f(\mathbf{x}) = \nabla \cdot \mathbf{v}$ as the **source field** of \mathbf{v} . Motivation: Loosely speaking, the integral of \mathbf{v} over any closed surface is determined by the net amount of \mathbf{v} poking outwardly through that surface. At the same time, this integral (Gauß’ law) equals the integral of f over the volume enclosed by the surface. Imagining outwardly directed field lines¹³ as emanating from some ‘sources’, we identify $f(\mathbf{x})dV$ as the source-contents of the volume element dV .



Figure 2.4: Schematic on the source and the circulation contents of vector fields

Similarly, the vector field $\mathbf{g} \equiv \nabla \times \mathbf{v}$ is called the **circulation field** of \mathbf{v} . The field \mathbf{v} contains circulating contributions if, for some closed contours γ , $\oint_{\gamma} ds \cdot \mathbf{v} \neq 0$. Since (Stokes) $\oint_{\gamma} ds \cdot \mathbf{v} = \int_{S(\gamma)} d\sigma \mathbf{n} \cdot \mathbf{g}$, a non-vanishing of \mathbf{g} implies that curves with this property exist (chose γ to be a small curve winding around a region in space where \mathbf{g} is non-vanishing and of definite sign.)

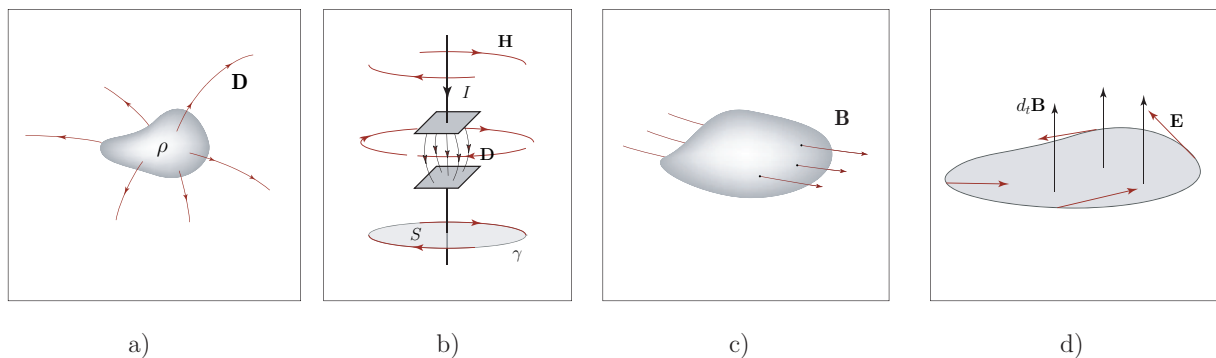


Figure 2.5: Schematic on the integral form of Maxwell's equations. a) Gauß's law, b) circulation of the magnetic field, c) absence of magnetic charges, d) law of induction

¹³Loosely speaking, the **field lines** of a vector field are curves in space locally tangent to the field (i.e. a visualization of the field is obtained by attaching arrows tangent to the field lines.) The more precise definition goes as follows: For each point \mathbf{x} in space define a curve $\Phi_{\mathbf{x}}(t)$ such that $d_t \Phi_{\mathbf{x}}(t)|_{t=0} = \mathbf{v}(\mathbf{x})$. $\Phi_{\mathbf{x}}(t)$ traces out the field line through \mathbf{x} . The map $\mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3, (\mathbf{x}, t) \mapsto \Phi_{\mathbf{x}}(t)$ is called the **flux** of the vector field.

Gauß' law (physics)

According to experimental observation,¹⁴ electric charges are the sources of the electric (displacement) field:

$$\int_S d\sigma \mathbf{D} \cdot \mathbf{n} = 4\pi \int_{V(S)} d^3x \rho = 4\pi Q_{V(S)}, \quad (2.23)$$

where S is an arbitrary closed surface, i.e. the amount of field lines leaving the volume $V(S)$ is determined by the charge $Q_{V(S)}$ inside $V(S)$ (cf. Fig. 2.5 a.) Eq. (2.23) is what physicists denote as **Gauß's**. The qualifying attribute '(physics)' is appropriate because the empirical law (2.23) and the mathematical law (2.18) need to be distinguished.

Applying (2.18) to the l.h.s. of this equation and using the arbitrariness of S (cf. the analogous argumentation on currents in the previous section), we obtain the Maxwell equation (2.1).

Circulation of the magnetic field

Consider a surface S in space pierced by electric current. Experimentally, one finds that the line integral of the magnetic¹⁵ field circulating around that surface equals that current. Likewise, a surface pierced by non-vanishing temporally fluctuating displacement field also supports a magnetic field circulation (cf. Fig. 2.5 b.) These observations are summarized in the law

$$\int_{S(\gamma)} d\sigma \left(\frac{4\pi}{c} \mathbf{j} + \frac{1}{c} d_t \mathbf{D} \cdot \mathbf{n} \right) = \int_{\gamma} ds \cdot \mathbf{H}. \quad (2.24)$$

Application of Stokes law to the r.h.s. obtains $\forall S : \int_S d\sigma (\nabla \times \mathbf{H} - c^{-1} d_t \mathbf{D} - 4\pi c^{-1} \mathbf{j}) \cdot \mathbf{n} = 0$, i.e. the Maxwell equation (2.2).

Absence of magnetic charges

To the best of our knowledge no isolated magnetic charges exist in the universe; the net flux¹⁶ of the magnetic field through any closed surface equals zero: $\forall S : \int_S d\sigma \mathbf{B} \cdot \mathbf{n} = 0$ (cf. Fig. 2.5 c.) Application of Gauß' law to this equation yields the Maxwell equation (2.4).

Law of induction

Consider a time dependent magnetic induction \mathbf{B} piercing a surface S bounded by a curve γ . One observes that the circulation of the electric field strength around γ is determined by the

¹⁴Experimentally, the field \mathbf{D} is determined by measuring the force $\mathbf{F} = \epsilon_0^{-1} q \mathbf{D}$ locally acting on a test charge q .

¹⁵The magnetic field can be determined, e.g., by measuring the torque exerted on a current loop.

¹⁶It is customary to denote the integral $\Phi_S(\mathbf{v}) = \int_S d\sigma \mathbf{n} \cdot \mathbf{v}$ of a vector field over a surface S as the **flux** of the field through S . Do not confuse this quantity with the flux Φ introduced previously!

time derivative of the flux of \mathbf{B} through S (cf. Fig. 2.5 d)):

$$\oint_{\gamma} d\mathbf{s} \cdot \mathbf{E} = -\frac{1}{c} \int_S d\sigma \mathbf{n} \cdot d_t \mathbf{B}.$$

Applying Stoke's theorem to the l.h.s. of this equation, we find $\forall S : \int \int_S d\sigma \mathbf{n} \cdot (\mathbf{E} + c^{-1} d_t \mathbf{B}) = 0$, equivalent to the Maxwell equation (2.3)

3.0 Electromagnetostatics

So far: discussed mathematical structure of Maxwell equations and outlined connection to experimentally observable phenomena. Now: develop solution theory of Maxwell equations.

3.1 Decoupling of the Maxwell equations in the static limit

Main source of complexity: coupling of the four equations to an in-separable entity. Drastic simplifications arise in cases where the fields are static: $\partial_t \mathbf{E} = \dots = \partial_t \mathbf{H} = 0$. What are necessary and sufficient conditions for time-independence of the fields?

Equation (2.1) implies that $\nabla \cdot \partial_t \mathbf{D} = 4\pi \partial_t \rho$, i.e. $\partial_t \mathbf{D} = 0$ necessarily implies $\partial_t \rho = 0$, the time-independence of all charge distributions. This in turn enforces (Eq. (2.10)) that $\nabla \cdot \mathbf{j} = 0$, i.e. in the static limit there are no sources of electric currents, all field lines of the current field are closed. Differentiating the Maxwell equation (2.2) w.r.t. time, and noting that $\partial_t \mathbf{H} = \partial_t \mathbf{D} = 0$ we further conclude that the current flow is stationary, $\partial_t \mathbf{j} = 0$. To summarize, we find that

The time-independence of the EM fields necessarily implies the stationarity of charge densities, $\partial_t \rho = 0$, and of the currents $\partial_t \mathbf{j} = 0$; in the static limit, the current-field is source-free, $\nabla \cdot \mathbf{j} = 0$.

Arguing in reverse, we consider the Maxwell equations for stationary sources $\partial_t \rho = 0$ and $\partial_t \mathbf{j} = 0$. In the following we will show that under these conditions stationary solutions $\mathbf{E}, \dots, \mathbf{H}$ can be found. Adopting stationarity of the fields as a working hypothesis, we obtain the **static limit of the Maxwell equations**

$$\begin{aligned}\nabla \cdot \mathbf{D} &= 4\pi\rho, \\ \nabla \times \mathbf{E} &= 0,\end{aligned}\tag{3.1}$$

$$\begin{aligned}\nabla \times \mathbf{H} &= \frac{4\pi}{c}\mathbf{j}, \\ \nabla \cdot \mathbf{B} &= 0.\end{aligned}\tag{3.2}$$

Importantly, the set of equations decomposes into two groups, one describing the ‘electric sector’ of the theory, with the charge density as a source, the other the magnetic sector, with the current density as a source.

3.2 Electrostatics

3.2.1 Formulation of the problem

We begin by exploring the theory of static electric fields, as described by the basic equations (3.1) (plus the vacuum relation $\mathbf{D} = \epsilon_0 \mathbf{E}$.) Exploiting that we are working in the Gauß system, we will set $\epsilon_0 = 1$ throughout and indicate the ϵ_0 -dependence of the theory only in final formulae.

For an electric field defined in a **simply connected** region of space,¹ the fact that \mathbf{E} is circulation-less, $\nabla \times \mathbf{E}$, implies² that \mathbf{E} can be written as the gradient of a scalar function ϕ :

$$\boxed{\mathbf{E} \equiv -\nabla\phi}, \quad (3.3)$$

where $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$, $\mathbf{x} \rightarrow \phi(\mathbf{x})$ is the **scalar potential** of the theory. Substituting this ansatz into the first of Eqs.(3.1), we obtain the **Poisson equation**

$$\boxed{\Delta\phi = -4\pi\rho.} \quad (3.4)$$

Once this equation is solved, the problem is under control for a (Eq. (3.3)) will yield the electric field strength \mathbf{E} .

▷ Info. For a given point in space, \mathbf{x}_0 , the equation $\phi(\mathbf{x}) = \phi(\mathbf{x}_0)$ defines a two-dimensional surface in the vicinity of \mathbf{x}_0 – an **equipotential surface**. The electric field $\mathbf{E}(\mathbf{x}_0)$ points in the direction of the steepest descent of ϕ at \mathbf{x} . i.e. it is perpendicular to the equipotential surfaces of the problem. Once the potential is known we may, thus, easily visualize the electric field.

¹A subset $U \subset \mathbb{R}^d$ is simply connected if every closed curve in it can be smoothly contracted to zero. For example, the full space \mathbb{R}^3 is simply connected, but $\mathbb{R}^3 \setminus \{0, 0, \mathbb{R}\}$, i.e. space with the z -axis removed, is not. In the latter case, curves winding around the z -axis cannot be shrunk to a point.

²Proof: For a given point in space \mathbf{r} , an arbitrary fixed reference point \mathbf{r}_0 , and an equally arbitrary curve γ connecting \mathbf{r}_0 and \mathbf{r} , define $\phi_\gamma(\mathbf{r}) \equiv -\int_\gamma ds \cdot \mathbf{E}$. Then $\nabla\phi(\mathbf{r}) = -\mathbf{E}(\mathbf{r})$. To see that $\phi_\gamma \equiv \phi$ is independent of the choice of γ , compute $\phi_\gamma - \phi_{\gamma'} = \oint_{\gamma-\gamma'} ds \cdot \mathbf{E} = \int_{S(\gamma-\gamma')} d\sigma \mathbf{n} \cdot \nabla \times \mathbf{E} = 0$, where $\gamma - \gamma'$ is the closed curve obtained by concatenating the reverse of γ' with γ . Notice, however, that the surface construction requires that $\nabla \times \mathbf{E}$ be defined everywhere on $S(\gamma - \gamma')$, and this is where the condition of simple connectivity comes in. As an example, consider the vector field $\mathbf{v} = \frac{-y\mathbf{e}_x + x\mathbf{e}_y}{(x^2+y^2)}$, defined everywhere except on the z -axis, where $x = y = 0$. It is straightforward to verify that $\nabla \times \mathbf{v} = 0$. However, \mathbf{v} is not a gradient field. This can be checked by verifying that $\oint \mathbf{s} \cdot \mathbf{v} = 2\pi$ along any path enclosing the z -axis. The fact that the domain of definition of \mathbf{v} is not-simply connected invalidates the above construction (surfaces bounded by curves encircling the z -axis are pierced by the z -axis, i.e. $\nabla \times \mathbf{v}$ is not defined on all of them.)

Thus, our central objective will be to develop solution strategies for the Poisson equation. For later reference we note that in the case of a vanishing charge distribution, $\rho = 0$, the Poisson equation is denoted as **Laplace equation**

$$\Delta\phi = 0. \quad (3.5)$$

▷ **Math.** As a first useful application of the δ -function, we introduce the concept of **Green functions**. Suppose we want to solve a linear differential equation $Lf = g$, where $f, g \in \mathcal{C}^n(U \subset \mathbb{R}^d, \mathbb{R})$ are functions (f is the solution we are after) and L is a **linear differential operator**, i.e. a an expression of the type

$$L = h^{(0)}(\mathbf{x}) + \sum_i h_i^{(1)}(\mathbf{x})\partial_{x_i} + \sum_{ij} h_{ij}^{(2)}(\mathbf{x})\partial_{x_i x_j}^2, \quad (3.6)$$

where the h 's are some functions and the derivatives act on the argument function f . Notice that $L(c_1 f_1 + c_2 f_2) = c_1 Lf_1 + c_2 Lf_2$, i.e. L acts linearly in function space (hence the denotation *linear differential operator*.) Many problems in electrodynamics (and for that matter quantum mechanics) effectively reduce to the solution of linear partial differential equation of the type above. For example, the Poisson equation (3.4) belongs to this category ($L \leftrightarrow -\Delta$, $f \leftrightarrow \phi$, $g \leftrightarrow 4\pi\rho$.)

The weak spot of a linear differential equation is just its linearity. If f_1 and f_2 solve the equation for g_1 and g_2 , respectively, the function $f_1 + f_2$ will be a solution for the inhomogeneity $g_1 + g_2$. This feature suggests to first solve the equation for functions g which are as simple as possible. Arguably the simplest function at all is the δ -function: except for one point in space it vanishes everywhere, i.e. it is maximally structureless. We may employ the δ -function to represent an arbitrary inhomogeneity as

$$g(\mathbf{x}) = \int d^d x' \delta(\mathbf{x} - \mathbf{x}') g(\mathbf{x}'),$$

i.e. as a sum (integral) over point-localized inhomogeneities centered at \mathbf{x}' of unit-normalization ($\int \delta = 1$) each carrying a weight $g(\mathbf{x}')$. It is, thus, a good idea to first solve the equation

$$LG(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'), \quad (3.7)$$

where the derivatives in L act only on the argument \mathbf{x} . The function — a distribution, in fact — G is known as the **Green function** of the differential equation. Once the Green function is known, the general solution of the differential equation can be obtained as by summation (integration) over the prototype solutions G :

$$f(\mathbf{x}) = \int d^3 x' G(\mathbf{x}, \mathbf{x}') g(\mathbf{x}').$$

Indeed, by direct substitution, $Lf(\mathbf{x}) = \int d^3 x' LG(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') = \int d^3 x' \delta(\mathbf{x} - \mathbf{x}') g(\mathbf{x}') = g(\mathbf{x})$. The linearity of the problem thus enables us to break down the solution of the general problem into the (generally much simpler) solution of the point-inhomogeneity problem, plus a subsequent integration.

Now, one may ask whether Eq. (3.7) does have a unique solution at all. Referring for a substantiated discussion to below, we here merely note that the solution becomes unique once appropriate boundary conditions for $G(\mathbf{x}, \mathbf{x}')|_{\mathbf{x} \rightarrow \partial U}$ at the boundaries of the domain of definition are specified. For unspecified boundary conditions, the solution is not unique: For any G , the function $G + F$ will be a solution as well, provided F solves the homogeneous problem $LF(\mathbf{x}) = 0$. Differential

equations whose solution requires the specification of data at the boundaries of the domain of support are known as **boundary value problems**.

3.2.2 Boundary value problem I: infinite space

Potential created by a static charge distribution

We next specialize to the case of the Poisson equation. Using the notation of the previous section, $L = -\Delta$, $f = \phi$ and $g = 4\pi\rho$. In this section, we will consider the Poisson equation in infinite three-dimensional case, i.e. the domain of definition of the differential equation is $U = \mathbb{R}^3$. As for the boundary conditions to be imposed on Φ , we require — on physical grounds — that the electric field \mathbf{E} created by any spatially localized charge distribution vanish at infinity (for otherwise infinitely separated charges would exert a force on each other.) This condition is satisfied if $\phi(\mathbf{r} \rightarrow \infty) \rightarrow 0$, i.e. ϕ vanishes at the boundaries of U , a ball of infinite radius. (We will see below, that this condition is, in fact, necessary to ensure the vanishing of the electric field.)

According to the logics developed in the previous problem, the solution of the problem is (up to an integration) equivalent to the computation of the **Green function of electrostatics**,

$$\Delta G(\mathbf{x}, \mathbf{x}') = -\delta(\mathbf{x} - \mathbf{x}'). \quad (3.8)$$

Physically, $G(\mathbf{x}, \mathbf{x}')$ is proportional to the potential of a point charge at \mathbf{x}' felt at \mathbf{x} . Notice that this potential can depend only on the difference vector, $\mathbf{x} - \mathbf{x}'$, i.e. $G(\mathbf{x}, \mathbf{x}') = G(\mathbf{x} - \mathbf{x}')$.

We next discuss two strategies for solving Eq.(3.8):

Solution strategy I: (Physics inspired but very restrictive.) Without loss of generality, we chose coordinates such that $\mathbf{x}' = 0$. Physically, we expect $G(\mathbf{x}) = G(r)$ to depend only on the distance to the origin.³ To determine the r -dependence of $G(r)$ let us integrate the vector field $\nabla G(r) = \partial_r G(r) \mathbf{e}_r$ (which is proportional to the electric field created by the point charge at the origin, why?) over a sphere S of radius R centered at the origin $\mathbf{x} = 0$ (cf. the example on p ??): $\int_S d\sigma \mathbf{n} \cdot \nabla G = 4\pi R^2 \partial_R G(R) = \int_{V(S)} d^3x \underbrace{\nabla \cdot \nabla}_{\Delta} G = -\int_{V(S)} \delta(\mathbf{x}) = -1$.

Solving for G we obtain $\partial_R G(R) = -1/4\pi R^2$ which integrates to $G(R) = 1/4\pi R$. Switching back to vectorial coordinates we obtain the important result

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{4\pi} \frac{1}{|\mathbf{x} - \mathbf{x}'|}. \quad (3.9)$$

Before discussing this any further, let us discuss the much more sophisticated

³More formally, this feature follows from the fact that both $\delta(\mathbf{x})$ and the Laplace operator Δ are invariant under rotations in space. Thus, the solution to $\Delta G(\mathbf{x}) = \delta(\mathbf{x})$ will be rotationally invariant, too.

▷ Math. Recall that the **Fourier transform** of a function $f \in C^n(\mathbb{R}^d, \mathbb{R})$ is defined by

$$\tilde{f}(\mathbf{k}) = \frac{1}{(2\pi)^d} \int d^d x f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (3.10)$$

(provided the integral exists.) Note: to enlarge the class of functions that can be Fourier transformed, it is customary to generalize the definition above according to

$$\tilde{f}(\mathbf{k}) = \lim_{\epsilon \rightarrow 0} \frac{1}{(2\pi)^d} \int d^d x f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x} - \epsilon \mathbf{x}^2}.$$

Remarks: (a) every function $f(\mathbf{x})$ that grows sub-exponentially for large $|\mathbf{x}|$ can be transformed. (b) It is customary to suppress the convergence generating factor ϵ if it is not needed. (c) For notational simplicity, we will often denote the Fourier transform by f instead of \tilde{f} .

The function $f(\mathbf{x})$ is retrieved by the inverse transformation⁴

$$f(\mathbf{x}) = \int d^d k e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{f}(\mathbf{k}).$$

▷ Example: The function $f(\mathbf{x}) = 1$ can not be transformed in the restrictive sense (3.10). However

$$\tilde{f}(\mathbf{k}) = \lim_{\epsilon \rightarrow 0} \frac{1}{(2\pi)^d} \int d^d x f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x} - \epsilon \mathbf{x}^2} = \lim_{\epsilon \rightarrow 0} \frac{1}{(4\pi\epsilon)^{d/2}} e^{-\mathbf{k}^2/4\epsilon} = \delta(\mathbf{k}),$$

where the last identity follows by comparison with (the d -dimensional generalization of) (2.14). Indeed, for a general function the substitution of the inverse Fourier transform into (3.10) demands $f(\mathbf{k}) \stackrel{!}{=} \frac{1}{(2\pi)^d} \int d^d x \int d^d k e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} f(\mathbf{k}')$ which implies the important identity

$$\frac{1}{(2\pi)^d} \int d^d x e^{i\mathbf{k}\cdot\mathbf{x}} = \delta(\mathbf{k}). \quad (3.11)$$

Table 3.2.2 summarizes the most important features of the Fourier transform. (The proofs follow readily from Eq. (3.10).):

In the last line, the **convolution** of two functions is defined as

$$(f * g)(\mathbf{x}) = \int d^d x' f(\mathbf{x} - \mathbf{x}') g(\mathbf{x}').$$

Solution strategy II: (more laborious but highly generalizable) Consider a general linear

⁴As with the $\mathbf{x} \rightarrow \mathbf{k}$ direction of the transformation, the inverse transform should be understood as

$$f(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \int d^d k e^{i\mathbf{k}\cdot\mathbf{x} - \epsilon \mathbf{k}^2} \tilde{f}(\mathbf{k}).$$

function	trafo.	function	trafo.
$f(\mathbf{x})$	$\tilde{f}(\mathbf{k})$	$f^*(\mathbf{x})$	$\tilde{f}^*(-\mathbf{k})$
$f(\mathbf{x}) + g(\mathbf{x})$	$\tilde{f}(\mathbf{k}) + \tilde{g}(\mathbf{k})$	$af(\mathbf{x})$	$a\tilde{f}(\mathbf{k})$
$\partial_{x_j} f(\mathbf{x})$	$ik_j \tilde{f}(\mathbf{k})$	$-ix_j f(\mathbf{x})$	$\partial_{k_j} \tilde{f}(\mathbf{k})$
1	$\delta(\mathbf{x})$	$\delta(\mathbf{x})$	$(2\pi)^{-d}$
$f(\mathbf{x} + \mathbf{a})$	$e^{-i\mathbf{k}\cdot\mathbf{a}} \tilde{f}(\mathbf{k})$	$e^{-i\mathbf{x}\cdot\mathbf{a}} f(\mathbf{x})$	$\tilde{f}(\mathbf{k} + \mathbf{a})$
$f(a\mathbf{x})$	$ a ^{-d} \tilde{f}(\mathbf{k}/a)$	$(f * g)(\mathbf{x})$	$(2\pi)^d \tilde{f}(\mathbf{k}) \tilde{g}(\mathbf{k})$

Table 3.1: Important features of the Fourier transform

partial differential equation $Lf = g$ with *constant coefficients*, i.e. one where the differential operator L (cf. Eq. (3.6)) contains only constant functions $h = \text{const.}$. (Our Poisson equation belongs to this family.) Differential equations of this type can readily be solved⁵ by Fourier transformation:

Consider the equation (summation convention)

$$\left(h^{(0)} + h_i^{(1)} \partial_{x_i} + h_{ij}^{(2)} \partial_{x_i x_j} + \dots \right) f(\mathbf{x}) = g(\mathbf{x}).$$

Fourier transforming both sides of this equation we obtain $(h^{(0)} + ih_i^{(1)} k_i - h_{ij}^{(2)} k_i k_j + \dots) \tilde{f}(\mathbf{k}) = \tilde{g}(\mathbf{k})$ which is solved by $\tilde{f}(\mathbf{k}) = \tilde{g}(\mathbf{k}) / (h^{(0)} + ih_i^{(1)} k_i - h_{ij}^{(2)} k_i k_j + \dots)$. A solution of the equation is thus given by

$$f(\mathbf{x}) = \int d^d k e^{i\mathbf{k}\cdot\mathbf{x}} \frac{\tilde{g}(\mathbf{k})}{h^{(0)} + ih_i^{(1)} k_i - h_{ij}^{(2)} k_i k_j + \dots},$$

i.e. the problem has been reduced to the computation of an integral. We write ‘a’ solution instead of ‘the’ solution because (for unspecified boundary conditions) to any f one may add a solution f_0 of the homogeneous equation $Lf_0 = 0$. Fourier transforming this latter equation, we obtain $(h^{(0)} + ih_i^{(1)} k_i - h_{ij}^{(2)} k_i k_j + \dots) \tilde{f}_0(\mathbf{k}) = 0$ which has the solution (think why)

$$\tilde{f}_0(\mathbf{k}) = h(\mathbf{k}) \delta(h^{(0)} + ih_i^{(1)} k_i - h_{ij}^{(2)} k_i k_j + \dots).$$

We illustrate the general procedure on the example of the Poisson equation (3.8). Fourier transforming the equation, and setting $\mathbf{x}' = 0$ as before, we obtain the

$$\tilde{G}(\mathbf{k}) = \frac{1}{(2\pi)^3 k^2}.$$

⁵In the jargon of differential equations, ‘solution’ means transformation to a form that may still involve an integral operation.

The next and final task is to compute the inverse transform of this expression:

$$\begin{aligned} G(\mathbf{x}) &= \frac{1}{(2\pi)^3} \int d^3k \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{k^2} = \frac{1}{(2\pi)^2} \int_0^\pi \underbrace{\sin\theta d\theta}_{-d(\cos\theta)} \int_0^\infty k^2 dk \frac{e^{ikr \cos\theta}}{k^2} = \\ &= \frac{1}{2\pi^2} \int_0^\infty dk \frac{\sin kr}{kr} = \frac{1}{2\pi^2 r} \int_0^\infty dk \frac{\sin k}{k} = \frac{1}{4\pi r}, \end{aligned}$$

where in the second equality we introduced polar coordinates in \mathbf{k} -space, $\mathbf{k} \leftrightarrow (k, \phi, \theta)$ (with the vector \mathbf{x} lying on the axis $\theta = 0$) and in the final equality used that $\int_0^\infty dk \frac{\sin k}{k} = \pi/2$. Generalizing to arbitrary values of the reference vector \mathbf{x}' , we obtain $G(\mathbf{x} - \mathbf{x}') = 1/4\pi|\mathbf{x} - \mathbf{x}'|$ in agreement with our earlier result (3.9).

According to the general scheme developed on p 27, the potential created by a general charge distribution is given by $\phi(\mathbf{x}) = 4\pi \int d^3x' G(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}')$, or

$$\boxed{\phi(\mathbf{x}) = \int d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}} \quad (3.12)$$

Using that $\nabla|\mathbf{x} - \mathbf{x}'|^{-1} = -(\mathbf{x} - \mathbf{x}')/|\mathbf{x} - \mathbf{x}'|^3$, the electric field created by the system of charges is given by

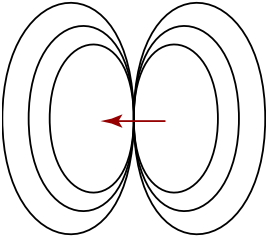
$$\mathbf{E}(\mathbf{x}) = \int d^3x' \rho(\mathbf{x}') \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3}. \quad (3.13)$$

▷ Example: The potential created by a set of N point charges q_i at positions \mathbf{x}_i is obtained by integration over $\rho(\mathbf{x}) = \sum_{i=1}^N q_i \delta(\mathbf{x} - \mathbf{x}_i)$,

$$\phi(\mathbf{x}) = \sum_{i=1}^N \frac{q_i}{|\mathbf{x} - \mathbf{x}_i|}. \quad (3.14)$$

▷ Example: Consider an electrostatic **dipole**, i.a. a system of two opposite charges of equal strength $\pm q/a$ at positions $\pm a\mathbf{n}'$, where \mathbf{n}' is a unit vector. (For convenience we place the center of the dipole at the coordinate $\mathbf{x} = 0$.) The corresponding charge distribution is given by $\rho(\mathbf{x}) = \frac{q}{a}(\delta(\mathbf{x} - a\mathbf{n}'/2) - \delta(\mathbf{x} + a\mathbf{n}'/2))$. Assuming that $|\mathbf{x}| \gg a$, the potential created by the dipole is given by

$$\phi(\mathbf{x}) = \frac{q}{a} \left(\frac{1}{|\mathbf{x} - a\mathbf{n}'/2|} - \frac{1}{|\mathbf{x} + a\mathbf{n}'/2|} \right) = \frac{q\mathbf{n}' \cdot \mathbf{x}}{|\mathbf{x}|^3} + \mathcal{O}((a^2/|\mathbf{x}|)^2)$$



In the limit $a \rightarrow 0$ — two infinitely charged and infinitesimally separated point charges — the source configuration becomes a point dipole. The product $qa' \equiv \mathbf{d}$ is denoted its **dipole moment**. The potential created by a point dipole at the coordinate origin is given by

$$\phi(\mathbf{x}) = \frac{\mathbf{d} \cdot \mathbf{x}}{|\mathbf{x}|^3}, \quad (3.15)$$

i.e. it is anisotropic and decays doubly as fast as the potential of a point charge (see the figure for a schematic of the equipotential lines of a dipole \mathbf{d} .) Taking the gradient of this expression, we obtain the electric field of the dipole,

$$\mathbf{E}(\mathbf{x}) = \frac{3\mathbf{n}(\mathbf{n} \cdot \mathbf{d}) - \mathbf{d}}{|\mathbf{x}|^3}, \quad (3.16)$$

where \mathbf{n} is the unit vector in \mathbf{x} -direction.

As a corollary, we note that our analysis above implies the identity,

$$\frac{1}{|\mathbf{x}|} \xrightarrow{\text{F.T.}} \frac{1}{2\pi^2 k^2}. \quad (3.17)$$

Using this formula we may solve the Poisson equation directly (i.e. without using Green functions) by Fourier transformation techniques: Fourier transformation of $\Delta\phi = -4\pi\rho$ yields $k^2\phi(\mathbf{k}) = 4\pi\rho(\mathbf{k})$ or $\phi(\mathbf{k}) = 4\pi k^{-2}\rho(\mathbf{k}) = (2\pi)^3(2\pi^2 k^2)^{-1}\rho(\mathbf{k})$. Now, we know that $(2\pi)^d$ times the product of two functions in Fourier space transforms to convolution of the real space representations. Thus,

$$\phi(\mathbf{x}) = \left(\frac{1}{|\mathbf{x}|} * \rho \right) (\mathbf{x}) = \int d^3x' \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}.$$

Energy considerations

There is no a priori rule how to attribute an ‘energy’ to an electromagnetic field. However, we *do* know of a mechanics-inspired prescription how to compute the energy of a system of (point) charges: Consider a system of N point charges at positions \mathbf{x}_i . To a configuration of $N = 1$ charges we assign the energy $E_1 = 0$.

▷ **Info.** There is, of course, some arbitrariness in assigning energy zero to a single non-vanishing charge in vacuum. This charge will create a finite electric field and it is, somehow, counter-intuitive to treat this system as one void of ‘energy’. However, every attempt to meaningfully assign a ‘self-energy’ to point charges in classical electrodynamics leads to unphysical divergences. The problem of the **infinite self-energy** finds its resolution only in quantum electrodynamics. All we can do in the classical context is to artificially normalize the infinite constant corresponding to the self-energy to zero.

The mechanical energy (i.e. the work to be done against the force field $\mathbf{F} = q_2\mathbf{E} = -q_2\nabla\phi$, where ϕ is the potential created by charge no.1) required to bring a second charge from infinity to its destination coordinate \mathbf{x}_2 is given by $E_2 = q_2\phi(\mathbf{x}_2 - \mathbf{x}_1) = q_2q_1/|\mathbf{x}_1 - \mathbf{x}_2|$. Bringing in a third charge against the potential of the two charges that have already been placed will cost the additional energy $q_3q_1/|\mathbf{x}_1 - \mathbf{x}_3| + q_3q_2/|\mathbf{x}_2 - \mathbf{x}_3|$, i.e. the total energy required to build up a three charge configuration is given by $E_3 = q_2q_1/|\mathbf{x}_1 - \mathbf{x}_2| + q_3q_1/|\mathbf{x}_1 - \mathbf{x}_3| + q_3q_2/|\mathbf{x}_2 - \mathbf{x}_3| = \sum_{1 \leq i < j \leq 3} q_iq_j/|\mathbf{x}_i - \mathbf{x}_j|$. Iterating this procedure, we conclude that the mechanical energy of an N -charge configuration is given by

$$E = \sum_{1 \leq i < j \leq N} \frac{q_iq_j}{|\mathbf{x}_i - \mathbf{x}_j|} = \frac{1}{2} \sum_{i \neq j}^N \frac{q_iq_j}{|\mathbf{x}_i - \mathbf{x}_j|}. \quad (3.18)$$

Thinking of a continuum charge distribution ρ as a (dense) accumulation of point charges, it is evident that the continuum generalization of this result, i.e. the **energy of a charge distribution** is given by

$$E = \frac{1}{2} \int d^3x d^3x' \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (3.19)$$

▷ Info. However, what happened to the exclusion rule $i \neq j$? The problem of self energies relates to the divergence of the denominators $1/|\mathbf{x} - \mathbf{x}'|$ as the points \mathbf{x} and \mathbf{x}' approach each other. (Gedanken experiment: Think of a point charge q as two charges $q/2$ infinitely closely spaced. The energy of this configuration, computed according to (3.18), diverges.) However, for any continuous charge distribution, the integral (3.19) is finite, i.e. there is no problem with the potentially dangerous regions $\mathbf{x} \simeq \mathbf{x}'$ where the arguments \mathbf{x} and \mathbf{x}' approach each other. To see this, consider the contribution to the integral from regions where $|\mathbf{x} - \mathbf{x}'| < a$ is small:

$$\int \int_{|\mathbf{x}-\mathbf{x}'|<a} d^3x d^3x' \frac{\rho(\mathbf{x})\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \simeq \int d^3x \rho^2(\mathbf{x}) \int_{|y|<a} d^3y \frac{1}{|y|} = \int d^3x \rho^2(\mathbf{x}) 4\pi \int_0^a y^2 \frac{dy}{y} = \int d^3x \rho^2(\mathbf{x}) 2\pi a^2,$$

which is finite: The 'smallness' of the three-dimensional volume element $y^2 dy \sin\theta d\theta d\phi$ overcompensates the divergence of the denominator. We thus conclude that (3.19) adequately describes the energy of the continuum charge distribution and that the problem of the self energy does not arise.

An interesting reformulation of the energy formula (3.19). Using Eq. (3.12),

$$\begin{aligned} E &= \frac{1}{2} \int d^d x \phi(\mathbf{x})\rho(\mathbf{x}) = -\frac{\epsilon_0}{8\pi} \int d^d x \phi(\mathbf{x})\Delta\phi(\mathbf{x}) \stackrel{(2.21)}{=} \frac{\epsilon_0}{8\pi} \int d^d x \nabla\phi(\mathbf{x})\nabla\phi(\mathbf{x}) = \\ &= \frac{1}{8\pi} \int d^3x \mathbf{D}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x}), \end{aligned}$$

where, for later reference, we have re-installed the dielectricity constant $\epsilon_0 = 1$. Remarkably, the last representation no longer makes reference to the matter-origin of the electric field; it

assigns an energy to the electric field *as such*. We will, thus, use

$$E = \frac{1}{8\pi} \int d^3x \mathbf{D} \cdot \mathbf{E}. \quad (3.20)$$

as a definition of the total **energy of the electric field**. The energy density, i.e., the energy per volume element is given by $u_E(\mathbf{x}) \equiv \frac{1}{8\pi}(\mathbf{D} \cdot \mathbf{E})(\mathbf{x})$. Later on we will see that many other characteristics of matter distributions — momentum, angular momentum, etc. — allow for a generalization to the electromagnetic context.

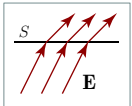
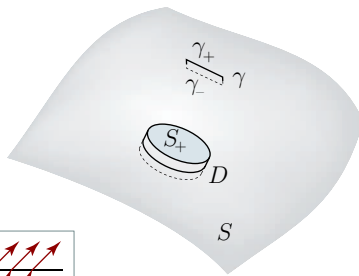
Surface charges and dipole layers

In many applications one is met with charge distributions residing on geometric structures of dimensionality lower than three — charged wires, or surfaces, planar accumulations of dipoles, etc. We here discuss two important subspecies, viz. surface charge distributions and dipole layers.

Surface charges: Consider a charged quasi-two dimensional surface. We idealize this object by a truly two-dimensional surface, S . (Exercise: invent a limiting process, similar in spirit to our discussion of the point dipole above, whereby the two-dimensional structure emerges from a charge distribution of finite thickness.) We define $\eta(\mathbf{x})d\sigma$ to be the charge residing in the surface element $d\sigma$ at \mathbf{x} . The electric field created by the surface can be obtained by obvious generalization of the general formula (3.13):

$$\mathbf{E}(\mathbf{x}) = \int_S d\sigma \eta(\mathbf{x}') \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^3}. \quad (3.21)$$

While Eq.(3.21) fully solves the problem, it does not tell us immediately how the electric field will actually behave in the immediate vicinity of the surface. Information of this type is conveniently obtained by running a few 'tests' based on the integral form of the Maxwell equations:



To this end, consider a little pill box D whose upper lid, S_+ (lower lid, S_-) is above (beneath) the surface S and whose height is assumed to be vanishingly small. We assume that the two surfaces S_{\pm} are aligned parallel to S and that they are so small that variations of the electric field along the lids S_{\pm} are negligible. However, crucially we anticipate jumps in the field vector upon crossing the surface, S . Under these conditions, the surface integral of \mathbf{E} over D can be evaluated as $\int_D \mathbf{E} \simeq \delta S(\mathbf{E}(\mathbf{x}_+) -$

$\mathbf{E}(\mathbf{x}_-)) \cdot \mathbf{n} \equiv \delta S \Delta \mathbf{E}_{\perp}(\mathbf{x})$, where δS is the area of S_{\pm} , $\mathbf{x}_{\pm} \in S_{\pm}$ are points infinitesimally above (below) the surface, \mathbf{n} is the unit normal vector to S (and S_{\pm}) at \mathbf{x} , and $\Delta \mathbf{E}_{\perp}(\mathbf{x})$ denotes the jump in the normal component of \mathbf{E} at \mathbf{x} . At the same time, by the Maxwell equation (2.23), the surface integral equals $\int_{V(D)} d^3x \nabla \cdot \mathbf{E} = 4\pi \int_{V(D)} d^3x \rho(\mathbf{x}) = 4\pi \int_{V(D) \cap S} d\sigma \eta(\mathbf{x}) \simeq$

$4\pi\delta S\eta(\mathbf{x})$. Comparing the two results we find

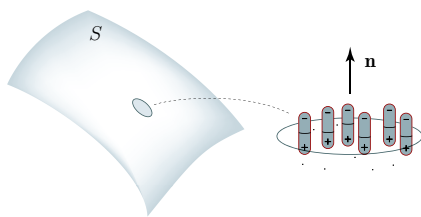
$$\Delta\mathbf{E}_\perp(\mathbf{x}) = 4\pi\eta(\mathbf{x}), \tag{3.22}$$

i.e. the component of the electric field normal to the surface jumps by an amount set by the surface charge.

To understand the behaviour of the tangential component, \mathbf{E}_\parallel , we integrate \mathbf{E} over a small closed contour γ comprising two stretches γ_\pm infinitesimally above/below S . Computing the line integral under the same assumptions as above, we obtain $\oint_\gamma d\mathbf{s} \cdot \mathbf{E} \simeq l(\mathbf{E}_\parallel(\mathbf{x}_+) - \mathbf{E}_\parallel(\mathbf{x}_-))$, where l denotes the length of γ_\pm and the contribution to the line integral from the stretches normal to S has been neglected. However, by Eq. (??), (or, equivalently, the condition $\nabla \times \mathbf{E} = 0$), this integral equals zero, i.e.

$$\Delta\mathbf{E}_\parallel(\mathbf{x}) = 0, \tag{3.23}$$

the tangential component of the electric field is continuous across S . Altogether, we obtain a profile of \mathbf{E} as indicated qualitatively in the figure.



dipole layers: Dipole layers are quasi two-dimensional configurations S of stacked dipoles. The individual dipole moments $\mathbf{d} \parallel \mathbf{n}$ are oriented normal to the surface (and uniformly oriented, $\mathbf{n} \cdot \mathbf{x} > 0$, say.)

▷ Info. Dipole layers are systems of considerable applied relevance: many **organic or biological molecules** carry a non-vanishing dipole element. Membranes formed by an effectively two-dimensional planar accumulation of such molecules represent dipolar layers.

Let $\eta(\mathbf{x}')\mathbf{n}d\sigma$ be the dipole moment carried by a surface element $d\sigma$ at \mathbf{x}' . According to (3.15), the potential created by this element is given by $d\phi(\mathbf{x}) = d\sigma\eta(\mathbf{x}')\frac{\mathbf{n}\cdot(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^3}$. The full potential created by the layer is given by

$$\phi(\mathbf{x}) = \int_S d\sigma\eta(\mathbf{x}')\frac{\mathbf{n}\cdot(\mathbf{x}-\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|^3}.$$

For a discussion of the potential ϕ , we refer to [1].



Figure 3.1: A charge/observation point configuration that qualifies for a multipole expansion. Discussion, see text.

(Cartesian) multipole expansion

Assume we observe a spatially localized charge distribution ρ from a very remote point \mathbf{x} . ('Remote' means that $|\mathbf{x}| \gg \max\{|\mathbf{x}'| \mid \rho(\mathbf{x}') \neq 0\}$, where it is assumed that the origin of the coordinate system is chosen somewhere in the support of ρ , i.e. $|\mathbf{x}|$ by far exceeds the extent of the charge distribution, cf. Fig. 3.1.) The first thing we will notice is, of course, the total charge carried by the distribution, $Q = \int d^3x' \rho(\mathbf{x}')$. Approaching the distribution or, equivalently, increasing the resolution at which we monitor its potential, we might notice that the charge is spread out inhomogeneously in space. E.g. even an overall neutral object might be intrinsically polarized. In a sense to be made precise below, we might, thus, attribute some vectorial dipole moment to the distribution. Approaching it further, we will be able to resolve yet finer details in the spatial pattern of charge. This simple picture suggests that the potential created by a charge distribution might be organized in a series whose individual terms contain information on the spatial structure of the charge at (a) an increasing level of detail whilst (b) decaying the more rapidly in space the higher their information content is. A series of this type indeed exists and it is called the **multipole expansion**.

The most frequently applied forms of multipole expansions applied in practice utilize non-cartesian coordinates and spherical harmonics, as discussed in the next section. We here consider a simpler version which nevertheless illustrates the main principles. Since $|\mathbf{x}| \gg |\mathbf{x}'|$ for all points \mathbf{x}' with $\rho(\mathbf{x}') \neq 0$, we may expand the distance function as

$$\frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} + \frac{x_i x'_i}{r^3} + \frac{1}{2} \frac{3x_i x_j - r^2 \delta_{ij}}{r^5} x'_i x'_j + \dots \quad (3.24)$$

Substituting this expansion into (3.12), we obtain the '**cartesian**' variant of the multipole expansion, $\phi = \phi^{(0)} + \phi^{(1)} + \phi^{(2)} + \dots \equiv \sum_{l=0}^{\infty} \phi^{(l)}$, where

$$\begin{aligned} \phi^{(0)}(\mathbf{x}) &= \frac{Q^{(0)}}{r}, & \phi^{(1)}(\mathbf{x}) &= Q_i^{(1)} \frac{x_i}{r^3}, & \phi^{(2)}(\mathbf{x}) &= \frac{Q_{ij}^{(2)}}{2} \frac{x_i x_j}{r^5}, \\ Q^{(0)} &= \int d^3x' \rho(\mathbf{x}'), \\ Q_i^{(1)} &= \int d^3x' \rho(\mathbf{x}') x'_i, \\ Q_{ij}^{(2)} &= \int d^3x' \rho(\mathbf{x}') (3x'_i x'_j - r'^2 \delta_{ij}), \end{aligned}$$

This defines the series we have been looking for: The potential has been represented as a series of ever more rapidly (the factor $r^{-(l+1)}$) decaying contributions. The entire information on the charge distribution is encoded in the **multipole coefficients** $Q^{(l)}$. In the next section, we will show that the number of these coefficients, and hence the information carried on the structure of the charge distribution, grows as $2l + 1$ with l .⁶

⁶There is $1 = 2 \times 0 + 1$ coefficient $Q^{(0)}$ for $l = 0$, $3 = 2 \times 1 + 1$ coefficients $Q_i^{(1)}$ for $l = 1$, and $5 = 2 \times 2 + 1$ coefficients $Q_{ij}^{(2)}$ for $l = 2$. To understand the number 5, note the symmetry $Q_{ij}^{(2)} = Q_{ji}^{(2)}$ and the 'sum rule' $\sum_i Q_{ii}^{(2)} = 0$.

To better understand the physical meaning of the series, let us investigate the first few contributions in more detail: The first contribution to the multipole expansion is given by

$$\phi^{(0)}(\mathbf{x}) = \frac{Q^{(0)}}{r},$$

where the **monopole moment** $Q^{(0)}$ is just the integrated charge of the distribution. This matches the expectation that from far away, the distribution looks like a structureless point charge.

To understand the meaning of the $l = 1$ term, let us define the **dipole moment** of the charge distribution as

$$\mathbf{d} \equiv \int d^3x' \rho(\mathbf{x}') \mathbf{x}'. \quad (3.25)$$

This vector generalizes the dipole moment introduced in the example on p 31 to arbitrary distributions. At the same time, its coefficients $d_i = Q_i^{(1)}$ equal the $l = 1$ coefficients of the expansion. Reflecting this equality, we call $Q_i^{(1)}$ the **dipole coefficients** of the expansion. The formula

$$\phi^{(1)}(\mathbf{x}) = Q_i^{(1)} \frac{x_i}{r^3} = \frac{\mathbf{d} \cdot \mathbf{x}}{r^3}, \quad (3.26)$$

then shows that to next to leading order, internal structure of a charge distribution becomes visible via its dipole moment. (This principle is often relevant in the observation of extraterrestrial objects such as exoplanets.)

⊖ ⊕ It is straightforward to construct charge distributions that carry neither monopole nor dipole moments (see the figure for an example.) For such distributions, the ⊕ ⊖ **quadrupole moments** $Q_{ij}^{(2)}$ define the dominant contribution to the series. The five independent moments are obtained by weighing the charge distribution by monomials of second order in the coordinates and integrating. The information stored in these coefficients describes the spatial structure of the distribution beyond the information carried by the dipole moment. The next order of the multipole moments, the seven independent **octupole moments** $Q_{ijk}^{(3)}$ create potentials decaying as r^{-4} and are, therefore, of lesser applied relevance. (The same applies to yet higher orders in the expansion.)

To summarize the utility of the multipole expansion,

- ▷ it allows one to express the potential $\phi(\mathbf{x})$ created by a charge distribution in relatively simple terms, viz. as a function of \mathbf{x} rather than by the general integral (3.12) (which has to be calculated for each point \mathbf{x} anew.)
- ▷ The multipole moments can be used to conveniently characterize the characteristics of a charge distribution. E.g. for a nucleus of approximately ellipsoidal shape, the dipole and quadrupole moments carry information on its deviations from a perfectly spherical geometry, etc.

- ▷ Notice, however, that the value obtained for the multipole moments depend on the choice of origin of the coordinate system. For example, for a point charge at \mathbf{x}_0 , all multipole moments will be non-vanishing unless we chose \mathbf{x}_0 to be the origin of the coordinate system (which would, of course, be the most reasonable choice.)

3.2.3 Poisson equation in non-cartesian coordinates

Whenever a problem in electrostatics possesses an in-built symmetry — rotational symmetry, cylindrical symmetry, etc. — its solution simplifies dramatically. To benefit maximally from the presence of a symmetry one ought to formulate the Poisson equation in problem adjusted coordinates, i.e. a coordinate system in which one or several coordinates remain invariant under the action of the symmetry operation. Coordinate systems of outstanding practical importance are the spherical and cylindrical coordinates.

We next formulate and solve the Poisson and Laplace equation in spherical coordinates. This discussion will provide the basis of the multipole expansion, a concept of importance in numerous applications in science and engineering ...

▷ Math. To prepare the analysis below, we need to introduce the notion of **complete systems of orthonormal functions**, a concept of great practical relevance in electrodynamics (plus quantum mechanics and many other areas of physics for that matter.)

A brief reminder of linear algebra: consider an n -dimensional real vector space V with scalar product $V \times V \rightarrow \mathbb{R}$, $(\mathbf{v}, \mathbf{w}) \mapsto \langle \mathbf{v}, \mathbf{w} \rangle$. Let $A : V \rightarrow V$ be an operator that is symmetric w.r.t. \langle , \rangle , i.e. $\forall \mathbf{v}, \mathbf{w} \in V : \langle \mathbf{v}, A\mathbf{w} \rangle = \langle A\mathbf{v}, \mathbf{w} \rangle$. Then, (i) A can be diagonalized (by an orthogonal transformation), i.e. there exists a basis of eigenstates \mathbf{v}_a , $a = 1, \dots, n$ with $A\mathbf{v}_a = \lambda_a \mathbf{v}_a$. (ii) We may chose the \mathbf{v}_a 's to be mutually orthogonal and normalized: $\langle \mathbf{v}_a, \mathbf{v}_b \rangle = \delta_{ab}$, i.e. $\{\mathbf{v}_a\}$ is an orthonormal basis. (iv) An arbitrary element $\mathbf{v} \in V$ can be expanded in the basis $\{\mathbf{v}_a\}$: $\mathbf{v} = \sum_{i=1}^n c_i \mathbf{v}_i$, where the coefficients c_i are given by $c_i = \langle \mathbf{v}_i, \mathbf{v} \rangle$.

Remarks: (a) The eigenbasis of A can be obtained by the Gram-Schmidt orthogonalization algorithm. (b) Nowhere do we require that the scalar product be the trivial one. I.e. $\langle \mathbf{v}, \mathbf{w} \rangle = v_i g_{ij} w_j$, where g_{ij} may be an arbitrary symmetric non-degenerate matrix. (d) Keep in mind: whenever we say "An operator A is symmetric", what we mean, in fact, is "it is symmetric w.r.t. a specifically chosen scalar product". This scalar product can, but need not, be the standard one. Only in that special case does the 'symmetric' operator assume the form of a symmetric matrix.

In functional analysis it is shown that, and under which conditions, the statements above afford a generalisation to the case of infinite dimensional vector spaces. We will here naively⁷ formulate this generalisation for those infinite dimensional spaces that are the most important with regard to applications in physics: function spaces.

As mentioned earlier, the space of differentiable⁸ functions $\mathcal{C}(U, \mathbb{R})$ defined on some subset $U \subset \mathbb{R}^d$ forms a vector space. In addition to differentiability we will usually also require integrability, i.e. we will consider functions $f \in L^2(U)$, where $L^2(U)$ is the space of square integrable functions

⁷... means without discussing the crucial convergence criteria required to firmly establish the diagonalisability of symmetric operators.

⁸In a somewhat sloppy notation we will use the symbol \mathcal{C} to denote spaces of functions that are as often differentiable as is required by the problem at hand.

($\int_U d^d x f(\mathbf{x})^2$ exists.) To keep the notation slim we will denote the intersection $L^2(U) \cap \mathcal{C}(U, \mathbb{R})$ again by $L^2(U)$. Remark: to bridge the gap between function spaces and finite dimensional vector spaces, imagine a function as the limit of a finite-dimensional object: We may approximate any continuous function $f : U \rightarrow \mathbb{R}$ by introducing a fine discretization (a 'grid') in U and assigning to each of its base points $\mathbf{x}_i \in U$ the discrete value $f(\mathbf{x}_i)$. In the limit of an infinitely dense grid, the discrete vector formed by all the values $\{f(\mathbf{x}_i)\}$ contains information equivalent to the full function. We may, thus, approximate any continuous functions by a sufficiently high-dimensional conventional vector. This analogy is often useful.

Now, the vector space $L^2(U)$ can be equipped with a real scalar product. Indeed, it is straightforward to verify that

$$\begin{aligned} \langle , \rangle : L^2(U) \times L^2(U) &\rightarrow \mathbb{R}, \\ (f, g) &\mapsto \langle f, g \rangle \equiv \int_U d^d x f(\mathbf{x})g(\mathbf{x}) \end{aligned}$$

satisfies the defining equations of a **scalar product**. (Do it!)

Importantly, however, the 'standard' scalar product above is not always the natural one. By way of example, consider the space of functions defined on the two-sphere S^2 . Let us parameterize S^2 by the standard set of coordinates, $U = \{(\theta, \phi) \in [0, \pi] \times [0, 2\pi]\}$.⁹ The 'natural' scalar product on the two-sphere is given by

$$\langle f, g \rangle = \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta f(\theta, \phi)g(\theta, \phi), \quad (3.27)$$

i.e. it accounts for the surface area element of the sphere.¹⁰

Now, consider a linear differential operator D acting in $L^2(U)$. (Alluding to the discretised picture above, we may think of D as a high-dimensional matrix; approximating derivative operations by finite differences, D does, literally, assume the form of a matrix acting on the discrete approximation of functions.) An interesting situation arises if D is symmetric w.r.t. the scalar product on the function space, i.e. if $\langle f, Dg \rangle = \langle Df, g \rangle$, or

$$\int d^d x f(\mathbf{x})\tilde{g}(\mathbf{x})(Dg)(\mathbf{x}) = \int d^d x (Df)(\mathbf{x})\tilde{g}(\mathbf{x})g(\mathbf{x}),$$

where \tilde{g} is the function defining the metric. Under these circumstances, the operator D possesses a complete system of **eigenfunctions** $\{f_n\}$, i.e. functions obeying the (differential) equation $Df_n(\mathbf{x}) = \lambda_n f_n(\mathbf{x})$. These functions are mutually orthogonal and can be normalized: $\langle f_n, f_{n'} \rangle = \delta_{nn'}$. Functions $f \in L^2(U)$ can be expanded in the set $\{f_n\}$,¹¹

$$f(\mathbf{x}) = \sum_n \langle f, f_n \rangle f_n(\mathbf{x}).$$

⁹Critical readers will object that U does not parameterize the full sphere; a line from the north pole to the south pole is excluded. This, however, does not play a role in our present discussion.

¹⁰Yet more generally, one might even consider scalar products of the form $\langle f, g \rangle = \int d^d x d^d x' f(\mathbf{x})\tilde{g}(\mathbf{x}, \mathbf{x}')g(\mathbf{x}')$, where $\tilde{g}(\mathbf{x}, \mathbf{x}') = \tilde{g}(\mathbf{x}', \mathbf{x})$ is a symmetric function of two arguments. However, constructs of this type usually do not appear in applications.

¹¹In cases where U is an unbounded set, the set of eigenvalues $\{\lambda_n\}$, the so-called **spectrum** of the operator D , may become (partially) dense. In this case, denoting the eigenfunctions $f(k)$, where k is a continuous index, the expansion assumes the form of an integral $f(\mathbf{x}) = \int dk \langle f(k), f \rangle f(k)$. The orthonormalisation condition assumes the form $\langle f(k), f(k') \rangle = \delta(k - k')$. (Why?)

▷ Example: Let $U = [0, 1]$. We consider the space of functions $L^2(U)$ subject to the constraint $f(0) = f(1) = 0$ ('Dirichlet boundary conditions'). Let $\langle f, g \rangle \equiv \int_0^1 dx f(x)g(x)$ be the standard metric and consider the differential operator $D = -d_x^2$. The eigenfunctions of this operator are given by $f_n(x) = \sqrt{2} \sin(\pi n x)$. They are orthonormalized, $\int_0^1 dx f_n(x) f_{n'}(x) = \delta_{nn'}$ and the expansion

$$f(x) = \sqrt{2} \sum_n c_n \sin(\pi n x), \quad c_n = \sqrt{2} \int_0^1 dx f(x) \sin(\pi n x)$$

is the familiar **Fourier series** expansion.

▷ Example: Let $U = [-1, 1]$ and $D = -d_x(1 - x^2)d_x$. The operator D is symmetric w.r.t. the standard scalar product on U . (Check it!) Its eigenfunctions $DP_l(x) = \lambda_l P_l(x)$ are known as **Legendre polynomials**. The first few Legendre polynomials are given by

$$\begin{aligned} P_0(x) &= 1 & P_1(x) &= x, \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) & P_3(x) &= \frac{1}{2}(5x^3 - 3x), \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) & \dots \end{aligned}$$

The generalization to polynomials of arbitrary degree is given by **Rodrigues' formula**

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l, \quad \lambda_l = l(l + 1). \quad (3.28)$$

Breaking with the general rule of scalar product normalization, the Legendre polynomials are normalized by $P_l(1) = 1$.

All we said below generalizes to the case of complex vector spaces. A **complex scalar product** on a finite dimensional complex vector space is defined by $\langle \mathbf{v}, \mathbf{w} \rangle = v_i^* g_{ij} w_j$, where $g = \{g_{ij}\}$ is an hermitian matrix: $g = g^\dagger$. Notice that $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle^*$. A general hermitian operator A , $\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle$ can (i) be diagonalized by an unitary transformation, i.e. there exists a basis of eigenstates $\{\mathbf{v}_a\}$, with $A\mathbf{v}_a = \lambda_a \mathbf{v}_a$, where $\lambda_a \in \mathbb{R}$. Again, we may chose these states to be orthonormal $\langle \mathbf{v}_a, \mathbf{v}_b \rangle = \delta_{ab}$. An arbitrary element $\mathbf{v} \in V$ can be expanded in the basis $\{\mathbf{v}_a\}$: $\mathbf{v} = \sum_{i=a}^n c_a \mathbf{v}_a$, where the coefficients c_a are given by $c_a = \langle \mathbf{v}_a, \mathbf{v} \rangle$.

The generalization to the infinite dimensional case is obvious: consider the **complex function space**, $L^2(U, \mathbb{C})$, i.e. the space of square integrable functions $f : U \rightarrow \mathbb{C}$. We define a scalar product by

$$\langle f, g \rangle = \int_U d^d x \tilde{g}(\mathbf{x}) f^*(\mathbf{x}) g(\mathbf{x}),$$

where the weight function \tilde{g} is real (why? Think of \tilde{g} as the generalization of a diagonal hermitian matrix.) A linear differential operator D is hermitian if $\forall f, g \in L^2(U, \mathbb{C}) : \langle f, Dg \rangle = \langle Df, g \rangle$. The eigenfunctions, f_n , of an hermitian differential operator form an (orthonormalizable) complete set. An arbitrary function can be expanded in the basis $\{f_n\}$ as $f = \sum \langle f_n, f \rangle f_n$.

▷ Example: On the interval $U = [0, 2\pi[$ consider the space of functions $f : U \rightarrow \mathbb{C}$ with 'periodic boundary conditions', $f(0) = f(x \rightarrow 2\pi)$. We chose the standard complex scalar product

$\langle f, g \rangle = \int_0^{2\pi} dx f^*(x)g(x)$. The orthonormalized eigenfunctions of the differential operator $D = -d_x^2$, $f_n(x) = (2\pi)^{-1/2} \exp(inx)$, $n \in \mathbb{Z}$ form a complete set. The expansion

$$f(x) = \frac{1}{(2\pi)^{1/2}} \sum_n c_n e^{inx}, \quad c_n = \frac{1}{(2\pi)^{1/2}} \int dx e^{-inx} f(x)$$

is the familiar **Fourier series expansion** on U .

The table below summarizes a number of relevant identities on linear symmetric operators in finite and infinite-dimensional vector spaces.

	finite dimensional	function spaces
real scalar product	$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_i v_i g_{ij} w_j$	$\langle f, g \rangle = \int d^d x \tilde{g}(\mathbf{x}) f(\mathbf{x}) g(\mathbf{x})$
complex scalar product	$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_i v_i^* g_{ij} w_j$	$\langle f, g \rangle = \int d^d x \tilde{g}(\mathbf{x}) f^*(\mathbf{x}) g(\mathbf{x})$
sym. of linear operator	$\langle \mathbf{v}, A\mathbf{w} \rangle = \langle A\mathbf{v}, \mathbf{w} \rangle$	$\langle f, Dg \rangle = \langle Df, g \rangle$
eigenvector/eigenfunction	$A\mathbf{v}_a = \epsilon_a \mathbf{v}_a$	$Df_n = \lambda_n f_n$
orthonormalisation	$\langle \mathbf{v}_a, \mathbf{v}_b \rangle = \delta_{ab}$	$\langle f_n, f_m \rangle = \delta_{nm}$
expansion	$\mathbf{v} = \sum_a \langle \mathbf{v}_a, \mathbf{v} \rangle \mathbf{v}_a$	$f = \sum \langle f_n, f \rangle f_n$

Table 3.2: Identities on complete set of states in finite and infinite dimensional vector spaces.

Solution of the Laplace equation (spherical coordinates)

Consider the Laplace operator in spherical coordinates,

$$\Delta = \Delta^\# + r^{-2} \Delta^s, \quad \Delta^\# = \frac{1}{r} \partial_r^2 r, \quad (3.29)$$

$$\Delta^s = \frac{1}{\sin \theta} \partial_\theta \sin \theta \partial_\theta + \frac{1}{\sin^2 \theta} \partial_\phi^2, \quad (3.30)$$

where the derivatives act on everything to the right. Our first objective is to compute the general solution of the Laplace equation

$$\Delta \Phi(r, \theta, \phi) = 0,$$

for unspecified boundary conditions. In spite of its formidable appearance, the Laplace operator has one weak spot: it is **separable**, i.e. the sum of three pieces each of which contains only

derivatives w.r.t. one of the three coordinates. This suggests to represent the solution as a product,

$$\Phi(r, \theta, \phi) = U(r)P(\theta)Q(\phi).$$

Substituting this ansatz into the Laplace equation and multiplying by r^2/UPQ , we obtain

$$U^{-1}(r^2\Delta^\#)U = -(PQ)^{-1}\Delta^sPQ.$$

Now the l.h.s./r.h.s. of this equation depends only on the variables $r/(\theta, \phi)$. The only way to satisfy this condition (think about it!) is to require that the two sides of the equation are equal *and* constant. Denoting this constant by $l(l+1)$,¹² we obtain the two equations

$$\Delta^\#U = l(l+1)r^{-2}U, \quad (3.31)$$

$$\Delta^sPQ = -l(l+1)PQ. \quad (3.32)$$

The first of these equations is controlled by the **radial component of the Laplace operator**. Substituting the monomial ansatz $U \sim r^\alpha$ into the radial equation, we obtain the solution

$$U = a^+r^l + a^-r^{-(l+1)}. \quad (3.33)$$

The presence of two undetermined constants signals that we have, indeed, found the most general equation of the (second order) differential equation for U . Notice that the first/second term on the r.h.s. diverges for $r \rightarrow \infty/r \rightarrow 0$, i.e. the choice of the constants a^\pm will be dictated by the boundary conditions imposed on the Laplace equation for large/small values of the radial coordinate.

Turning to the **angular component of the Laplace operator**, we may apply an argument similar to the one used above to separate the problem into the solution of two ordinary differential equations: multiplication of Eq. (3.32) by $\sin^2\theta PQ$ obtains

$$P^{-1}(\sin\theta\partial_\theta\sin\theta\partial_\theta + l(l+1)\sin^2\theta)P = -Q^{-1}\partial_\phi^2Q,$$

i.e. an equation whose l.h.s./r.h.s. depends only on θ/ϕ . Requiring constancy of both sides, we obtain the two ordinary differential equations

$$d_\phi^2Q + m^2Q = 0, \quad (3.34)$$

$$[(\sin\theta)^{-1}d_\theta\sin\theta d_\theta + l(l+1) - m^2(\sin\theta)^{-2}]P = 0, \quad (3.35)$$

where $m^2 = \text{const.}$. The first of these equations is solved by $Q = \exp(im\phi)$. In order for Q to be single-valued, $Q(\phi \rightarrow 2\pi) = Q(0)$, we must require $m \in \mathbb{Z}$.¹³ Changing variables, $x = \cos\theta$, the remaining, polar equation assumes the form

$$d_x [(1-x^2)d_xP] + l(l+1) - (1-x^2)^{-1}m^2] P = 0.$$

¹²Later on, we shall see that l is integer. The positivity of the constant $l(l+1) \geq 0$ is required by the fact that the operator $-r^2\Delta^\#$ (regarded as an operator acting in the space of square integrable functions) is positive definite.

¹³Notice that the set of functions $\{Q_m\}$ is a complete set on the interval $[0, 2\pi]$, cf. the example on p 40.

For $m = 0$, this equation collapses to the differential equation $[d_x(1-x^2)d_x + l(l+1)]P = 0$ discussed in an example on p 40; its solutions are the Legendre polynomials $P_l(x) \equiv P_l^{m=0}$.

As for the general case, $m \neq 0$, it turns out that solutions that are finite on the interval $[-1, 1]$ exist only if (a) l is integer and (b) limited to values $l \geq |m|$. The solutions corresponding to the allowed values $m = -l, -l+1, \dots, l$ of the azimuthal index, P_l^m , are called **Legendre functions**. For positive m they are defined by

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x). \quad (3.36)$$

The solutions for negative values of m are given by $P_l^{m<0} = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m$.

Summarizing, we have found that the so-called **spherical harmonics**

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) \exp(im\phi) \quad (3.37)$$

solve the angular part of the problem, $\Delta^s Y_{lm} = -l(l+1)Y_{lm}$. A number of remarks on this important family of functions:

We first note that the angular part of the Laplace operator, Δ^s is symmetric w.r.t. the natural scalar product on the two-sphere, (3.27). (prove it!) This means that its eigenfunctions, viz. the family of functions $Y_{lm} \propto P_l^m(\cos \theta) Q_m(\phi)$ form a complete set of orthogonal functions on the sphere. Indeed, one can prove that

$$\int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta Y_{lm}^*(\theta, \phi) Y_{l'm'}(\theta, \phi) = \delta_{ll'} \delta_{mm'}. \quad (3.38)$$

The unit-normalization of the r.h.s. determines the normalization factor in (3.37). The eigenvalues $\lambda_l = -l(l+1)$ are $(2l+1)$ -fold degenerate, i.e. each Y_l^m , $m = -l, \dots, l$ has eigenvalue $l(l+1)$. Arbitrary functions on the sphere can be expanded in spherical harmonics as¹⁴

$$f(\theta, \phi) = \sum_{lm} a_{lm} Y_{lm}(\theta, \phi), \quad a_{lm} = \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta Y_{lm}^*(\theta, \phi) f(\theta, \phi). \quad (3.39)$$

Eq. (3.39) entails the completeness relation

$$\sum_{l=0}^{\infty} \sum_{m=-l}^l Y_{lm}^*(\theta', \phi') Y_{lm}(\theta, \phi) = \delta(\phi - \phi') (\sin \theta)^{-1} \delta(\theta - \theta'). \quad (3.40)$$

¹⁴The possibility to expand functions in this way motivates the denotation 'spherical harmonics': 'spherical' for we are on the sphere, 'harmonics' because the functions Y_{lm} assume the role of the harmonic functions $\sim \exp(ikx)$ on flat integration spaces.

For the sake of concreteness, the explicit form of a few spherical harmonics is given below:

$$\begin{aligned}
 l = 0 & : Y_{00} = \frac{1}{\sqrt{4\pi}}, \\
 l = 1 & : Y_{1,0} = \sqrt{\frac{3}{4\pi}} \cos \theta, \\
 & Y_{1,\pm 1} = \sqrt{\frac{3}{8\pi}} \sin \theta \exp(\pm i\phi), \\
 l = 2 & : Y_{2,0} = \sqrt{\frac{5}{16\pi}} (3 \cos^2 \theta - 1), \\
 & Y_{2,\pm 1} = \mp \sqrt{\frac{15}{8\pi}} \cos \theta \sin \theta \exp(\pm i\phi), \\
 & Y_{2,\pm 2} = \mp \sqrt{\frac{15}{32\pi}} \sin^2 \theta \exp(\pm 2i\phi). \tag{3.41}
 \end{aligned}$$

For later reference, we notice that

$$Y_{l0}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi}} P_l(\cos \theta), \quad Y_{lm}(0, \phi) = \delta_{m,0} \sqrt{\frac{2l+1}{4\pi}}. \tag{3.42}$$

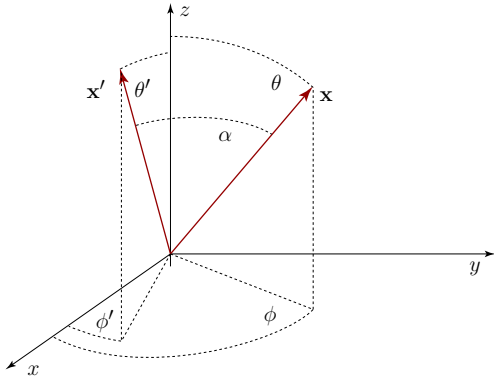
Combining Eq.(3.37) with our earlier result (3.33) for the radial function, we conclude that the **general solution of the Laplace equation in spherical coordinates** is given by

$$\Phi(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l (a_{lm}^+ r^l + a_{lm}^- r^{-(l+1)}) Y_{lm}(\theta, \phi), \tag{3.43}$$

where the value of the constants a^\pm is determined by the boundary conditions of the problem.

Addition theorem

At this point, it is a good idea to take a step back and ask where our so far analysis has got us: Firstly, we have not yet solved the Poisson equation — the prime information carrier on a source-ful system — in spherical coordinates. Second, it has not yet become clear enough why it is worthwhile to invest the considerable effort to formulate the theory in spherical coordinates at all. In this and the following section, we will address these two remaining issues in turn.



Rather than solving the **Poisson equation** directly in spherical coordinates, we will build on previous results, i.e. use that the Green function of electrostatics (in unbounded space) is given by Eq.(3.9). In the math block below we prove that the Green function of infinite-space electrostatics, characteristic distance function $G(\mathbf{x}, \mathbf{x}') = (4\pi|\mathbf{x} - \mathbf{x}'|)^{-1}$ affords the expansion

$$G(\mathbf{x}, \mathbf{x}') = \sum_{l=0}^{\infty} \frac{r_{<}^l r_{>}^{-(l+1)}}{2l+1} \sum_{m=-l}^l Y_{lm}^*(\theta, \phi) Y_{lm}(\theta', \phi'), \quad (3.44)$$

where $r_{<}/r_{>}$ denotes the smaller/larger of the radial arguments r and r' . Eq.(3.44) is known as the **addition theorem**. In the next section we will apply it to demonstrate the utility of the spherical harmonics in applied electrostatics.

▷ **Math.** Before turning to the actual **proof of the addition theorem**, notice that the structure of the r.h.s. of Eq. (3.44) is essentially fixed by the condition that for $\mathbf{x} \neq \mathbf{x}'$ the l.h.s. solves the Laplace equation, $\Delta_{\mathbf{x}}|\mathbf{x} - \mathbf{x}'|^{-1} = \Delta_{\mathbf{x}'}|\mathbf{x} - \mathbf{x}'|^{-1} = 0$, where $\Delta_{\mathbf{x}/\mathbf{x}'}$ is the Laplace operator acting on the argument \mathbf{x}/\mathbf{x}' . This means that as a function of both arguments \mathbf{x} and \mathbf{x}' , the r.h.s. must be of the structure (3.43). To avoid divergences at large and small values of the arguments, resp., the expansion in the larger/smaller of the two radial coordinates must not engage the dangerous powers $r^l/r^{-(l+1)}$. This fixes the structure of the radial term $r_{<}^l r_{>}^{-(l+1)}$.

However, what the schematic argument above cannot tell us is the precise form of the expansion coefficients. To actually prove the theorem, we consider the Green function of electrostatics as defined by (3.8). Let us assume that G admits an expansion of the type

$$G(\mathbf{x}, \mathbf{x}') = \sum_{l=0}^{\infty} \sum_{m=-l}^l g_l(r, r') Y_{lm}^*(\theta, \phi) Y_{lm}(\theta', \phi'). \quad (3.45)$$

Substituting this ansatz into the defining equation, using the polar representation, $\Delta = \Delta^{\#} + r^{-2}\Delta^s$, and that $\Delta^s Y_{lm} = -l(l+1)Y_{lm}$ we obtain

$$\sum_{l=0}^{\infty} \sum_{m=-l}^l \left(\Delta^{\#} - r^{-2}l(l+1) \right) g_l(r, r') Y_{lm}^*(\theta, \phi) Y_{lm}(\theta', \phi') = -\delta(\mathbf{x} - \mathbf{x}').$$

Now, $\delta(\mathbf{x} - \mathbf{x}') = (r-r')^{-2} \delta(r-r') \delta(\phi-\phi') (\sin(\theta-\theta'))^{-1} \delta(\theta-\theta')$ ^(3.40) $= r^{-2} \delta(r-r') \sum_{l=0}^{\infty} \sum_{m=-l}^l Y_{lm}^*(\theta', \phi') Y_{lm}(\theta, \phi)$. Substituting this representation of the δ -function, we obtain

$$\sum_{l=0}^{\infty} \sum_{m=-l}^l \left[\left(\Delta^{\#} - r^{-2}l(l+1) \right) g_l(r, r') + \delta(r-r') \right] Y_{lm}^*(\theta, \phi) Y_{lm}(\theta', \phi') = 0.$$

The completeness of the functions Y_{lm} implies¹⁵ the vanishing of the 'coefficients' [...] for all values

¹⁵This can be seen by multiplying the equation by $Y_{l'm'}^*(\theta', \phi')$ and integrating (θ', ϕ') over the spherical measure. Eq.(3.38) then implies [...] $Y_{lm}(\theta, \phi) = 0$ which means that the prefactor [...] = 0.

of (l, m) separately:

$$\left(-\Delta^\# + r^{-2}l(l+1)\right)g_l(r, r') = r^{-2}\delta(r - r').$$

Consider this equation for $r \neq r'$. Then (cf. Eq.(3.33)), the function $g_l(r, \cdot)$ must be a solution of the radial Laplace equation. We thus conclude that

$$\begin{aligned} r < r' : \quad g_l(r, r') &= a_{1,l}^+ r^l + a_{1,l}^- r^{-(l+1)}, \\ r > r' : \quad g_l(r, r') &= a_{2,l}^+ r^l + a_{2,l}^- r^{-(l+1)}, \end{aligned}$$

where the coefficients will be functions of the second argument, r' . (Due to the presence of a δ -function on the r.h.s. of the differential equation, we do not expect the coefficients to be continuous across the point $r = r'$.) Since, however, the coefficients do not depend on the angular variables, we may determine their value by considering specific configurations of (θ, \dots, ϕ') . Let us choose, then, $\theta = \dots = \phi' = 0$. (Vectors \mathbf{x} and \mathbf{x}' aligned and pointing in z -direction.) Then $|\mathbf{x} - \mathbf{x}'|^{-1} = |r - r'|^{-1}$ becomes a scalar distance function. For both $r > r'$ and $r < r'$, this function be expanded in a power series: $|r - r'|^{-1} \stackrel{r > r'}{\cong} r^{-1} \sum_{l=0}^{\infty} (r'/r)^l$ and $|r - r'|^{-1} \stackrel{r < r'}{\cong} r'^{-1} \sum_{l=0}^{\infty} (r/r')^l$. Substituting these series into the ansatz (3.45), using (3.42), and (3.9), we obtain

$$\frac{1}{4\pi} \sum_l r^l r'^{-(l+1)} = \sum_l (a_{1,l}^+ r^l + a_{1,l}^- r^{-(l+1)}) \frac{2l+1}{4\pi},$$

and a similar formula for $r > r'$. Comparison of the two sides of the equation leads to the identification $a_{1,l}^+ = r'^{-(l+1)}/(2l+1)$, $a_{2,l}^- = r^l/(2l+1)$ and $a_{2,l}^+ = a_{1,l}^- = 0$. Summarizing these results in a single formula, we arrive at

$$g_l(r, r') = \frac{r_{<}^l r_{>}^{-(l+1)}}{2l+1}.$$

Finally, substituting the function g back into our ansatz for the Green function and remembering Eq.(3.9) we obtain the addition theorem.

3.2.4 Spherical multipole Expansion

The spherical harmonics and the addition theorem (3.44) provide us with the means to define a spherical version of the multipole expansion. This is important because it is tailored to situations with spherical or cylindrical geometry and defines the multipole expansion most frequently used in practice (for example in atomic and nuclear physics, in accelerator engineering, and many other disciplines.)

Conceptually, the spherical multipole expansion follows the same logics as that formulated in section 3.2.2. The difference is that this time we start out from the addition theorem discussed in the previous section: Substituting Eq.(3.44) into (3.12), we obtain

$$\begin{aligned} \phi(\mathbf{x}) &= 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{r^{-(l+1)}}{2l+1} Y_{lm}(\theta, \phi) q_{lm}, \\ q_{lm} &= \int_0^{\infty} r'^2 dr' \int_0^{2\pi} d\phi' \int_0^{\pi} \sin \theta' d\theta' \rho(r', \phi', \theta') r'^l Y_{lm}^*(\theta', \phi'), \end{aligned} \tag{3.46}$$

where the coefficients q_{lm} define the **spherical multipole moments** of the charge distribution. The potential is now represented as a series of ever more rapidly (the factor $r^{-(l+1)}$) decaying contributions, and the entire information on the charge distribution is encoded in the multipole coefficients q_{lm} . The number of these coefficients, and hence the information carried on the structure of the charge distribution, grows as $2l + 1$ with l .

Since $Y_{00} = (4\pi)^{-1/2}$, the first contribution to the multipole expansion is given by

$$\phi^{(0)}(\mathbf{x}) = (4\pi)^{1/2} \frac{q_{00}}{r}.$$

As in the cartesian case, the **monopole moment** $q_{00} = (4\pi)^{-1/2}Q$ is but $((4\pi)^{-1/2} \times)$ the integrated charge of the distribution.

To understand the $l = 1$ term, first consider Eq. (3.41) and note that the $l = 1$ spherical harmonics can be obtained from the representation of a unit vector in spherical coordinates $\mathbf{n} \equiv (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T$ as

$$Y_{1,m} = \sqrt{\frac{3}{4\pi}} \sum_{a=1}^3 U_{m,a} n_a, \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i & 0 \\ 0 & 0 & \sqrt{2} \\ 1 & -i & 0 \end{pmatrix},$$

where U is unitary, $U^\dagger U = \mathbf{1}$. Comparing to the definition of the Cartesian dipole moment (3.25), we conclude that

$$\begin{pmatrix} q_{1,-1} \\ q_{1,0} \\ q_{1,+1} \end{pmatrix} = \sqrt{\frac{3}{4\pi}} U^* \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix},$$

i.e. up to the (complex conjugate of the) unitary transformation U , the three component vector $\mathbf{q}_1 \equiv \{q_{1,m}\}$ equals the dipole moment of the distribution. Substituting these relations into the definition of the $l = 1$ term, we find

$$\phi^{(1)}(\mathbf{x}) = \frac{4\pi}{3} \sum_{m=-1}^1 r^{-2} Y_{1,m}(\theta, \phi) q_{1,m} = r^{-3} (U_{ma}^* x_a) (U_{mb} d_b) = \frac{\mathbf{d} \cdot \mathbf{x}}{r^3},$$

where in the last step the unitarity of U was used. This shows how the $l = 1$ term represents the dipole moment of the distribution.

Similar reasoning may be applied to work out the relation between the spherical $l = 2$ terms $q_{2,m} = \int d^3 x' \rho(\mathbf{x}') r'^2 Y_{2,m}$ and the cartesian **quadrupole moments**. For example, applying (3.41) to express the spherical harmonics in cartesian coordinates one finds,

$$q_{22} = \sqrt{\frac{15}{32\pi}} \int d^3 x' (x'_1 - ix'_2)^2 \rho(\mathbf{x}'),$$

i.e. the integral of a weight function quadratic in the components of \mathbf{x}' multiplied by the charge distribution. The four other quadrupole moments can be represented in a similar

manner. (Due to $Y_{lm} = Y_{l-m}^*$ it suffices to compute the coefficients q_{lm} for positive m .) This shows that, up to linear combination, the $l = 2$ moments are the quadrupole moments of the distribution.

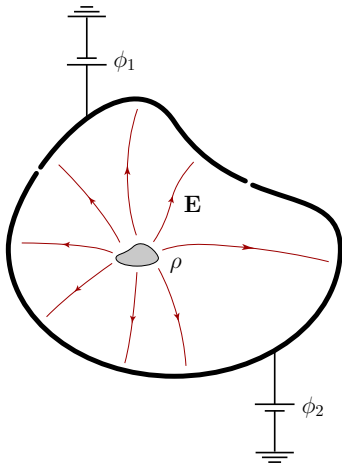
From the discussion above, one may get the impression that the spherical multipole expansion is more complicated than the cartesian one. Still it is the from most widely used in practice. The reason is that, from an applied perspective, one frequently wants to describe deviations of charge distributions from perfectly isotropic forms. Problems of this type are most naturally articulated, and answered, in the language of spherical coordinates and harmonics.

3.2.5 Boundary value problem II: finite space

Boundary conditions

Eq.(3.12) represents a solution to the Poisson equation with vanishing-potential boundary conditions at spatial infinity. In the text above we spoke of *the* solution, yet we haven't actually shown its uniqueness. A related and more general question: suppose we wish to compute the solution of the Poisson equation in some bounded region of space B . Further suppose we wish to specify the functional profile of the potential at the boundaries ∂B of that volume. How would one do that? I.e. how much information do we *need* to specify to make the solution in B unique and when would the problem become over-specified. Further, given a complete specification of the boundary conditions, how does one solve the Poisson equation in practice? These are the questions we will address in this section.

▷ Info. The boundary value problem sketched above is not quite as academic as it might seem. For example, in electrical engineering, one often needs to compute the electric field created by charge distributions inside space domains limited by metallic boundaries (see the figure.)



Now, the potential inside an ideally conducting metallic surface must be constant. Reason: was there a potential gradient tangential to the surface, the corresponding electric field would make the mobile charges inside the metal re-arrange so as to eliminate that very potential drop. (Notice that the constancy of the potential implies that electric field lines are perpendicular to metallic surfaces.) However, by attaching electrodes to the boundary sheets, one can put these at any desired value of the potential (relative to the ground potential $\phi = 0$.) Further, the region of interest may be bounded by several, mutually isolated boundary regions, i.e. the boundary potential may be described by some piecewise constant function $\phi|_{\partial B}$. By taking the — gedanken — limit of a very large number of (mutually isolated) boundary sheets, we arrive at the

prototypical boundary value problem specified by an arbitrary function $\phi|_{\partial B}$. Boundary conditions set in this way are denoted as **Dirichlet boundary conditions**. We will see below that the Dirichlet problem affords a unique solution (means extension of ϕ inside the domain.) That this must be so follows from our expectation that the setup outlined above is physically meaningful, i.e. will generate a unique electric field.

However, there are alternative ways of formulating a physically meaningful boundary problem: imagine a domain of space bounded by a layer of surface charges. In section 3.2.2 the field component normal to a surface charge layer, $\partial_n\phi$ is determined by the (local) amount of surface charge. This suggests that the specification of $\partial_n\phi$ everywhere at ∂B — so-called **Neumann boundary conditions** — also defines a uniquely solvable problem. However, the simultaneous specification of ϕ and its normal derivative (or any other additional information for that matter) over-specifies the problem.

We begin our analysis by proving the uniqueness of the solution to the Poisson equation with either Dirichlet or Neumann boundary conditions: Let the closed surface ∂B be the boundary of some region of space, B . (Parts, or even the whole of ∂B may be at spatial infinity.) On ∂B we specify either

- ▷ the potential $\phi|_{\partial B}$ (Dirichlet boundary conditions), or
- ▷ its normal derivative $\partial_n\phi|_{\partial B}$ (Neumann conditions.)

(Yet more generally, we may specify ϕ on a subset $S \subset \partial B$ and $\partial_n\phi$ on the complementary domain $\partial B \setminus S$.) Then the solution to the Poisson equation is uniquely determined.

To prove this statement, let us assume that two solutions ϕ_1 and ϕ_2 obeying the same boundary conditions had been found. Applying Green's first identity (2.21) to the difference $\Gamma \equiv \phi_1 - \phi_2$, we obtain

$$\int_B d^3x (\Gamma \Delta \Gamma - \nabla \Gamma \cdot \nabla \Gamma) = \int_{\partial B} d\sigma \Gamma \partial_n \Gamma.$$

Using that $\Delta \Gamma = -\Delta \phi_1 + \Delta \phi_2 = 4\pi(\rho - \rho) = 0$ and that for all points in ∂B , either $\partial_n \Gamma = 0$ or $\Gamma = 0$, we conclude $\int d^3x' |\nabla \Gamma|^2 = 0$. Since $|\nabla \Gamma| \geq 0$, this relation implies $\nabla \Gamma = 0$ everywhere in B . Thus $\Gamma = \text{const.}$, i.e. the two potentials are equal up to an inessential constant (which for Dirichlet boundary conditions must be zero at the boundary, and hence everywhere.)

Formal solution of the boundary value problem

We next employ the method of Green functions to construct a formal solution of the boundary value problem. The attribute 'formal' signals that the construction will leave us with much of the hard work of obtaining a solution of the Poisson for concretely specified boundaries. What the method will yield, though, is a reduction of the problem to the solution of two simpler ones: (i) the solution of the Poisson equation for a δ -source and simplified boundary data, plus (ii) an integral.

Let the Green function $G(\mathbf{x}, \mathbf{x}')$ be a solution of the δ -source Poisson equation in \mathbf{x}' : $\Delta' G(\mathbf{x}, \mathbf{x}') = -\delta(\mathbf{x} - \mathbf{x}')$ (i.e. deviating from our earlier definition, we let the Laplace operator act on the coordinate \mathbf{x}' . However, this is just a matter of convention; G is still a Green function.) Now, a particular solution to this equation — the one with vanishing boundary conditions at infinity — is $G_0(\mathbf{x}, \mathbf{x}') = (4\pi|\mathbf{x} - \mathbf{x}'|)^{-1}$. To implement other types of boundary

conditions, we may use the freedom to add to any particular solution an arbitrary solution of the homogeneous equation (the Laplace equation), $F(\mathbf{x}, \mathbf{x}')$, where $\Delta'F(\mathbf{x}, \mathbf{x}') = 0$. Thus, the most general form of the Green function is given by

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{4\pi} \frac{1}{|\mathbf{x} - \mathbf{x}'|} + F(\mathbf{x}, \mathbf{x}'). \quad (3.47)$$

Temporarily leaving the auxiliary function F unspecified, we next employ Greens' second identity (2.22) to relate the Green function to the potential ϕ . Choosing $\theta(\mathbf{x}') = G(\mathbf{x}, \mathbf{x}')$ and $\psi(\mathbf{x}') = \phi(\mathbf{x}')$, we then obtain the equation

$$\phi(\mathbf{x}) = 4\pi \int_B d^3x' G(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}') + \int_{\partial B} d\sigma' [G(\mathbf{x}, \mathbf{x}')\partial_{n'}\phi(\mathbf{x}') - \phi(\mathbf{x}')\partial_{n'}G(\mathbf{x}, \mathbf{x}')]. \quad (3.48)$$

While for boundaries ∂B sent to infinity we are back to formula (3.12), for finite volumes, the surface terms on the r.h.s. implement the relevant boundary information. Specifically, for **Dirichlet boundary conditions**, we will try to determine a function F such that $G(\mathbf{x}, \mathbf{x}')|_{\mathbf{x}' \in \partial B} = 0$ obeys vanishing (Dirichlet) boundary conditions. Then,

$$\phi(\mathbf{x}) = 4\pi \int_B d^3x' G(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}') - \int_{\partial B} d\sigma' \phi(\mathbf{x}')\partial_{n'}G(\mathbf{x}, \mathbf{x}') \quad (3.49)$$

describes the potential in terms of the Green function and the pre-specified Dirichlet boundary data $\phi(\mathbf{x}' \in \partial B)$. Notice the logics of our strategy: by utilizing the Green function, we reduce the problem (general sources/general Dirichlet boundary data) to the much simpler one (δ -source/vanishing Dirichlet boundary data). Once this latter problem has been solved, the solution of a general problem of the first kind can be obtained by integration.

Turning to the case of **Neumann boundary conditions** and following the same strategy as above, one might be tempted to set $\partial_{n'}G(\mathbf{x}, \mathbf{x}')|_{\mathbf{x}' \in \partial B} = 0$. However, this requirement would be too strong: application of Stokes theorem obtains $\int_{\partial B} d\sigma' \partial_{n'}G(\mathbf{x}, \mathbf{x}') = \int_B d^3x' \Delta'G(\mathbf{x}, \mathbf{x}') = -1$, i.e. the normal derivative must not be globally vanishing. However, we may well set it to the constant value $\partial_{n'}G(\mathbf{x}, \mathbf{x}')|_{\mathbf{x}' \in \partial B} = -A^{-1}$, where A is the area of ∂B .

Substituting this ansatz into (3.48), we obtain

$$\phi(\mathbf{x}) = 4\pi \int_B d^3x' G(\mathbf{x}, \mathbf{x}')\rho(\mathbf{x}') + \int_{\partial B} d\sigma' \partial_{n'}\phi(\mathbf{x}')G(\mathbf{x}, \mathbf{x}') + \langle \phi \rangle_{\partial B}, \quad (3.50)$$

where the constant $\langle \phi \rangle_{\partial B} \equiv A^{-1} \int_{\partial B} d\sigma' \phi(\mathbf{x}')$ is the average value of the potential on the boundary. Since this constant will not affect the electric field, its presence is not of importance. (However, the need to solve (3.50) self consistently may create some headaches when it comes to actual computations.¹⁶)

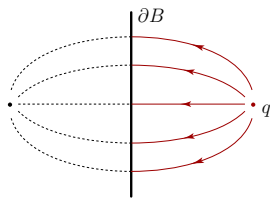
Before testing the method above on a concrete example, let us briefly discuss the **physical meaning of the function F in (3.47)**. Being a solution of the Laplace equation (inside B),

¹⁶Fortunately, in most Neumann-like problems, parts of the bounding surface are sent to infinity. In this case, $\langle \phi \rangle_{\partial B} = 0$ and the problem does not arise.

we may think of the function F as the potential created by charges *outside* B . For any specific choice of the reference point \mathbf{x} of the unit charge determining G , these outer charges modify the ‘bare’ potential $\sim |\mathbf{x} - \mathbf{x}'|^{-1}$ so as to conform with the boundary conditions. This way of thinking of the function F is not only of conceptual value; It provides one with a hint as to how to determine the function G in concrete applications. Especially in problems of a high degree of symmetry may try to introduce some ‘image charges’ outside B whose position/strength is chosen so as to adjust the boundary behaviour of ϕ appropriately.

Concrete solution of boundary value problems

First the bad news: there are no generally applicable solution strategies to obtain the Dirichlet or Neumann Green function; in most problems of practical interest one will need to resort to numerical procedures. However, in cases where one is dealing with a problem of exceptionally high symmetry, one may try to follow the logics outlined in the end of the previous section to manually construct a solution: Placing a system of fictitious charges — so called **image charges** — outside the reference volume B one may try to modify the potential inside B so as to conform with the boundary conditions. Let us introduce this method on two simple examples:



▷ Example: Consider a **charge distribution in front of an infinite conducting plane**. Choosing coordinates such that the plane coincides with the $x = 0$ plane and assuming that at spatial infinity it bends through the half space $x > 0$ into a closed surface ∂B we may apply the general formalism above. Consider, thus, a unit charge at some point \mathbf{x} in front of the surface. We wish to compute the potential created by this charge (aka the Green function) where $\phi|_{\partial B} = 0$. To this end, we use the ansatz (3.47), where we think of the function F as the potential created by a mirror or image charge kept at the other side of the plane. Symmetry considerations suggest¹⁷ to choose the image charge to (a) be of equal and opposite strength -1 , and (b) to place it at a coordinate vector $\mathbf{x}^s \equiv (-x, y, z)$ (i.e. the reflection of \mathbf{x} at the grounded plane.) With this choice, the Green function assumes the form

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{4\pi} \left(\frac{1}{|\mathbf{x}' - \mathbf{x}|} - \frac{1}{|\mathbf{x}' + \mathbf{x}^s|} \right). \quad (3.51)$$

This function indeed solves the Poisson in the positive half plane (the interior of ∂B .) On the plane $x = 0$ it vanishes, i.e. it obeys vanishing Dirichlet boundary conditions as required. According to the general construction recipe, the potential of a general charge distribution in the positive half plane is given by

$$\phi(\mathbf{x}) \stackrel{x>0}{=} 4\pi \int_{x>0} d^3x' G(\mathbf{x}, \mathbf{x}') \rho(\mathbf{x}').$$

It is important to keep in mind that this result holds only in the positive half plane (the interior of B); the image charge is purely fictitious, i.e. there are no true charges in the negative half plane

¹⁷As mentioned above, the method of image charges is applicable to problems with an exceptionally high degree of symmetry. Rather than ‘computing’ a system of suitable charges it is more customary to guess one.

implying that $\phi(\mathbf{x})|_{x<0} = 0$.¹⁸

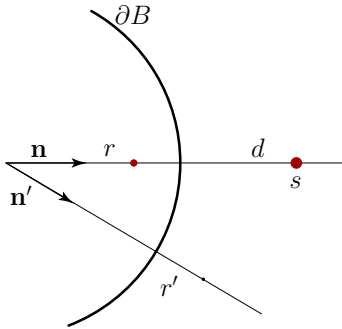
One may ask whether the second term (3.51) can be given a more 'physical' interpretation than that of the potential created by a phantom charge. Indeed, one should expect that a charge placed in front of a grounded conducting plane will lead to the generation of some surface charge (i.e. physically realized by electrons that are either attracted or repelled by the reference charge.) To compute the induced charge, we may employ Eq.(3.22), i.e. determine the jump in the normal component of the electric field across the surface. For the sake of simplicity, let us consider a point charge $q > 0$ sitting at $\mathbf{x} = (a, 0, 0)$ in front of the plane. According to Eq. (3.51), the potential created by this charge is given by

$$\phi(\mathbf{x}) = \begin{cases} q [((x-a)^2 + y^2 + z^2)^{-1/2} - ((x+a)^2 + y^2 + z^2)^{-1/2}], & x > 0, \\ 0, & x < 0. \end{cases}$$

Differentiating this result we obtain the surface charge density $\eta(y, z) = \Delta E_{\perp}/4\pi = -\partial_x \phi/4\pi = -(qa/2\pi)(a^2 + y^2 + z^2)^{-3/2}$. Notice that the (negative) induced charge is maximal at the point closest to the test charge and decays as $\sim r^{-3}$ for points in the plane far away from the point of closest proximity. The total amount of induced charge is given by

$$q_{\text{ind}} = \int_{\partial B} dydz \eta(y, z) = \frac{qa}{2\pi} \int_{\partial B} \frac{dydz}{(a^2 + y^2 + z^2)^{3/2}} = -q.$$

The totality of the induced charge completely 'screens' the test charge. Contrary to the fictitious image charge, the induced charge *is* physical (can be measured.). Also notice that the induced charge uniquely determines $\partial_n \phi$, i.e. it is surely impossible to freely impose both ϕ and $\partial_n \phi$.



▷ Example: As a second example, consider a charge distribution inside a conducting sphere of radius R (kept at potential $\phi = 0$.) Again, we attempt to construct the Green function of the problem by placing image charges into the region outside the system boundary. Let us *assume* that a single unit charge will be sufficient to equilibrate the potential on the sphere. By symmetry it is clear that this charge must be sit somewhere on the axis connecting the reference unit charge and the origin of the sphere.

Denoting the unit vector in the direction of the unit charge by \mathbf{n} and strength and distance of the image charge by s and d , respectively, the trial Green function is given by

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{4\pi} \left(\frac{1}{|r\mathbf{n} - r'\mathbf{n}'|} - \frac{s}{|d\mathbf{n} - r'\mathbf{n}'|} \right),$$

where \mathbf{n} is the unit vector in direction \mathbf{x} . Setting $r' = R$ and reorganizing factors, this becomes

$$G(\mathbf{x}, \mathbf{x}')|_{r'=R} = \frac{1}{4\pi} \left(\frac{1}{R|\mathbf{n}' - (r/R)\mathbf{n}|} - \frac{s}{d|\mathbf{n} - (R/d)\mathbf{n}'|} \right) \stackrel{!}{=} 0.$$

It is straightforward to verify that this expression vanishes if we set $d = R(R/r) > R$ and $s = R/r > 1$. The potential created by an arbitrary charge distribution inside the sphere is then given by (3.49).

¹⁸While the absence of charges merely implies $\phi(\mathbf{x})|_{x<0} = \text{const.}$, the vanishing of constant follows from the presence of a grounded ($\phi = 0$) boundary.

The analysis of the induced surface charge is left as an exercise.

3.3 Magnetostatics

3.3.1 Vector potential and solution of Maxwell equations

Before turning to the magnetic sector (3.2) of the static theory, let us recapitulate the solution strategy in the electric theory. Using the fact that $\nabla \times \mathbf{E} = 0$ (three scalar equations ‘constraining’ the form of the vector field \mathbf{E}), we wrote $\mathbf{E} = -\nabla\phi$, i.e. we obtained \mathbf{E} as the derivative of an unconstrained scalar field. Similarly, in the magnetic case, $\nabla \cdot \mathbf{B} = 0$ constrains the form of the field \mathbf{B} . In analogy to the ansatz above we might, thus, attempt to represent \mathbf{B} as $\mathbf{B} = D(\text{an unconstrained field})$, where D is a suitably chosen differential operator. Since, however, $\nabla \cdot \mathbf{B} = 0$ represents only one scalar constraint equation, the unconstrained auxiliary field will be of a more complex structure than the scalar potential ϕ . Indeed, we are going to show below that the equation $\nabla \cdot \mathbf{B} = 0$ implies that the magnetic field can be written as $\mathbf{B} = \nabla \times \mathbf{A}$, i.e. as the curl of an auxiliary field known as the (magnetic) **vector potential**. In magnetostatics, the vector potential plays a role analogous to the scalar potential of electrostatics.

▷ Math. To understand the mathematical principles behind the **existence of a vector potential**, consider the following problem (whose solution is of considerable importance in its own right): Let \mathbf{f} be a source-free vector field, $\nabla \cdot \mathbf{f} = 0$, and g be a scalar field. Is there a vector field \mathbf{X} such that $g = \nabla \cdot \mathbf{X}$ are the sources of \mathbf{X} and $\mathbf{f} = \nabla \times \mathbf{X}$ its circulation? If so, is \mathbf{X} unique?

To answer these questions, let us make an ansatz $\mathbf{X} = \mathbf{X}_s + \mathbf{X}_c$, where

$$\begin{aligned} \nabla \times \mathbf{X}_s &= 0, & \nabla \cdot \mathbf{X}_s &= g, \\ \nabla \cdot \mathbf{X}_c &= 0, & \nabla \times \mathbf{X}_c &= \mathbf{f}, \end{aligned}$$

i.e. we decompose the sought-for vector field into a ‘source component’ \mathbf{X}_s (curl-free but source-ful) and a complementary ‘curl component’ \mathbf{X}_c (curl but no sources.)

We know that the circulation-less component $\mathbf{X}_s = -\nabla\phi$ can be written as the gradient of a scalar function.¹⁹ This function must then be determined such that it satisfies the source condition, $\nabla \cdot \mathbf{X}_s = -\Delta\phi = g$. Turning to the complementary component \mathbf{X}_c , let us assume that the vanishing of the divergence can be resolved by making the ansatz, $\mathbf{X}_c = \nabla \times \mathbf{A}'$, i.e. we assume that the vanishing of the divergence implies the existence of a (vector) potential, very much like the vanishing of the curl implied the existence of a scalar potential. Surely this ansatz is consistent with $\nabla \cdot \mathbf{X}_c = 0$. To relate \mathbf{A}' to the circulation of \mathbf{X}_c , we compute $\nabla \times \mathbf{X}_c = \nabla \times \nabla \times \mathbf{A}' = -\Delta\mathbf{A}' + \nabla(\nabla \cdot \mathbf{A}') \stackrel{!}{=} \mathbf{f}$. The problem posed by the solution of this equation can be simplified by noting that the addition of the gradient of an arbitrary scalar function ψ to the vector potential, $\mathbf{A}' \rightarrow \mathbf{A} \equiv \mathbf{A}' + \nabla\psi$, does not change \mathbf{X}_c , i.e. the transformed vector potential is as good a vector potential as the original one. We now chose the function ψ such that $\nabla \cdot \mathbf{A} = \nabla \cdot \mathbf{A}' + \Delta\psi = 0$, i.e. we use it to render

¹⁹ . . . , where the minus sign has been introduced to stay close to the conventions of electrostatics.

\mathbf{A} divergence-less. Our source equation then reduces to $-\Delta\mathbf{A} = \mathbf{f}$ or three independent Poisson equations $\Delta A_i = -f_i$ for the components of \mathbf{A} . These equations are solved (assuming infinite space boundary conditions for simplicity) by the formula $A_i(\mathbf{x}) = \frac{1}{4\pi} \int d^3x' |\mathbf{x} - \mathbf{x}'|^{-1} f_i(\mathbf{x}')$.²⁰

Expressed in terms of the scalar and the vector potential, $\mathbf{X} = -\nabla\phi + \nabla \times \mathbf{A}$. However, this choice of \mathbf{X} is not unique: Without changing the circulation, we may add to \mathbf{X} the gradient of a scalar function, $\mathbf{X} \rightarrow \mathbf{X} + \nabla h$. If h solves the Laplace equation, $\Delta h = 0$, the source content, too, remains unchanged. I.e. the vector field \mathbf{X} is unique up to solutions of the Laplace equation.

The mathematical considerations outlined above solve much of the problem posed by the solution of Eqs.(3.2). The fact that the magnetic induction \mathbf{B} is source free implies that it can be written as the curl of a vector potential, $\mathbf{B} = \nabla \times \mathbf{A}$.²¹ The choice of the vector potential is not unique: without altering \mathbf{B} , we may add to \mathbf{A} the gradient of an arbitrary function,

$$\mathbf{A} \rightarrow \mathbf{A} + \nabla f. \quad (3.52)$$

▷ Info. The transformation Eq.(3.52) is the most elementary example of a **gauge transformation**. Later on we will identify it as a specific case of the most general gauge transformation of electrodynamics. In fact, gauge transformations play a pivotal role in theories far more general than electrodynamics; they form the conceptual basis of various types of gauge theories. The list of important gauge theories includes quantum electrodynamics (QED), quantum chromodynamics (QCD), and many more.

The prototypical transformation (3.52) anticipates two crucial features of every gauge theory: (i) the freedom to choose amongst different realizations of an auxiliary field (Generally, fields possessing a gauge freedom are called **gauge fields**.) can be used to obtain a particularly simple, or problem adjusted representation of the theory. (ii) As the gauge transformation affects only the gauge field (\mathbf{A}) but not the derived 'physical field' (\mathbf{B}) all final results obtained by the theory must be independent of the choice of the gauge. This condition of **gauge invariance** imposes a strong consistency check on the validity of the theory and of physical approximation schemes.

Throughout it will be convenient to choose a gauge wherein $\nabla \cdot \mathbf{A} = 0$, the so-called **Coulomb gauge**. (Recall the logics of the discussion above: for an arbitrary configuration \mathbf{A}' , we may choose a gauge function f such that $\Delta f = -\nabla \cdot \mathbf{A}'$ solves the Poisson equation. Then, the gauged vector potential $\mathbf{A} \equiv \mathbf{A}' + \nabla f$ is source-free. Without loss of generality we may

²⁰Note that this solution is indeed divergenceless, as required:

$$\begin{aligned} \nabla \cdot \mathbf{A} &= \partial_i A_i = \frac{1}{4\pi} \int d^3x' \partial_{x_i} |\mathbf{x} - \mathbf{x}'|^{-1} f_i(\mathbf{x}') = \frac{1}{4\pi} \int d^3x' (-\partial_{x_i} |\mathbf{x} - \mathbf{x}'|^{-1}) f_i(\mathbf{x}') = \\ &= \frac{1}{4\pi} \int d^3x' |\mathbf{x} - \mathbf{x}'|^{-1} \partial_{x_i'} f_i(\mathbf{x}') = 0, \end{aligned}$$

where $\partial_{x_i'} f_i(\mathbf{x}') = \nabla \cdot \mathbf{f}(\mathbf{x}') = 0$ by assumption.

²¹Alternatively, one may consider $\mathbf{B} = \nabla \times \mathbf{A}$ as an *ansatz* for the solution, \mathbf{B} , of the Maxwell differential equations. The construction below demonstrates that this ansatz works.

thus assume that \mathbf{A} has been source-free from the outset.) Using the vacuum identification $\mathbf{H} = \mathbf{B}$ and substituting the ansatz $\mathbf{B} = \nabla \times \mathbf{A}$ into the Maxwell equation $\nabla \times \mathbf{H} = 4\pi c^{-1} \mathbf{j}$ we then obtain $\nabla \times (\nabla \times \mathbf{A}) = -\Delta \mathbf{A} + \nabla(\nabla \cdot \mathbf{A}) = -\Delta \mathbf{A} = 4\pi c^{-1} \mathbf{j}$, i.e. the vector potential is obtained by solving the three independent Poisson equations $\Delta A_i = -4\pi c^{-1} j_i$. These equations are mathematically equivalent to their electrostatic analog $\Delta \phi = -4\pi \rho$ discussed above. Making the proper identifications and assuming infinite space boundary conditions we thus obtain the solution

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3 x' \frac{\mathbf{j}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (3.53)$$

Acting on this equation with the differential operator $\nabla \times$ and using the auxiliary identity $\nabla \times (f(\mathbf{x})\mathbf{v}(\mathbf{x}')) = \nabla f(\mathbf{x}) \times \mathbf{v}(\mathbf{x}')$ we find that the magnetic field created by the current distribution is given by

$$\mathbf{B}(\mathbf{x}) = \frac{1}{c} \int d^3 x' \frac{\mathbf{j}(\mathbf{x}') \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}. \quad (3.54)$$

After the experimentalists who first quantitatively explored the connection between magnetic fields and currents, Eq.(3.54) is called the **law of Biot and Savart**.

Two more remarks on the result (3.53) are in place: Firstly, notice that the Eq.(3.53) was derived for vanishing boundary conditions at infinity, $A(\mathbf{x})|_{|\mathbf{x}| \rightarrow \infty} \rightarrow 0$. In principle, however, our complete discussion of Poisson equations subject to more complex boundary conditions in section 3.2.5 may be carried over to the magnetic case. Yet, unlike with electrostatics (where both Dirichlet and Neumann boundary conditions could be given a clear physical meaning) we do not yet know how to interpret 'magnetic boundary conditions' on the behaviour of \mathbf{A} . For a discussion of this point we refer to chapter 5 below. Secondly, we need to verify that the vector potential (3.53) indeed obeys our working assumption, i.e. the Coulomb gauge condition $\nabla \cdot \mathbf{A} = 0$. To this end, we note that the Fourier transform of the Poisson equation $\Delta \mathbf{A} = -\frac{4\pi}{c} \mathbf{j}$ is given by $-k^2 \mathbf{A}(\mathbf{k}) = -\frac{4\pi}{c} \mathbf{j}(\mathbf{k})$. Solving for $\mathbf{A}(\mathbf{k})$ we obtain $\mathbf{A}(\mathbf{k}) = \frac{4\pi}{c} k^{-2} \mathbf{j}(\mathbf{k})$.²² Taking the scalar product of this equation with the vector \mathbf{k} and noting that $\nabla \cdot \mathbf{j}$ Fourier transforms to $\mathbf{k} \cdot \mathbf{j} = 0$, we obtain $\mathbf{k} \cdot \mathbf{A} = 0$ or $\nabla \cdot \mathbf{A} = 0$.

3.3.2 Magnetic multipole expansion

To better understand the physical meaning of Eqs.(3.53) and (3.54) above, let us consider a situation where the current flow in the system is confined to a small region B in space. We want to compute the magnetic field created by the currents at points far away from B . As in our discussion in section 3.2.4, we might apply the addition theorem to expand the distance function in (3.53) in spherical harmonics, thus generating a magnetic variant of the multipole expansion.

▷ Info. Indeed, expansions of this type are of considerable applied relevance: for example, the

²²To establish the equivalence between the Fourier representation and the real space representation above, recall the remarks made in the paragraph below Eq. (3.17).

construction of magnetic traps is based on **quadrupole magnets**, i.e. magnets whose multipole expansion starts at sub-leading order. Similarly, the particle beams in accelerators are commonly focused by sequences of quadrupole magnets (occasionally even octupole magnets or magnets of yet higher order.)

However, presently, we shall restrict ourselves to a simpler ‘brute force’ Taylor expansion of the distance-denominator, similar to the one outlined in section 3.2.2. To present the results of this expansion in a convenient manner, we will make use of the auxiliary identity

$$\int d^3x (f \mathbf{j} \cdot \nabla g + g \mathbf{j} \cdot \nabla f) = 0, \quad (3.55)$$

which holds for arbitrary functions f, g and source-free vector fields \mathbf{j} . This equation is proven by a straightforward integration by parts. Application of (3.55) to the cases $g = x_i$, $f = 1$ and $f = x_i$, $g = x_j$ yields the two relations

$$\int d^3x j_i = 0, \quad \int d^3x' (x_i j_j + j_j x_i) = 0. \quad (3.56)$$

Substituting the expansion (3.24) into (3.53), we obtain the series $\mathbf{A} = \mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \dots$, which should be compared to the scalar potential series (??). Specifically, the analog of the electrostatic monopole term is given by

$$\mathbf{A}^{(0)} = \frac{1}{cr} \int d^3x' \mathbf{j}(\mathbf{x}') = 0,$$

where we used the first of the two relations (3.56). The vanishing of the first moment reflects the **absence of magnetic monopoles**.

Turning to the second moment, we obtain

$$\begin{aligned} A_i^{(1)} &= \frac{x_j}{cr^3} \int d^3x' x'_j j_i \stackrel{(3.56)}{=} \frac{x_j}{2cr^3} \int d^3x' (x'_j j_i - x'_i j_j) = \frac{x_j \epsilon_{kji} \epsilon_{kj'i'}}{2cr^3} \int d^3x' x'_j j_{i'} = \\ &= \frac{x_j \epsilon_{kji}}{2cr^3} \int d^3x' (\mathbf{x}' \times \mathbf{j})_k = -\frac{1}{2cr^3} \left[\mathbf{x} \times \int d^3x' \mathbf{x}' \times \mathbf{j} \right]_i. \end{aligned}$$

In close analogy to the electric dipole moment, we define the **magnetic dipole moment**

$$\mathbf{m} \equiv \frac{1}{2c} \int d^3x \mathbf{x} \times \mathbf{j}. \quad (3.57)$$

Expressed in terms of this quantity, the dipolar contribution to the vector potential is given by

$$\boxed{\mathbf{A}^{(1)} = \frac{\mathbf{m} \times \mathbf{x}}{r^3}.} \quad (3.58)$$

In principle, the series expansion of the vector potential may now be driven to higher (quadrupole) orders. However, for the sake of simplicity, we will restrict our discussion to the discussion of the dominant contribution to the series, Eq.(3.58).

Taking the curl of (3.58), and noting that $\nabla \times (\mathbf{x} \times \mathbf{v}) = -(\nabla \cdot \mathbf{x})\mathbf{v} = -3\mathbf{v}$ we obtain the magnetic field created by the current dipole,

$$\mathbf{B}(\mathbf{x}) = \frac{3\mathbf{n}(\mathbf{n} \cdot \mathbf{m}) - \mathbf{m}}{r^3}, \quad (3.59)$$

where \mathbf{n} is the unit vector in \mathbf{x} -direction. Notice the structural equivalence of this expression to its electric analog, Eq.(3.16).

Before proceeding, let us briefly discuss the magnetic dipole moments of two particular types current of distributions of considerable practical relevance:

- ▷ Often, the current sources are realized as (approximately) **closed current loops**, each carrying a constant current I . In this case (cf. Eq.(2.17)) the volume integral (3.57) reduces to

$$\mathbf{m} = -\frac{I}{2c} \oint_{\gamma} d\mathbf{s} \times \mathbf{x}. \quad (3.60)$$

Particularly simple expressions are obtained for planar current loops, i.e. in cases where the curve γ lies in a plane. Introducing coordinates such that the unit vector normal to the plane $\mathbf{n} = \mathbf{e}_3$ coincides with the coordinate vector \mathbf{e}_3 , we then obtain for the i th component of the magnetic moment

$$m_i = -\frac{I}{2c} \mathbf{e}_i \cdot \oint_{\gamma} d\mathbf{s} \times \mathbf{x} = -\frac{I}{2c} \oint_{\gamma} \mathbf{e}_i \cdot (d\mathbf{s} \times \mathbf{x}) = -\frac{I}{2c} \oint_{\gamma} d\mathbf{s} \cdot (\mathbf{x} \times \mathbf{e}_i).$$

Application of Stokes law then yields $m_i = -\frac{I}{2c} \int_{S(\gamma)} \int d\sigma \mathbf{e}_3 \cdot (\nabla \times (\mathbf{x} \times \mathbf{e}_i))$. Noting that $\mathbf{e}_3 \cdot (\nabla \times (\mathbf{x} \times \mathbf{e}_i)) = -2\delta_{3i}$, we obtain

$$\mathbf{m} = \frac{AI}{c} \mathbf{e}_3,$$

i.e. a magnetic moment perpendicular to the plane supporting the current loop and proportional to the area of the loop.

- ▷ Imagine a **system of point particles** at coordinates \mathbf{x}_i where each particle carries charge q_i , is of mass m_i , and moves with velocity \mathbf{v}_i . In this case, $\mathbf{j} = \sum_i \mathbf{v}_i q_i \delta(\mathbf{x} - \mathbf{x}_i)$. Equation (3.57) reduces to

$$\frac{1}{2c} \sum_i q_i \mathbf{x}_i \times \mathbf{v}_i = \frac{1}{2c} \sum_i \frac{q_i}{m_i} \mathbf{l}_i, \quad (3.61)$$

where $\mathbf{l} = \mathbf{x} \times (m\mathbf{v})$ is the angular momentum carried by a point particle. [Only parts of the magnetic moment carried by genuine elementary particles are due to their orbital angular momentum \mathbf{l} . A second contribution stems from their 'intrinsic' angular momentum or **spin**. Specifically, for an electron at rest, $\mathbf{m} = 2.002 \times \frac{e}{2cm} \mathbf{S}$, where $|\mathbf{S}| = 1/2$ is the electron spin and the pre-factor 2.002 is known as the **g-factor** of the electron.]

3.3.3 Magnetic forces

To understand the physical meaning of the connections (currents \rightsquigarrow fields) derived above, let us explore an intuitively accessible quantity, viz. the mechanical forces created by a current distribution. Assuming that the charge density of the mobile charge carriers in the wire is given by ρ , and that these charges move with a velocity \mathbf{v} , the current density in the wire is given by $\mathbf{j} = \rho\mathbf{v}$.²³ Comparison with the Lorentz force law (2.6) shows that in a magnetic field the wire will be subject to a force density $= c^{-1}\rho\mathbf{v} \times \mathbf{B}$. Specifically, for a thin wire of cross section dA , and carrying a current $I = \rho dAv$ (v is the component of $\mathbf{v} = v\mathbf{e}_{\parallel}$ along the wire.) the force acting on a segment of length dl will be $d\mathbf{f} = \mathbf{f}dAdl = c^{-1}\rho dAdl(v\mathbf{e}_{\parallel}) \times \mathbf{B} = c^{-1}I ds \times d\mathbf{B}$. Summarizing, the force acting on a line element ds of a wire carrying a current I is given by

$$\boxed{d\mathbf{f} = \frac{I}{c} ds \times \mathbf{B}.} \quad (3.62)$$

In the following, we discuss a few applications of the prototypical force–formula (3.62):

Forces on current loops

Consider a **single closed current loop** γ carrying a current I . The current flow will generate a magnetic field (3.54) which will in turn act on the line elements of the loop, as described by (3.62). One may thus wonder whether the loop exerts a net force on itself. Integrating over (3.62), we find that this force is given by

$$\begin{aligned} \mathbf{F} &= \oint_{\gamma} d\mathbf{f} = \frac{I}{c} \oint_{\gamma} ds \times \mathbf{B} \stackrel{(3.54)}{=} \frac{I^2}{c^2} \oint_{\gamma} \oint_{\gamma} \frac{ds \times (ds' \times (\mathbf{x} - \mathbf{x}'))}{|\mathbf{x} - \mathbf{x}'|^3} = \\ &= \frac{I^2}{c^2} \oint_{\gamma} \oint_{\gamma} \left(\frac{ds'(ds \cdot (\mathbf{x} - \mathbf{x}'))}{|\mathbf{x} - \mathbf{x}'|^3} - \frac{(\mathbf{x} - \mathbf{x}') ds \cdot ds'}{|\mathbf{x} - \mathbf{x}'|^3} \right) = \\ &= -\frac{I^2}{c^2} \oint_{\gamma} ds' \oint_{\gamma} ds \cdot \nabla \frac{1}{|\mathbf{x} - \mathbf{x}'|} = 0. \end{aligned}$$

The second term in the second line vanishes because it changes sign under coordinate interchange $\mathbf{x} \leftrightarrow \mathbf{x}'$. In the third line we used that $(\mathbf{x} - \mathbf{x}')|\mathbf{x} - \mathbf{x}'|^{-3} = \nabla|\mathbf{x} - \mathbf{x}'|^{-1}$ and that the integral of the gradient of a function over a closed loop vanishes. We conclude that a *closed current loop does not exert a net force onto itself*.

However, distinct current distributions do, in general, act by magnetic forces onto each other. By way of example, consider **two infinite parallel wires** kept at a distance d and carrying currents I_1 and I_2 , resp. Choosing coordinates as shown in the figure and parameterizing

²³The amount of charge flowing through a small surface element $d\sigma$ during a time interval dt is $dQ = d\sigma\rho\mathbf{n} \cdot \mathbf{v}dt$. Dividing by dt we find that the current through $d\sigma$ is given by $I_{d\sigma} = d\sigma\mathbf{n} \cdot (\rho\mathbf{v})$ which means that $\mathbf{j} = \rho\mathbf{v}$ is the current density.

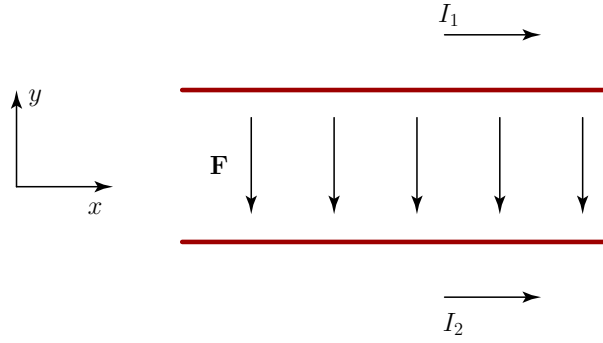


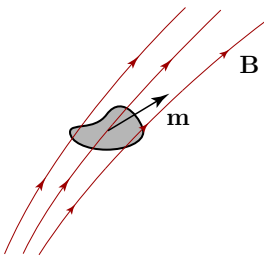
Figure 3.2: Two infinite wires exerting a force on each other

the line elements as $ds_{1,2} = dx_{1,2} \mathbf{e}_x$, we obtain the force \mathbf{F}_{12} wire #2 exerts on wire #1 as

$$\begin{aligned} \mathbf{F}_{12} &= \frac{I_1 I_2}{c^2} \int dx_1 \int dx_2 \frac{\mathbf{e}_x \times (\mathbf{e}_x \times (\mathbf{e}_x(x_1 - x_2) - \mathbf{e}_y d))}{|(x_1 - x_2)^2 + d^2|^{3/2}} = \\ &= \mathbf{e}_y \frac{I_1 I_2 d}{c^2} \int dx_1 \int dx_2 \frac{1}{(x_2^2 + d^2)^{3/2}} = \mathbf{e}_y \frac{2I_1 I_2}{c^2 d} \int dx_1, \end{aligned}$$

where in the second line we used $\int_{-\infty}^{\infty} dx (x^2 + d^2)^{-3/2} = 2/d^2$. Reflecting their infinite length, the wires exert an infinite force on each other. However, the force per unit length $\frac{d\mathbf{F}}{dx} = \frac{2I_1 I_2}{c^2 d} \mathbf{e}_y$ is finite and depends inverse quadratically on the distance between the wires. The force is attractive/repulsive for currents flowing in the same/opposite direction. (One of the more prominent manifestations of the attractive magnetic forces between parallel current flows is the phenomenon of current implosion in hollow conductors.)

Forces on local current distributions



Consider a spatially localized current distribution in an external magnetic field. We wish to compute **total force** acting on the distribution

$$\mathbf{F} = \int d^3x \mathbf{f} = \frac{1}{c} \int d^3x \mathbf{j} \times \mathbf{B}. \quad (3.63)$$

Choosing the origin of the coordinate system somewhere inside the current distribution, and assuming that the magnetic field varies slowly across the extent of the current flow, one may Taylor expand \mathbf{B} in the Lorentz force formula:

$$\mathbf{F} = \frac{1}{c} \int d^3x \mathbf{j}(\mathbf{x}) \times [\mathbf{B}(0) + \mathbf{x} \cdot \nabla \mathbf{B}(0) + \dots].$$

($\mathbf{x} \cdot \nabla \mathbf{B}$ is a vector with components $x_i \partial_i B_j$.) The auxiliary identity (3.56) implies that the

first contribution to the expansion vanish. The i th component of the force is thus given by

$$\begin{aligned}
 F_i &\simeq \frac{1}{c} \int d^3x \epsilon_{ijk} j_j x_l \partial_l B_k \stackrel{(3.56)}{=} \frac{1}{2c} \int d^3x \epsilon_{ijk} \epsilon_{mjl} (\mathbf{j} \times \mathbf{x})_m \partial_l B_k = \\
 &= \frac{1}{2c} \int d^3x (\delta_{kl} \delta_{im} - \delta_{km} \delta_{il}) (\mathbf{j} \times \mathbf{x})_m \partial_l B_k = \frac{1}{2c} \int d^3x ((\mathbf{j} \times \mathbf{x})_i \partial_k B_k - (\mathbf{j} \times \mathbf{x})_k \partial_i B_k) = \\
 &= \partial_i \Big|_{\mathbf{x}=0} (\mathbf{m} \cdot \mathbf{B}(\mathbf{x})),
 \end{aligned}$$

or

$$\boxed{\mathbf{F} = \nabla \Big|_{\mathbf{x}=0} (\mathbf{m} \cdot \mathbf{B}(\mathbf{x})),} \quad (3.64)$$

where in the second line we used $\nabla \cdot \mathbf{B} = \partial_l B_l = 0$. Thus, (a) the force is strongest if the magnetic moment is aligned with the magnetic field, and (b) proportional to the rate at which the external magnetic field varies. Specifically, for a uniform field, no forces act.

While the force on a magnetic moment relies on the presence of a field gradient, even a constant field does exert a finite **torque**

$$\mathbf{N} = \int d^3x \mathbf{x} \times \mathbf{f} = \frac{1}{c} \int d^3x \mathbf{x} \times (\mathbf{j} \times \mathbf{B}). \quad (3.65)$$

Approximating the field by its value at the origin,

$$\begin{aligned}
 N_i &= \frac{1}{c} \int d^3x \epsilon_{ijk} \epsilon_{klm} x_j j_l B(0)_m = \frac{1}{c} \int d^3x (\delta_{il} \delta_{jm} - \delta_{im} \delta_{lj}) x_j j_l B(0)_m = \\
 &= \frac{1}{c} \int d^3x (x_j j_i B(0)_j - x_j j_j B(0)_i) \stackrel{(3.56)}{=} \frac{1}{2c} \int d^3x \epsilon_{lji} (\mathbf{x} \times \mathbf{j})_l B(0)_j = \\
 &= (\mathbf{m} \times \mathbf{B}(0))_i.
 \end{aligned}$$

Thus, the torque acting on the moment,

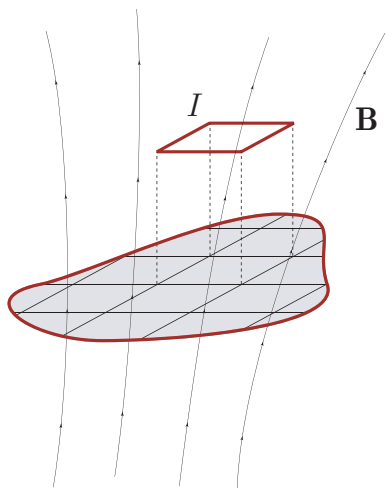
$$\boxed{\mathbf{N} = \mathbf{m} \times \mathbf{B}(0),} \quad (3.66)$$

is perpendicular to both the external field and the moment vector. It acts so as to align the field and the moment.

4.0 Electrodynamics

4.1 Magnetic field energy

As a prelude to our discussion of the full set of Maxwell equations, let us address a question which, in principle, should have been answered in the previous chapter: What is the energy stored in a static magnetic field? In section 3.2.2, the analogous question for the electric field was answered in a constructive manner: we computed the *mechanical* energy required to build up a system of charges. It turned out that the answer could be formulated entirely in terms of the electric field, without explicit reference to the charge distribution creating it. By symmetry, one might expect a similar prescription to work in the magnetic case. Here, one would ask for the energy needed to build up a current distribution against the magnetic field created by those elements of the current distribution that have already been put in place. A moments thought, however, shows that this strategy is not quite as straightforwardly implemented as in the electric case: no matter how slowly we move a ‘current loop’ in an magnetic field, an electric field acting on the charge carriers maintaining the current in the loop will be induced — the induction law. Work will have to be done to maintain a constant current and it is this work function that essentially enters the energy balance of the current distribution.



To make this picture quantitative, consider a current loop carrying a current I . We may think of this loop as being consecutively built up by importing small current loops (carrying current I) from infinity (see the figure.) The currents flowing along adjacent segments of these loops will eventually cancel so that only the net current I flowing around the boundary remains. Let us, then, compute the work that needs to be done to bring in one of these loops from infinity.

Consider, first, an ordinary point particle kept at a (mechanical) potential U . The rate at which this potential changes if the particle changes its position is $d_t U = d_t U(\mathbf{x}(t)) = \nabla U \cdot d_t \mathbf{x} = -\mathbf{F} \cdot \mathbf{v}$, where \mathbf{F} is the force acting on the particle. Specifically, for the charged particles moving in our prototypical current loops, $\mathbf{F} = q\mathbf{E}$, where \mathbf{E} is the electric field induced by variation of the magnetic flux through the loop as it approaches from infinity.¹ Now, what is \mathbf{E} in a situa-

¹In the expression describing the potential change, the Lorentz force contribution drops out because $(\mathbf{v} \times$

tion where there are no free electric charges acting as sources? Our test particle moves in a magnetic field, and due to this motion the strength of the magnetic field experienced by it changes in time. This causes an electric field by the law of induction. To relate the field \mathbf{E} to the time variation of a 'magnetic quantity', consider the line integral around an arbitrary loop, $\int ds \cdot \mathbf{E} = -c^{-1} \int_{\delta S} d\sigma \mathbf{n} \cdot d_t \mathbf{B} = -c^{-1} \int_{\delta S} d\sigma \mathbf{n} \cdot d_t(\nabla \times \mathbf{A}) = -c^{-1} \int ds \cdot d_t \mathbf{A}$. Since this holds regardless of the geometry of the loop, we have $\mathbf{E} = -c^{-1} d_t \mathbf{A}$, where $d_t \mathbf{A}$ is the change in vector potential due to the movement of the loop. We thus express the change in potential energy as $d_t U = \frac{q}{c} d_t \mathbf{A} \cdot \mathbf{v}$. Now let us generalize this expression from the potential change of a single particle, to that, dU , of the particles moving in a volume element d^3x inside a current carrying loop. Since the charge carried by a volume element in the loop is given by ρd^3x , and $\rho \mathbf{v} = \mathbf{j}$, we get

$$d_t dU = \frac{d^3x}{c} d_t \mathbf{A} \cdot \mathbf{j}. \quad (4.1)$$

Finally, integration over space and time yields the total energy change associated with the motion of the loop (or more generally a current distribution) as

$$E = \frac{1}{c} \int d^3x \mathbf{j} \cdot \mathbf{A}. \quad (4.2)$$

Although derived for the specific case of a current loop, Eq.(4.2) holds for general current distributions subject to a magnetic field. (For example, for the current density carried by a point particle at $\mathbf{x}(t)$, $\mathbf{j} = q\delta(\mathbf{x} - \mathbf{x}(t))\mathbf{v}(t)$, we obtain $E = q\mathbf{v}(t) \cdot \mathbf{A}(\mathbf{x}(t))$, i.e. the familiar Lorentz-force contribution to the Lagrangian of a charged particle.)

Now, assume that we shift the loop at *fixed current* against the magnetic field. The change in potential energy corresponding to a small shift is given by $\delta E = c^{-1} \int d^3x \mathbf{j} \cdot \delta \mathbf{A}$, where $\nabla \times \delta \mathbf{A} = \delta \mathbf{B}$ denotes the change in the field strength. Using that $\nabla \times \mathbf{H} = 4\pi c^{-1} \mathbf{j}$, we represent δE as

$$\begin{aligned} \delta E &= \frac{1}{4\pi} \int d^3x (\nabla \times \mathbf{H}) \cdot \delta \mathbf{A} = \frac{1}{4\pi} \int d^3x \epsilon_{ijk} (\partial_j H_k) \delta A_i = \\ &= -\frac{1}{4\pi} \int d^3x H_k \epsilon_{ijk} \partial_j \delta A_i = \frac{1}{4\pi} \int d^3x \mathbf{H} \cdot \delta \mathbf{B}, \end{aligned}$$

where in the integration by parts we noted that due to the spatial decay of the fields no surface terms at infinity arise. Due to the linear relation $\mathbf{H} = \mu_0^{-1} \mathbf{B}$, we may write $\mathbf{H} \cdot \delta \mathbf{B} = \delta(\mathbf{H} \cdot \mathbf{B})/2$, i.e. $\delta E = \frac{1}{8\pi} \delta \int \mathbf{B} \cdot \mathbf{E}$. Finally, summing over all shifts required to bring the current loop in from infinity, we obtain

$$\boxed{E = \frac{1}{8\pi} \int d^3x \mathbf{H} \cdot \mathbf{B}} \quad (4.3)$$

for the **magnetic field energy**. Notice (a) that we have again managed to express the energy of the system entirely in terms of the fields, i.e. without explicit reference to the sources creating these fields and (b) the structural similarity to the electric field energy (3.20).

$\mathbf{B} \cdot \mathbf{v} = 0$: magnetic forces do not change the energy of particles.

4.2 Electromagnetic gauge field

Consider the full set of Maxwell equations in vacuum ($\mathbf{E} = \mathbf{D}$ and $\mathbf{B} = \mathbf{H}$),

$$\begin{aligned}\nabla \cdot \mathbf{E} &= 4\pi\rho, \\ \nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{E} &= \frac{4\pi}{c} \mathbf{j}, \\ \nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial}{\partial t} \mathbf{B} &= 0, \\ \nabla \cdot \mathbf{B} &= 0.\end{aligned}\tag{4.4}$$

As in previous sections we will try to use constraints inherent to these equations to compactify them to a smaller set of equations. As in section 3.3, the equation $\nabla \cdot \mathbf{B} = 0$ implies

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}.}\tag{4.5}$$

(However, we may no longer expect \mathbf{A} to be time independent.) Now, substitute this representation into the law of induction: $\nabla \times (\mathbf{E} + c^{-1} \partial_t \mathbf{A}) = 0$. This implies that $\mathbf{E} + c^{-1} \partial_t \mathbf{A} = -\nabla \phi$ can be written as the gradient of a scalar potential, or

$$\boxed{\mathbf{E} = -\nabla \phi - \frac{1}{c} \partial_t \mathbf{A}.}\tag{4.6}$$

We have, thus, managed to represent the electromagnetic fields as in terms of derivatives of a generalized four-component potential $A = \{A_\mu\} = (\phi, -\mathbf{A})$. (The negative sign multiplying \mathbf{A} has been introduced for later reference.) Substituting Eqs. (4.5) and (4.6) into the inhomogeneous Maxwell equations, we obtain

$$\begin{aligned}-\Delta \phi - \frac{1}{c} \partial_t \nabla \cdot \mathbf{A} &= 4\pi\rho, \\ -\Delta \mathbf{A} + \frac{1}{c^2} \partial_t^2 \mathbf{A} + \nabla(\nabla \cdot \mathbf{A}) + \frac{1}{c} \partial_t \nabla \phi &= \frac{4\pi}{c} \mathbf{j}\end{aligned}\tag{4.7}$$

These equations do not look particularly inviting. However, as in section 3.3 we may observe that the choice of the generalized vector potential \mathbf{A} is not unique; this freedom can be used to transform Eq.(4.7) into a more manageable form: For an arbitrary function $f : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}$, $(\mathbf{x}, t) \mapsto f(\mathbf{x}, t)$. The transformation $\mathbf{A} \rightarrow \mathbf{A} + \nabla f$ leaves the magnetic field unchanged while the electric field changes according to $\mathbf{E} \rightarrow \mathbf{E} - c^{-1} \partial_t \nabla f$. If, however, we synchronously redefine the scalar potential as $\phi \rightarrow \phi - c^{-1} \partial_t f$, the electric field, too, will not be affected by the transformation. Summarizing, the **generalized gauge transformation**

$$\begin{aligned}\mathbf{A} &\rightarrow \mathbf{A} + \nabla f, \\ \phi &\rightarrow \phi - \frac{1}{c} \partial_t f,\end{aligned}\tag{4.8}$$

leaves the electromagnetic fields (4.5) and (4.6) unchanged. (In the four component shorthand notation introduced above, the gauge transformation assumes easy-to-memorize the form $A_\mu \rightarrow A_\mu - \partial_\mu f$, where the four-derivative operator $\{\partial_\mu\} = (\partial_0, \nabla)$ and $x_0 = ct$ as before.)

The gauge freedom may be used to transform the vector potential into one of several convenient representations. Of particular relevance to the solution of the time dependent Maxwell equations is the so-called **Lorenz gauge**²

$$\boxed{\nabla \cdot \mathbf{A} + \frac{1}{c} \partial_t \phi = 0.} \quad (4.9)$$

This equation, too, affords a compact four-vector notation: for a general vector $\{v^\mu\} = (v_0, \mathbf{v})$ we define a vector $\{v_\mu\} \equiv (v_0, -\mathbf{v})$, i.e. an object with 'lowered components' that differs from the one with 'raised components' by a sign change in the space-like sector. Using this notation, the Lorenz condition assumes the form $\partial_\mu A^\mu = 0$. It is always possible to satisfy this condition by a suitable gauge transformation. Indeed, if A' does not obey the Lorenz condition, we may define $A_\mu = A'_\mu - \partial_\mu f$ to obtain $0 \stackrel{!}{=} \partial_\mu A^\mu = \partial_\mu A'^\mu - \partial_\mu \partial^\mu f$. If we chose f so as to satisfy the equation $\partial_\mu f^\mu = \partial_\mu A'^\mu$, the Lorenz equation is satisfied. Expressed in terms of space and time components, this latter equation assumes the form

$$\left(\Delta - \frac{1}{c^2} \partial_t^2 \right) f = - \left(\nabla \cdot \mathbf{A}' + \frac{1}{c} \partial_t \phi' \right),$$

i.e. a wave equation with inhomogeneity $-(\nabla \cdot \mathbf{A} + c^{-1} \partial_t \phi)$. We shall see momentarily that such equations can always be solved, i.e. an implementation of the Lorenz gauge condition is possible.

In the Lorenz gauge, the Maxwell equations assume the simplified form

$$\begin{aligned} \left(\Delta - \frac{1}{c^2} \partial_t^2 \right) \phi &= -4\pi\rho, \\ \left(\Delta - \frac{1}{c^2} \partial_t^2 \right) \mathbf{A} &= -\frac{4\pi}{c} \mathbf{j}. \end{aligned} \quad (4.10)$$

In combination with the gauge condition (4.9), Eqs. (4.10) are fully equivalent to the set of Maxwell equations (4.4). Again, the four-vector notation may be employed to compactify the notation still further. With $\partial_\mu \partial^\mu = -\Delta + c^{-2} \partial_t^2$ and $j_\mu = (c\rho, -\mathbf{j})$, the potential equations assume the form

$$\boxed{\partial_\nu \partial^\nu A_\mu = \frac{4\pi}{c} j_\mu, \quad \partial_\mu A^\mu = 0.} \quad (4.11)$$

Before turning to the discussion of the solution of these equations a few general remarks are in order:

- ▷ The Lorenz condition is the prevalent gauge choice in electrodynamics because (a) it brings the Maxwell equations into a maximally simple form and (b) will turn out below to be invariant under the most general class of coordinate transformations, the Lorentz

²Named after the Danish physicist *Ludwig Lorenz* (1829-91) whose name is often confused with that of Hendrik Lorentz (1853-1928), the inventor of the Lorentz transformation, and many other concepts in the field.

transformations see below. (It is worthwhile to note that the Lorenz condition does not unambiguously fix the gauge of the vector potential. Indeed, we may add to any Lorenz gauge vector potential A_μ a function $A_\mu \rightarrow A_\mu + f_\mu$ satisfying the homogeneous wave equation $\partial^\mu \partial_\mu f = 0$ without altering the condition $\partial_\mu A^\mu = 0$.) Other gauge conditions frequently employed include

- ▷ the **Coulomb gauge** or **radiation gauge** $\nabla \cdot \mathbf{A} = 0$ (employed earlier in section 3.3.) The advantage of this gauge is that the scalar Maxwell equation assumes the simple form of a Poisson equation $\Delta \phi(\mathbf{x}, t) = -4\pi\rho(\mathbf{x}, t)$ which is solved by $\phi(\mathbf{x}, t) = \int d^3x' \rho(\mathbf{x}', t)/|\mathbf{x} - \mathbf{x}'|$, i.e. by an instantaneous 'Coulomb potential' (hence the name Coulomb gauge.) This gauge representation has also proven advantageous within the context of quantum electrodynamics. However, we won't discuss it any further in this text.

4.3 Electromagnetic waves in vacuum

As a warmup to our discussion of the full problem posed by the solution of Eqs. (4.10), we consider the vacuum problem, i.e. a situation where no sources are present, $j_\mu = 0$. Taking the curl of the second of Eqs. (4.10) and using (4.5) we then find

$$\left(\Delta - \frac{1}{c^2} \partial_t^2\right) \mathbf{B} = 0.$$

Similarly, adding to the gradient of the first equation c^{-1} times the time derivative of the second, and using Eq.(4.6), we obtain

$$\left(\Delta - \frac{1}{c^2} \partial_t^2\right) \mathbf{E} = 0,$$

i.e. in vacuum both the electric field and the magnetic field obey homogeneous wave equations.

4.3.1 Solution of the homogeneous wave equations

The homogeneous wave equations are conveniently solved by Fourier transformation. To this end, we define a four-dimensional variant of Eq. (3.10),

$$\begin{aligned} \tilde{f}(\omega, \mathbf{k}) &= \int d^3x dt f(t, \mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x} + i\omega t}, \\ f(t, \mathbf{x}) &= \int (d^3k)(d\omega) f(\omega, \mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t}, \end{aligned} \quad (4.12)$$

where $(d^d s) \equiv d^d s / (2\pi)^d$. The one difference to Eq.(3.10) is that the sign-convention in the exponent of the (ω/t) -sector of the transform differs from that in the (\mathbf{k}, \mathbf{x}) -sector. We next

subject the homogeneous wave equation

$$\left(\Delta - \frac{1}{c^2} \partial_t^2\right) \psi(t, \mathbf{x}) = 0,$$

(where ψ is meant to represent any of the components of the \mathbf{E} - or \mathbf{B} -field) to this transformation and obtain

$$\left(k^2 - \left(\frac{\omega}{c}\right)^2\right) \psi(\omega, \mathbf{k}) = 0.$$

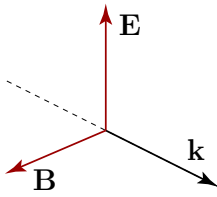
Evidently, the solution ψ must vanish for all values (ω, \mathbf{k}) , except for those for which the factor $k^2 - (\omega/c)^2 = 0$. We may thus write

$$\psi(\omega, \mathbf{k}) = c_+(\mathbf{k})\delta(\omega - kc) + c_-(\mathbf{k})\delta(\omega + kc),$$

where $c_{\pm} \in \mathbb{C}$ are arbitrary complex functions of the **wave vector** \mathbf{k} . Substituting this representation into the inverse transform, we obtain the general **solution of the scalar homogeneous wave equation**

$$\psi(t, \mathbf{x}) = \int d^3k \left(c_+(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x} - ckt)} + c_-(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x} + ckt)} \right). \quad (4.13)$$

The general solution is obtained by linear superposition of elementary **plane waves**, $e^{i(\mathbf{k}\cdot\mathbf{x} \mp ckt)}$, where each wave is weighted with an arbitrary coefficient $c_{\pm}(\mathbf{k})$. The elementary constituents are called waves because for any fixed instance of space, \mathbf{x} /time, t they harmonically depend on the complementary argument time, t /position vector, \mathbf{x} . The waves are planar in the sense that for all points in the plane fixed by the condition $\mathbf{x} \cdot \mathbf{k} = \text{const.}$ the phase of the wave is identical, i.e. the set of points $\mathbf{k} \cdot \mathbf{x} = \text{const.}$ defines a 'wave front' perpendicular to the wave vector \mathbf{k} . The spacing between consecutive wave fronts with the same phase $\arg(\exp(i(\mathbf{k} \cdot \mathbf{x} - ckt)))$ is given by $\Delta x = \frac{2\pi}{k} \equiv \lambda$, where λ is the **wave length** of the wave and $\lambda^{-1} = 2\pi/k$ its **wave number**. The temporal oscillation period of the wave fronts is set by $2\pi/ck$.



Focusing on a fixed wave vector \mathbf{k} , we next generalize our results to the vectorial problem posed by the homogeneous wave equations. Since every component of the fields \mathbf{E} and \mathbf{B} is subject to its own independent wave equation, we may write down the prototypical solutions

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}, \quad \mathbf{B}(\mathbf{x}, t) = \mathbf{B}_0 e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}, \quad (4.14)$$

where we introduced the abbreviation $\omega = ck$ and $\mathbf{E}_0, \mathbf{B}_0 \in \mathbb{C}^3$ are constant coefficient vectors. The Maxwell equations $\nabla \cdot \mathbf{B} = 0$ and (vacuum) $\nabla \cdot \mathbf{E} = 0$ imply the condition $\mathbf{k} \cdot \mathbf{E}_0 = \mathbf{k} \cdot \mathbf{B}_0 = 0$, i.e. the coefficient vectors are orthogonal to the wave vector. Waves of this type, oscillating in a direction perpendicular to the wave vector, are called **transverse waves**. Finally, evaluating Eqs.(4.14) on the law of induction $\nabla \times \mathbf{E} + c^{-1} \partial_t \mathbf{B} = 0$, we obtain

the additional equation $\mathbf{B}_0 = \mathbf{n}_k \times \mathbf{E}_0$, i.e. the vector \mathbf{B}_0 is perpendicular to both \mathbf{k} and \mathbf{E}_0 , and of equal magnitude as \mathbf{E}_0 . Summarizing, the vectors

$$\mathbf{k} \perp \mathbf{E} \perp \mathbf{B} \perp \mathbf{k}, \quad |\mathbf{B}| = |\mathbf{E}| \quad (4.15)$$

form an orthogonal system and \mathbf{B} is uniquely determined by \mathbf{E} (and vice versa).

At first sight, it may seem that we have been too liberal in formulating the solution (4.14): while the physical electromagnetic field is a *real* vector field, the solutions (4.14) are manifestly complex. The simple solution to this conflict is to identify $\text{Re } \mathbf{E}$ and $\text{Re } \mathbf{B}$ with the physical fields.³

4.3.2 Polarization

In the following we will discuss a number of physically different realizations of plane electromagnetic waves. Since \mathbf{B} is uniquely determined by \mathbf{E} , we will focus attention on the latter. Let us choose a coordinate system such that $\mathbf{e}_3 \parallel \mathbf{k}$. We may then write

$$\mathbf{E}(\mathbf{x}, t) = (E_1 \mathbf{e}_1 + E_2 \mathbf{e}_2) e^{ikx_3 - i\omega t},$$

where $\mathbf{k} \cdot \mathbf{x} = kx_3$ was used. Depending on the choice of the complex coefficients E_i , a number of physically different wave-types can be distinguished. To understand the different options, decompose $E_i = |E_i| \exp(i\phi_i)$ into a real amplitude, E_i and phase ϕ_i , to obtain

$$\text{Re } \mathbf{E}(\mathbf{x}, t) = \mathbf{e}_1 |E_1| \cos(kx_3 - \omega t + \phi_1) + \mathbf{e}_2 |E_2| \cos(kx_3 - \omega t + \phi_2). \quad (4.16)$$

Now let us look at a number of different limiting cases.

Linearly polarized waves

For identical phases $\phi_1 = \phi_2 = \phi$, we obtain

$$\text{Re } \mathbf{E} = (|E_1| \mathbf{e}_1 + |E_2| \mathbf{e}_2) \cos(kx_3 - \omega t + \phi),$$

i.e. a vector field **linearly polarized** in the direction of the vector $|E_1| \mathbf{e}_1 + |E_2| \mathbf{e}_2$.

▷ Info. Linear polarization is a hallmark of many artificial light sources, e.g. **laser light** is usually linearly polarized. Likewise, the radiation emitted by many **antennae** shows approximately linear polarization.

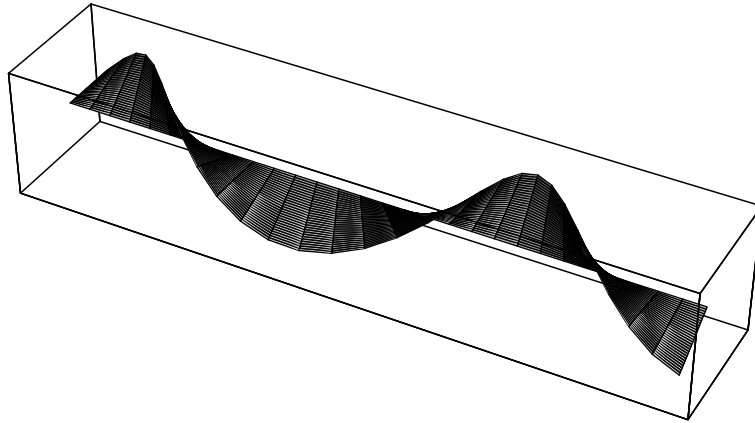


Figure 4.1: Schematic of a circularly polarized electric and magnetic field.

Circularly polarized waves

Next consider the case of a phase difference, $\phi_1 = \phi_2 \mp \pi/2 \equiv \phi$ and equal amplitudes $|E_1| = |E_2| = E$:

$$\text{Re } \mathbf{E}(\mathbf{x}, t) = E (\mathbf{e}_1 \cos(kx_3 - \omega t + \phi) \mp \mathbf{e}_2 \sin(kx_3 - \omega t + \phi)).$$

Evidently, the tip of the vector \mathbf{E} moves on a circle — \mathbf{E} is circularly polarized. Regarded as a function of \mathbf{x} (for any fixed instance of time), \mathbf{E} traces out a spiral whose characteristic period is set by the wave number $\lambda = 2\pi/k$, (see Fig. 4.1.) The sense of orientation of this spiral is determined by the sign of the phase mismatch. For $\phi_1 = \phi_2 + \pi/2$ ($\phi_1 = \phi_2 - \pi/2$) we speak of a right (left) circularly polarized wave or a wave of positive (negative) **helicity**.

▷ Info. As with classical point particles, electromagnetic fields can be subjected to a quantization procedure. In quantum electrodynamics it is shown that the quanta of the electromagnetic field, the **photons** carry definite helicity, i.e. they represent the minimal quantum of circularly polarized waves.

Elliptically polarized waves

Circular and linear polarization represent limiting cases of a more general form of polarization. Indeed, the minimal geometric structure capable of continuously interpolating between a line segment and a circle is the ellipse. To conveniently see the appearance of ellipses, consider the basis change, $\mathbf{e}_\pm = \frac{1}{\sqrt{2}}(\mathbf{e}_1 \pm i\mathbf{e}_2)$, and represent the electric field as

$$\mathbf{E}(\mathbf{x}, t) = (E_+ \mathbf{e}_+ + E_- \mathbf{e}_-) e^{ikx_3 - i\omega t}.$$

³One may ask why, then, did we introduce complex notation at all. The reason is that working with exponents of phases is way more convenient than explicitly having to distinguish between the sin and cos functions that arise after real parts have been taken.

In this representation, a circularly polarized wave corresponds to the limit $E_+ = 0$ (positive helicity) or $E_- = 0$ (negative helicity). Linear polarization is obtained by setting $E_+ = \pm E_-$. It is straightforward to verify that for generic values of the ratio $|E_-/E_+| \equiv r$ one obtains an **elliptically polarized** wave where the ratio between the major and the minor axis is set by $|1+r/1-r|$. The tilting angle α of the ellipse w.r.t. the 1-axis is given by $\alpha = \arg(E_-/E_+)$.

This concludes our discussion of the polarization of electromagnetic radiation. It is important to keep in mind that the different types of polarization discussed above represent limiting cases of what in general is only partially or completely un-polarized radiation. Radiation of a given value of the frequency, ω , usually involves the superposition of waves of different wave vector \mathbf{k} (at fixed wave number $k = |\mathbf{k}| = \omega c^{-1}$.) Only if the amplitudes of all partial waves share a definite phase/amplitude relation, do we obtain a polarized signal. The degree of polarization of a wave can be determined by computing the so-called Stokes parameters. However, we will not discuss this concept in more detail in this text.

4.4 Green function of the wave equation

4.4.1 Computation of the Green function

To prepare our discussion of the full problem (4.11), let us consider the inhomogeneous scalar wave equation

$$\left(\Delta - \frac{1}{c^2} \partial_t^2 \right) f = -4\pi g, \quad (4.17)$$

where the inhomogeneity g and the solution f are scalar function. As with the Poisson equation (3.4) the weak spot of the wave equation is its linearity. We may, therefore, again employ the concept of Green functions to simplify the solution of the problem. The Green function $G(\mathbf{x}, t; \mathbf{x}', t')$ is defined by obvious generalization of the Green function of electrostatics, (3.8) to a problem with space-time degrees of freedom.

$$\left(\Delta - \frac{1}{c^2} \partial_t^2 \right) G(\mathbf{x}, t; \mathbf{x}', t') = -\delta(\mathbf{x} - \mathbf{x}')\delta(t - t'). \quad (4.18)$$

Once the solution of this equation has been obtained (which requires specification of a set of boundary conditions), the solution of (4.17) becomes a matter of a straightforward integration:

$$f(\mathbf{x}, t) = 4\pi \int d^3x' dt' G(\mathbf{x}, t; \mathbf{x}', t') g(\mathbf{x}', t'). \quad (4.19)$$

Assuming vanishing boundary conditions at infinity $G(\mathbf{x}, t; \mathbf{x}', t') \xrightarrow{|\mathbf{x}-\mathbf{x}'|, |t-t'| \rightarrow \infty} 0$, we next turn to the solution of Eq.(4.18).

▷ **Math.** In view of the physical importance of the Green function, we present two alternative ways of deriving it, both emphasizing different aspects. The first utilizes methods of the **theory of complex functions**.

It starts from the solution of Eq. (4.18) in Fourier space. Application of Eq. (4.12) to both sides of the equation yields (show it)

$$\tilde{G}(\mathbf{k}, \omega) = \frac{1}{k^2 - c^{-2}\omega^2},$$

so 'all' that's left is the computation of the inverse transform

$$G(\mathbf{x}, t) = \int (d^3k) \int (d\omega) \frac{e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t}}{k^2 - c^{-2}\omega^2}.$$

Let us do the integral over frequency first, and consider the integral

$$I(k) \equiv -c^2 \int \frac{d\omega}{2\pi} \frac{e^{-i\omega t}}{(\omega - kc)(\omega + kc)}.$$

It now looks like we are hitting a disaster. The integrand contains two poles and looks ill-defined. We deal with this problem in a cavalier manner by interpreting the integral as one over a *complex* variable ω along the real axis. When running upon the poles, we decide to surround them along infinitesimal semi-circles. This, of course, introduces an ambiguity, surrounding in the upper or lower complex half planes. For reasons that will become evident momentarily, choose the upper plane. At infinity, $|\omega| \rightarrow \infty$, we close our integration contour along a huge semicircle in the *lower* complex plane. (We have to close in the lower plane, because in the upper plane $\exp(-i\omega t)$ would be divergent for positive t .) This defines the re-interpretation

$$I^+(k) \equiv -c^2 \Theta(t) \oint \frac{d\omega}{2\pi} \frac{e^{-i\omega t}}{(\omega - kc)(\omega + kc)},$$

where the superscript + indicates that we have met a choice, and the Heaviside function $\Theta(t)$ reminds us that the construction works for $t > 0$. We can now do the integral via the theorem of residues to obtain (exercise)

$$I^+(k) = \frac{ic}{2k} \Theta(t) \left(e^{-ikct} - e^{+ikct} \right).$$

Now substitute this result back into our integral,

$$G^+(\mathbf{x}, t) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} I^+(k).$$

Following a recipe previously applied in the computation of the Green function of electrostatics, we introduce polar coordinates and integrate over angles to obtain

$$\begin{aligned} G^+(\mathbf{x}, t) &= \frac{1}{(2\pi)^2 x i} \int_0^\infty k dk \left(e^{ikx} - e^{-ikx} \right) I^+(k) = \\ &= \frac{\Theta(t)c}{2(2\pi)^2 x} \int_0^\infty dk \left(e^{ikx} - e^{-ikx} \right) \left(e^{-ikct} - e^{+ikct} \right) = \\ &= \frac{\Theta(t)c}{4\pi x} (\delta(x - ct) - \delta(x + ct)) = \\ &= \frac{\Theta(t)}{4\pi x} \delta(t - x/c), \end{aligned}$$

where in the second last equality we integrated over the exponentials to obtain δ -functions, and in the second used that $\delta(x+ct) = 0$, because both, x and t are positive, and expressed the δ -function as a function of t , cancelling a factor of c .

We first note that for the chosen set of boundary conditions, the Green function $G(\mathbf{x} - \mathbf{x}', t - t')$ will depend on the difference of its arguments only. We next Fourier transform Eq.(4.18) in the temporal variable, i.e. we act on the equation with the integral transform $\tilde{f}(\omega) = \int dt \exp(i\omega t) f(t)$ (whose inverse is $f(t) = \int (d\omega) \exp(-i\omega t) \tilde{f}(\omega)$.) The temporally transformed equation is given by

$$(\Delta + k^2) G(\mathbf{x}, \omega) = -\delta(\mathbf{x}), \quad (4.20)$$

where we defined $k \equiv \omega/c$. If it were not for the constant k (and a trivial scaling factor $(2\pi)^{-1}$ on the l.h.s.) this equation, known in the literature as the **Helmholtz equation**, would be equivalent to the Poisson equation of electrostatics. Indeed, it is straightforward to verify that the solution of (4.20) is given by

$$G^\pm(\mathbf{x}, \omega) = \frac{1}{8\pi^2} \frac{e^{\pm ik|\mathbf{x}|}}{|\mathbf{x}|}, \quad (4.21)$$

where the sign ambiguity needs to be fixed on physical grounds.

▷ Math. To **prove Eq.(4.21)**, we introduce polar coordinates centered around \mathbf{x}' and act with the spherical representation of the Laplace operator (cf. section 8.1.2) on $G^\pm(r) = \frac{1}{8\pi^2} e^{\pm ikr}/r$. Noting that the radial part of the Laplace operator, $\Delta^\#$ is given by $r^{-2} \partial_r r^2 \partial_r = \partial_r^2 + (2/r) \partial_r$ and $\Delta^\#(4\pi r)^{-1} = -\delta(\mathbf{x})$ (the equation of the Green function of electrostatics), we obtain

$$\begin{aligned} (\Delta + k^2) G^\pm(r, \omega) &= \frac{1}{8\pi^2} (\partial_r^2 + 2r^{-1} \partial_r + k^2) \frac{e^{\pm ikr}}{r} = \\ &= \frac{e^{\pm ikr}}{8\pi^2} \left((\partial_r^2 + 2r^{-1} \partial_r) r^{-1} + 2e^{\mp ikr} (\partial_r r^{-1}) \partial_r e^{\pm ikr} + e^{\mp ikr} r^{-1} (\partial_r^2 + k^2 + 2r^{-1} \partial_r) e^{\pm ikr} \right) = \\ &= -\frac{1}{2\pi} \delta(\mathbf{x}) \mp 2ikr^{-2} \pm 2ikr^{-2} = -\frac{1}{2\pi} \delta(\mathbf{x}), \end{aligned}$$

as required.

Doing the inverse temporal Fourier transform, we obtain

$$G^\pm(\mathbf{x}, t) = \int d\omega e^{-i\omega t} G_{\omega/c}(\mathbf{x}) = \int d\omega e^{-i\omega t} \frac{1}{8\pi^2} \frac{e^{\pm i\omega c^{-1}|\mathbf{x}|}}{|\mathbf{x}|} = \frac{1}{4\pi} \frac{\delta(t \mp c^{-1}|\mathbf{x}|)}{|\mathbf{x}|},$$

or

$$\boxed{G^\pm(\mathbf{x} - \mathbf{x}', t - t') = \frac{1}{4\pi} \frac{\delta(t - t' \mp c^{-1}|\mathbf{x} - \mathbf{x}'|)}{|\mathbf{x} - \mathbf{x}'|}}. \quad (4.22)$$

For reasons to become clear momentarily, we call G^+ (G^-) the **retarded (advanced) Green function** of the wave equation.

4.4.2 Physical meaning of the Green function

Retarded Green function

To understand the physical significance of the retarded Green function G^+ , we substitute the r.h.s. of Eq.(4.22) into (4.19) and obtain

$$f(\mathbf{x}, t) = \int d^3x' \frac{g(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}. \quad (4.23)$$

For any fixed instance of space and time, (\mathbf{x}, t) , the solution $f(\mathbf{x}, t)$ is affected by the sources $g(\mathbf{x}', t')$ at all points in space and fixed *earlier* times $t' = t - |\mathbf{x} - \mathbf{x}'|/c$. Put differently, a time $t - t' = |\mathbf{x} - \mathbf{x}'|/c$ has to pass before the amplitude of the source $g(\mathbf{x}', t')$ may cause an effect at the observation point \mathbf{x} at time t — the signal received at (\mathbf{x}, t) is subject to a **retardation mechanism**. When the signal is received, it is so at a strength $g(\mathbf{x}', t)/|\mathbf{x} - \mathbf{x}'|$, similarly to the Coulomb potential in electrostatics. (Indeed, for a time independent source $g(\mathbf{x}', t) = g(\mathbf{x}')$, we may enter the wave equation with a time-independent ansatz $f(\mathbf{x})$, whereupon it reduces to the Poisson equation.) Summarizing, the sources act as ‘instantaneous Coulomb charges’ which are (a) of strength $g(\mathbf{x}', t')$ and felt at times $t = t' + |\mathbf{x} - \mathbf{x}'|/c$.

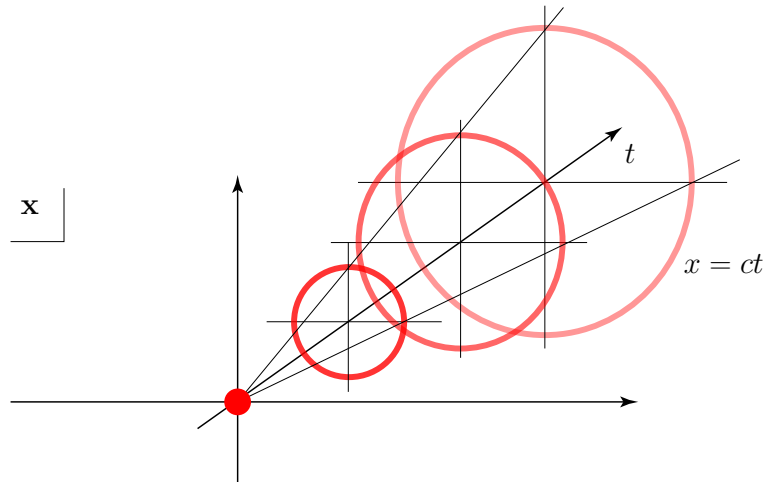


Figure 4.2: Wave front propagation of the pulse created by a point source at the origin schematically plotted as a function of two-dimensional space and time. The width of the front is set by $\delta x = c\delta t$, where δt is the duration of the time pulse. Its intensity decays as $\sim x^{-1}$.

▷ Example: By way of example, consider a point source at the origin, which ‘broadcasts’ a signal for a short instance of time at $t' \simeq 0$, $g(\mathbf{x}', t') = \delta(\mathbf{x}')F(t')$ where the function F is sharply peaked around $t' = 0$ and describes the temporal profile of the source. The signal is then given by $f(\mathbf{x}, t) = F(t - |\mathbf{x}|/c)/|\mathbf{x}|$, i.e. we obtain a pattern of out-moving **spherical waves**, whose amplitude diminishes as $|\mathbf{x}|^{-1}$ or, equivalently $\sim 1/tc$ (see Fig. 4.2.)

Advanced Green function

Consider now the solution we would have obtained from the advanced Green function, $f(\mathbf{x}, t) = \int d^3x' \frac{g(\mathbf{x}', t + |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}$. Here, the signal responds to the behaviour of the source in *the future*. The principle of cause-and-effect or **causality** is violated implying that the advanced Green function, though mathematically a legitimate solution of the wave equation, does not carry physical meaning. Two more remarks are in order: (a) when solving the wave equation by techniques borrowed from the theory of complex functions, the causality principle is built in from the outset, and the retarded Green function automatically selected. (b) Although the advanced Green function does not carry immanent physical meaning, its not a senseless object altogether. However, the utility of this object discloses itself only in quantum theories.

4.4.3 Electromagnetic gauge field and Green functions

So far we have solved the wave equation for an arbitrary scalar source. Let us now specialize to the wave equations (4.11) for the components of the electromagnetic potential, A_μ . To a first approximation, these are four independent scalar wave equations for the four sources j_μ . We may, therefore, just copy the prototypical solution to obtain the **retarded potentials of electrodynamics**

$$\begin{aligned}\phi(\mathbf{x}, t) &= \int d^3x' \frac{\rho(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}, \\ \mathbf{A}(\mathbf{x}, t) &= \frac{1}{c} \int d^3x' \frac{\mathbf{j}(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}.\end{aligned}\quad (4.24)$$

There is, however, one important consistency check that needs to be performed: recalling that the wave equations (4.11) hold only in the Lorenz gauge $\partial_\mu A^\mu = 0 \Leftrightarrow c^{-1} \partial_t \phi + \nabla \cdot \mathbf{A} = 0$, we need to check that this condition is actually fulfilled by the solutions (4.24). Relatedly, we have to keep in mind that the sources are not quite independent; they obey the continuity equation $\partial_\mu j^\mu = 0 \Leftrightarrow \partial_t \rho + \nabla \cdot \mathbf{j} = 0$.

As in section 3.3, it will be most convenient to probe the gauge behaviour of the vector potential in a fully developed Fourier language. Also, we will use a four vector notation throughout. In this notation, the Fourier transformation (4.12) assumes the form

$$\tilde{f}(k) = \frac{1}{(2\pi)^4} \int d^4x f(x) e^{ix_\mu k^\mu}, \quad (4.25)$$

where a factor of c^{-1} has been absorbed in the integration measure and $k_\mu = (k_0, \mathbf{k})$ with $k_0 = \omega/c$.⁴ The Fourier transform of the scalar wave equation (4.17) becomes $k_\mu k^\mu \psi(k) = -4\pi g(k)$. Specifically, the Green function obeys the equation $k_\mu k^\mu G(k) = -(2\pi)^{-4}$ which is solved by $G(k) = -((2\pi)^4 k_\mu k^\mu)^{-1}$. (One may check explicitly that this is the Fourier transform of our solution (4.22), however for our present purposes there is no need to do

⁴Recall that $x_0 = ct$, i.e. $k_\mu x^\mu = k_0 x_0 - \mathbf{k} \cdot \mathbf{x} = \omega t - \mathbf{k} \cdot \mathbf{x}$.

so.) The solution of the general scalar wave equation (4.19) obtains by *convoluting* the Green function and the source,

$$f(x) = 4\pi(G * g)(x) \equiv 4\pi \int d^4x' G(x - x')g(x').$$

Using the convolution theorem, this transforms to $f(k) = (2\pi)^4 4\pi G(k)g(k) = -\frac{4\pi g(k)}{k_\mu k^\mu}$. Specifically, for the vector potential, we obtain

$$A_\mu = -4\pi \frac{j_\mu}{k_\nu k^\nu}. \quad (4.26)$$

This is all we need to check that the gauge condition is fulfilled: The Fourier transform of the Lorenz gauge relation $\partial_\mu A^\mu = 0$ is given by $k_\mu A^\mu = k^\mu A_\mu = 0$. Probing this relation on (4.26), we obtain $k^\mu A_\mu \propto k^\mu j_\mu = 0$, where we noted that $k^\mu j_\mu = 0$ is the continuity relation.

We have thus shown that the solution (4.24) thus conforms with the gauge constraints. As an important byproduct, our proof above reveals an intimate connection between the gauge behaviour of the electromagnetic potential and current conservation.

Eqs. (4.24) generalize Eqs.(3.12) and (3.53) to the case of dynamically varying sources. In the next sections, we will explore various aspects of the physical contents of these equations.

4.5 Field energy and momentum

4.5.1 Field energy

In sections 3.2.2 and 4.1, we discussed the energy of static electric and magnetic fields. In this section we will identify the results obtained for the electric and the magnetic field energy, Eqs.(3.20) and (4.3), respectively, as building blocks of a larger picture.

We begin by performing a few basic operations on the Maxwell equations: multiplying Eq.(2.2) by $\mathbf{E} \cdot$ and Eq. (2.3) by $-\mathbf{H} \cdot$, and adding the results to each other, we obtain

$$\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \frac{1}{c}(\mathbf{E} \cdot \partial_t \mathbf{D} + \mathbf{H} \cdot \partial_t \mathbf{B}) = \frac{4\pi}{c} \mathbf{j} \cdot \mathbf{E}.$$

Using that for general vector field \mathbf{v} and \mathbf{w} , (check!) $\nabla \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{w} \cdot (\nabla \times \mathbf{v}) - \mathbf{v} \cdot (\nabla \times \mathbf{w})$, as well as $\mathbf{E} \cdot \partial_t \mathbf{D} = \partial_t(\mathbf{E} \cdot \mathbf{D})/2$ and $\mathbf{H} \cdot \partial_t \mathbf{B} = \partial_t(\mathbf{H} \cdot \mathbf{B})/2$, this equation can be rewritten as

$$\frac{1}{8\pi} \partial_t (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) + \frac{c}{4\pi} \nabla \cdot (\mathbf{E} \times \mathbf{H}) + \mathbf{j} \cdot \mathbf{E} = 0. \quad (4.27)$$

To understand the significance of Eq.(4.31), let us integrate it over a test volume V :

$$d_t \int_V d^3x \frac{1}{8\pi} (\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) = - \int_{S(V)} d\sigma \mathbf{n} \cdot \frac{c}{4\pi} (\mathbf{E} \times \mathbf{H}) - \int_V d^3x \mathbf{j} \cdot \mathbf{E}, \quad (4.28)$$

where use of Gauß's theorem has been made. The first term is the sum over the electric and the magnetic field energy density, as derived earlier for static field configurations. We now interpret this term as the energy stored in general electromagnetic field configurations and

$$w \equiv \frac{1}{8\pi}(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H}) \quad (4.29)$$

as the **electromagnetic energy density**. What is new in *electrodynamics* is that the field energy may change in time. The r.h.s. of the equation tells us that there are two mechanisms whereby the electromagnetic field energy may be altered. First, there is a surface integral over the so-called **Poynting vector field**

$$\mathbf{S} \equiv \frac{c}{4\pi} \mathbf{E} \times \mathbf{H}. \quad (4.30)$$

We interpret the integral over this vector as the energy current passing through the surface S and the Poynting vector field as the **energy current density**. Thus, the pair (energy density, w)/(Poynting vector \mathbf{S}) plays a role similar to the pair (charge density, ρ)/(current density, \mathbf{j}) for the matter field. However, unlike with matter, the energy of the electromagnetic field is not conserved. Rather, the **balance equation of electromagnetic field energy** above states that

$$\partial_t w + \nabla \cdot \mathbf{S} = -\mathbf{j} \cdot \mathbf{E}. \quad (4.31)$$

The r.h.s. of this equation contains the matter field \mathbf{j} which suggests that it describes the conversion of electromagnetic into mechanical energy. Indeed, the temporal change of the energy of a charged point particle in an electromagnetic field (cf. section 4.1 for a similar line of reasoning) is given by $d_t U(\mathbf{x}(t)) = -\mathbf{F} \cdot \dot{\mathbf{x}} = -q(\mathbf{E} + c^{-1}\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} = -q\mathbf{E} \cdot \mathbf{v}$, where we observed that the magnetic component of the Lorentz force is perpendicular to the velocity and, therefore, does not do work on the particle. Recalling that the current density of a point particle is given by $\mathbf{j} = q\delta(\mathbf{x} - \mathbf{x}(t))\mathbf{v}$, this expression may be rewritten as $d_t U = -\int d^3x \mathbf{j} \cdot \mathbf{E}$. However, the energy change of the particle is provided by the electromagnetic energy, so the latter changes by $+\int d^3x \mathbf{j} \cdot \mathbf{E}$. The r.h.s. of Eq. (4.30) is the generalization of this expression to arbitrary current densities.

Energy conservation implies that the work done on a current of charged particles has to be taken from the electromagnetic field. This explains the appearance of the mechanical energy on the r.h.s. of the balance equation (4.31).

4.5.2 Field momentum

*Unlike in previous sections on conservation laws, we here identify $\mathbf{B} = \mathbf{H}$, $\mathbf{D} = \mathbf{E}$, i.e. we consider the vacuum case.*⁵

⁵The discussion of the field momentum in matter (cf. ??) turns out to be a delicate matter, wherefore we prefer to stay on the safe ground of the vacuum theory.

As a second example of a physical quantity that can be exchanged between matter and electromagnetic fields, we consider momentum. According to Newton's equations, the change of the total mechanical momentum \mathbf{P}_{mech} carried by the particles inside a volume B is given by the integrated force density, i.e.

$$d_t \mathbf{P}_{\text{mech}} = \int_B d^3x \mathbf{f} = \int_B d^3x \left(\rho \mathbf{E} + \frac{1}{c} \mathbf{j} \times \mathbf{B} \right),$$

where in the second equality we inserted the Lorentz force density. Using Eqs. (2.1) and (2.2) to eliminate the sources we obtain

$$d_t \mathbf{P}_{\text{mech}} = \frac{1}{4\pi} \int_B d^3x \left((\nabla \cdot \mathbf{E}) \mathbf{E} - \mathbf{B} \times (\nabla \times \mathbf{B}) + c^{-1} \mathbf{B} \times \dot{\mathbf{E}} \right).$$

Now, using that $\mathbf{B} \times \dot{\mathbf{E}} = d_t(\mathbf{B} \times \mathbf{E}) - \dot{\mathbf{B}} \times \mathbf{E} = d_t(\mathbf{B} \times \mathbf{E}) - c \mathbf{E} \times (\nabla \times \mathbf{E})$ and adding $0 = \mathbf{B}(\nabla \cdot \mathbf{B})$ to the r.h.s., we obtain the symmetric expression

$$d_t \mathbf{P}_{\text{mech}} = \frac{1}{4\pi} \int_B d^3x \left((\nabla \cdot \mathbf{E}) \mathbf{E} - \mathbf{E} \times (\nabla \times \mathbf{E}) + \mathbf{B}(\nabla \cdot \mathbf{B}) - \mathbf{B} \times (\nabla \times \mathbf{B}) + c^{-1} d_t(\mathbf{B} \times \mathbf{E}) \right),$$

which may be reorganized as

$$\begin{aligned} d_t \left(\mathbf{P}_{\text{mech}} - \frac{1}{4\pi c} \int_B d^3x \mathbf{B} \times \mathbf{E} \right) &= \\ &= \frac{1}{4\pi} \int_B d^3x \left((\nabla \cdot \mathbf{E}) \mathbf{E} - \mathbf{E} \times (\nabla \times \mathbf{E}) + (\nabla \cdot \mathbf{B}) \mathbf{B} - \mathbf{B} \times (\nabla \times \mathbf{B}) \right). \end{aligned}$$

This equation is of the form $d_t(\text{something}) = (\text{something else})$. Comparing to our earlier discussion of conservation laws, we are led to interpret the 'something' on the l.h.s. as a conserved quantity. Presently, this quantity is the sum of the total mechanical momentum density \mathbf{P}_{mech} and the integral

$$\boxed{\mathbf{P}_{\text{field}} = \int d^3x \mathbf{g}, \quad \mathbf{g} \equiv \frac{1}{4\pi c} \mathbf{E} \times \mathbf{B} = \frac{\mathbf{S}}{c^2}.} \quad (4.32)$$

The structure of this expression suggests to interpret $\mathbf{P}_{\text{field}}$ as the **momentum carried by the electromagnetic field** and \mathbf{g} as the momentum density (which happens to be given by c^{-2} times the Poynting vector.) If our tentative interpretation of the equation above as a conservation law is to make sense, we must be able to identify its r.h.s. as a surface integral. This in turn requires that the components of the (vector valued) integrand be representable as a $X_j \equiv \partial_i T_{ij}$ i.e. as the divergence of a vector-field \mathbf{T}_j with components T_{ij} . (Here, $j = 1, 2, 3$ plays the role of a spectator index.) If this is the case, we may, indeed, transform the integral to a surface integral, $\int_B d^3x X_j = \int_B d^3x \partial_i T_{ij} = \int_{\partial B} d\sigma \mathbf{n} \cdot \mathbf{T}_j$. Indeed, it is not difficult to verify that

$$[(\nabla \cdot \mathbf{X}) - \mathbf{X} \times (\nabla \times \mathbf{X})]_j = \partial_i \left[X_i X_j - \frac{\delta_{ij}}{2} \mathbf{X} \cdot \mathbf{X} \right]$$

Identifying $\mathbf{X} = \mathbf{E}, \mathbf{B}$ and introducing the components of the **Maxwell stress tensor** as

$$T_{ij} = \frac{1}{4\pi} \left[E_i E_j + B_i B_j - \frac{\delta_{ij}}{2} (\mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B}) \right], \quad (4.33)$$

The r.h.s. of the conservation law assumes the form $\int_B d^3x \partial_i T_{ij} = \int_{\partial B} d\sigma n_i T_{ij}$, where n_i are the components of the normal vector field of the system boundary. The **law of the conservation of momentum** thus assumes the final form

$$d_t(\mathbf{P}_{\text{mech}} + \mathbf{P}_{\text{field}})_j = \int_{\partial B} d\sigma n_i T_{ij}. \quad (4.34)$$

Physically, $dt d\sigma n_i T_{ij}$ is the (j th component of the) momentum that gets pushed through $d\sigma$ in time dt . Thus, $d\sigma n_i T_{ij}$ is the momentum per time, or force exerted on $d\sigma$ and $n_i T_{ij}$ the force per area or **radiation pressure** due to the change of linear momentum in the system.

It is straightforward to generalize the discussion above to the conservation of **angular momentum**: The angular momentum carried by a mechanical system of charged particles may be converted into angular momentum of the electromagnetic field. It is evident from Eq. (4.32) that the angular momentum density of the field is given by

$$\mathbf{l} = \mathbf{x} \times \mathbf{g} = \frac{1}{4\pi c} \mathbf{x} \times (\mathbf{E} \times \mathbf{B}).$$

However, in this course we will not discuss angular momentum conservation any further.

4.5.3 Energy and momentum of plane electromagnetic waves

Consider a plane wave in vacuum. Assuming that the wave propagates in 3-direction, the physical electric field, \mathbf{E}_{phys} is given by

$$\mathbf{E}_{\text{phys.}}(\mathbf{x}, t) = \text{Re } \mathbf{E}(\mathbf{x}, t) = \text{Re} \left((E_1 \mathbf{e}_1 + E_2 \mathbf{e}_2) e^{ikx_3 - i\omega t} \right) = r_1 u_1(x_3, t) \mathbf{e}_1 + r_2 u_2(x_3, t) \mathbf{e}_2,$$

where $u_i(x_3, t) = \cos(\phi_i + kx_3 - \omega t)$ and we defined $E_i = r_i \exp(i\phi_i)$ with real r_i . Similarly, the magnetic field $B_{\text{phys.}}$ is given by

$$\mathbf{B}_{\text{phys.}}(\mathbf{x}, t) = \text{Re}(\mathbf{e}_3 \times \mathbf{E}(\mathbf{x}, t)) = r_1 u_1(x_3, t) \mathbf{e}_2 - r_2 u_2(x, t) \mathbf{e}_1.$$

From these relations we obtain the energy density and Poynting vector as

$$w(\mathbf{x}, t) = \frac{1}{8\pi} (\mathbf{E}^2 + \mathbf{B}^2) = \frac{1}{4\pi} ((r_1 u_1)^2 + (r_2 u_2)^2)(x_3, t)$$

$$\mathbf{S}(\mathbf{x}, t) = cw(\mathbf{x}, t) \mathbf{e}_3,$$

where we omitted the subscript 'phys.' for notational simplicity.

▷ EXERCISE. Check that w and \mathbf{S} above comply with the conservation law (4.31).

4.6 Electromagnetic radiation

To better understand the physics of the expressions (4.24), let us consider sources confined within a region in space of typical extension d . Without loss of generality, we assume the time dependence of the sources to be harmonic, with a certain characteristic frequency ω , $j_\mu(x) = j_\mu(\mathbf{x}) \exp(-i\omega t)$. (The signal generated by sources of more general time dependence can always be obtained by superposition of harmonic signals.) As we shall see, all other quantities of relevance to us (potentials, fields, etc.) will inherit the same time dependence, i.e. $X(\mathbf{x}, t) = X(\mathbf{x}) \exp(-i\omega t)$. Specifically, Eq. (4.24) then implies that the electromagnetic potentials, too, oscillate in time, $A_\mu(x) = A_\mu(\mathbf{x}) \exp(-i\omega t)$. Substituting this ansatz into Eq. (4.24), we obtain $A_\mu(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{j_\mu(\mathbf{x}') e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}$.

As in our earlier analyses of multipole fields, we assume that the observation point \mathbf{x} is far away from the source, $r = |\mathbf{x}| \gg d$. Under these circumstances, we may focus attention on the spatial components of the vector potential,

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x' \frac{\mathbf{j}(\mathbf{x}') e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x}-\mathbf{x}'|}, \quad (4.35)$$

where we have substituted the time-oscillatory ansatz for sources and potential into (4.24) and divided out the time dependent phases. From (4.35) the magnetic and electric fields are obtained as

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad \mathbf{E} = ik^{-1} \nabla \times (\nabla \times \mathbf{A}), \quad (4.36)$$

where in the second identity we used the source-free variant of the law of magnetic circulation, $c^{-1} \partial_t \mathbf{E} = -ik \mathbf{E} = \nabla \times \mathbf{B}$.

We may now use the smallness of the ratio d/r to expand $|\mathbf{x}-\mathbf{x}'| \simeq r - \mathbf{n} \cdot \mathbf{x}' + \dots$, where \mathbf{n} is the unit vector in \mathbf{x} -direction. Substituting this expansion into (4.35) and expanding the result in powers of r^{-1} , we obtain a series similar in spirit to the electric and magnetic multipole series discussed above. For simplicity, we here focus attention on the dominant contribution to the series, obtained by approximating $|\mathbf{x}-\mathbf{x}'| \simeq r$. For reasons to become clear momentarily, this term,

$$\mathbf{A}(\mathbf{x}) \simeq \frac{e^{ikr}}{cr} \int d^3x' \mathbf{j}(\mathbf{x}'), \quad (4.37)$$

generates **electric dipole radiation**. In section 3.3.2 we have seen that for a static current distribution, the integral $\int \mathbf{j}$ vanishes. However, for a dynamic distribution, we may engage the Fourier transform of the continuity relation, $i\omega \rho = \nabla \cdot \mathbf{j}$ to obtain

$$\int d^3x j_i = \int d^3x \nabla x_i \cdot \mathbf{j} = - \int d^3x' x_i (\nabla \cdot \mathbf{j}) = -i\omega \int d^3x' x_i \rho.$$

Substituting this result into (4.37) and recalling the definition of the electric dipole moment of a charge distribution, $\mathbf{d} \equiv \int d^3x \mathbf{x} \cdot \rho$, we conclude that

$$\boxed{\mathbf{A}(\mathbf{x}) \simeq -ik \mathbf{d} \frac{e^{ikr}}{r}}, \quad (4.38)$$

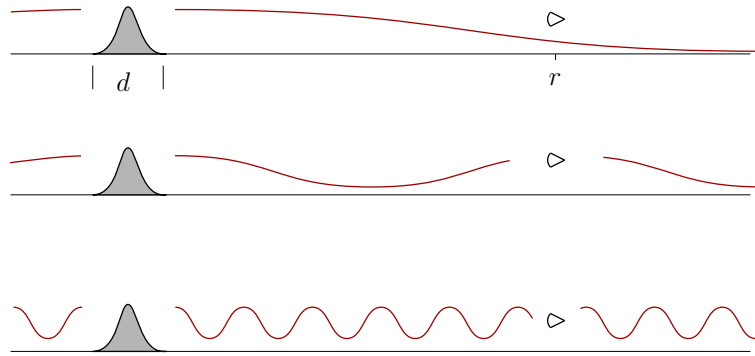


Figure 4.3: Three qualitatively distinct regimes of the radiation field. Discussion, see text.

is indeed controlled by the electric dipole moment.

Besides d and r another characteristic length scale in the problem is the characteristic **wave length** $\lambda \equiv 2\pi/k = 2\pi c/\omega$. For simplicity, we assume that the wave length $\lambda \gg d$ is much larger than the extent of the source.

▷ Info. This latter condition is usually met in practice. E.g. for radiation in the MHz range, $\omega \sim 10^6 \text{ s}^{-1}$ the wave length is given by $\lambda = 3 \cdot 10^8 \text{ m s}^{-1} / 10^6 \text{ s}^{-1} \simeq 300 \text{ m}$ much larger than the extension of typical antennae.

We then need to distinguish between three different regimes (see Fig. 4.3): the near zone, $d \ll r \ll \lambda$, the intermediate zone, $d \ll \lambda \sim r$, and the far zone, $d \ll \lambda \ll r$. We next discuss these regions in turn.

Near zone

For $r \ll \lambda$ or $kr \ll 1$, the exponential in (4.37) may be approximated by unity and we obtain

$$\mathbf{A}(\mathbf{x}, t) \simeq -i \frac{k}{r} \mathbf{d} e^{-i\omega t},$$

where we have re-introduced the time dependence of the sources. Using Eq.(4.36) to compute the electric and magnetic field, we get

$$\mathbf{B}(\mathbf{x}, t) = i \frac{k}{r^2} \mathbf{n} \times \mathbf{d} e^{-i\omega t}, \quad \mathbf{E} = \frac{\mathbf{n}(\mathbf{n} \cdot \mathbf{d}) - \mathbf{d}}{r^3} e^{-i\omega t}.$$

The electric field equals the dipole field (3.16) created by a dipole with time dependent moment $\mathbf{d} \exp(i\omega t)$. This motivates the denotation ‘electric dipole radiation’. The magnetic field is by a factor $kr \ll 1$ smaller than the electric field, i.e. in the near zone, the electromagnetic field is dominantly electric. In the limit $k \rightarrow 0$ the magnetic field vanishes and the electric field reduces to that of a static dipole.

The analysis of the **intermediate zone**, $kr \sim 1$ is complicated in as much as all powers in an expansion of the exponential in (4.37) in kr must be retained. For a discussion of the resulting field distribution, we refer to [1].

Far zone

For $kr \gg 1$, it suffices to let the derivative operations on the argument kr of the exponential function. Carrying out the derivatives, we obtain

$$\mathbf{B} = k^2 \mathbf{n} \times \mathbf{d} \frac{e^{i(kr - \omega t)}}{r}, \quad \mathbf{E} = -\mathbf{n} \times \mathbf{B}. \quad (4.39)$$

These asymptotic field distributions have much in common with the vacuum electromagnetic fields above. Indeed, one would expect that far away from the source (i.e. many wavelengths away from the source), the electromagnetic field resembles a spherical wave (by which we mean that the surfaces of constant phase form spheres.) For observation points sufficiently far from the center of the wave, its curvature will be nearly invisible and we the wave will look approximately planar. These expectations are met by the field distributions (4.39): neglecting all derivatives other than those acting on the combination kr (i.e. neglecting corrections of $\mathcal{O}(kr)^{-1}$), the components of \mathbf{E} and \mathbf{B} obey the wave equation (the Maxwell equations in vacuum.⁶ The wave fronts are spheres, outwardly propagating at a speed c . The vectors $\mathbf{n} \perp \mathbf{E} \perp \mathbf{B} \perp \mathbf{n}$ form an orthogonal set, as we saw is characteristic for a vacuum plane wave.

Finally, it is instructive to compute the **energy current** carried by the wave. To this end, we recall that the physically realized values of the electromagnetic field obtain by taking the real part of (4.39). We may then compute the Poynting vector (4.30) as

$$\mathbf{S} = \frac{ck^4}{4\pi r^2} \mathbf{n} (\mathbf{d}^2 - (\mathbf{n} \cdot \mathbf{d})^2) \cos^2(kr - \omega t) \xrightarrow{\langle \dots \rangle_t} \frac{ck^4}{8\pi r^2} \mathbf{n} (\mathbf{d}^2 - (\mathbf{n} \cdot \mathbf{d})^2),$$

Where the last expression is the energy current temporally averaged over several oscillation periods ω^{-1} . The energy current is maximal in the plane perpendicular to the dipole moment of the source and decays according to an inverse square law. It is also instructive to compute the **total power** radiated by the source, i.e. the energy current integrated over spheres of constant radius r . (Recall that the integrated energy current accounts for the change of energy inside the reference volume per time, i.e. the power radiated by the source.) Choosing the z -axis of the coordinate system to be colinear with the dipole moment, we have $\mathbf{d}^2 - (\mathbf{n} \cdot \mathbf{d})^2 = d^2 \sin^2 \theta$ and

$$P \equiv \int_S d\sigma \mathbf{n} \cdot \mathbf{S} = \frac{ck^4 d^2}{8\pi r^2} r^2 \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi \sin^2 \theta = \frac{ck^4 d^2}{3}.$$

Notice that the radiated power does not depend on the radius of the reference sphere, i.e. the work done by the source is entirely converted into radiation and not, say, in a steadily increasing density of vacuum field energy.

⁶Indeed, one may verify (do it!) that the characteristic factors $r^{-1} e^{i(kr - \omega t)}$ are exact solutions of the wave equations; they describe the space-time profile of a **spherical wave**.

▷ Info. As a concrete example of a radiation source, let us consider a center-fed **linear antenna**, i.e. a piece of wire of length a carrying an AC current that is maximal at the center of the wire. We model the current flow by the prototypical distribution $I(z) = I_0(1 - 2|z|a^{-1})\Theta(a - 2|z|)\exp(-i\omega t)$, where we $a \ll \lambda = c/\omega$ and alignment with the z -axis has been assumed. Using the continuity equation, $\partial_z I(z, t) + \partial_t \rho(z, t)$, we obtain the charge density (charge per unit length) in the wire as $\rho(z) = iI_0 2(\omega a)^{-1} \text{sgn}(z)\Theta(2|z| - a)\exp(-i\omega t)$. The dipole moment of the source is thus given by $\mathbf{d} = \mathbf{e}_z \int_{-a/2}^{a/2} dz z \rho(z, t) = iI_0(a/2\omega)\exp(-i\omega t)$, and the radiated power by

$$P = \frac{(ka)^2}{12c} I_0^2.$$

The coefficient $R_{\text{rad}} \equiv (ka)^2/12c$ of the factor I_0^2 is called **radiation resistance** of the antenna. To understand the origin of this denotation, notice that $[R_{\text{rad}}] = [c^{-1}]$ has indeed the dimension of resistivity (exercise.) Next recall that the power required to drive a current through a conventional resistor R is given by $P = UI = RI^2$. Comparison with the expression above suggests to interpret R_{rad} as the 'resistance' of the antenna. However, this resistance has nothing to do with dissipative energy losses *inside* the antenna. (i.e. those losses that hold responsible for the DC resistivity of metals.) Rather, work has to be done to feed energy into electromagnetic radiation. This work determines the radiation resistance. Also notice that $R_{\text{rad}} \sim k^2 \sim \omega^2$, i.e. the radiation losses increase quadratic in frequency. This latter fact is of immanent technological importance.

Our results above relied on a first order expansion in the ratio d/r between the extent of the sources and the distance of the observation point. We saw that at this order of the expansion, the source coupled to the electromagnetic field by its electric dipole moment. A more sophisticated description of the field may be obtained by driving the expansion in d/r to higher order. E.g. at next order, the electric quadrupole moment and the magnetic dipole moment enter the stage. This leads to the generation of **electric quadrupole radiation** and **magnetic dipole radiation**. For an in-depth discussion of these types of radiation we refer to [1].

5.0 Macroscopic electrodynamics

5.1 Macroscopic Maxwell equations

Suppose we want to understand the electromagnetic fields in an environment where extended pieces of matter are present. From a purist point of view, we would need to regard the $\mathcal{O}(10^{23})$ carriers of electric and magnetic moments — electrons, protons, neutrons — comprising the medium as sources. Throughout, we will denote the fields $\mathbf{e}, \dots, \mathbf{h}$ created by this system of sources as **microscopic fields**. (Small characters are used to distinguish the microscopic fields from the effective macroscopic fields to be introduced momentarily.) The ‘microscopic Maxwell equations’ read as

$$\begin{aligned}\nabla \cdot \mathbf{e} &= 4\pi\rho, \\ \nabla \times \mathbf{b} - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{e} &= \frac{4\pi}{c} \mathbf{j}, \\ \nabla \times \mathbf{e} + \frac{1}{c} \frac{\partial}{\partial t} \mathbf{b} &= 0, \\ \nabla \cdot \mathbf{b} &= 0.\end{aligned}$$

Now, for several reasons, it does not make much sense to consider these equations as such: First, it is clear that any attempts to get a system of $\mathcal{O}(10^{23})$ highly dynamical sources under control are bound to fail. Second, as a matter of principle, the dynamics of the microscopic sources is governed by quantum effects, i.e. we must not describe them in terms of a classical vector field \mathbf{j} . Finally, we aren’t even interested in knowing the microscopic fields. Rather, (in classical electrodynamics) we want to understand the behavior of fields on ‘classical’ length scales which generally exceed the atomic scales by far.¹ Let us, then, introduce **macroscopic fields** by averaging the microscopic fields as

$$\mathbf{E} \equiv \langle \mathbf{e} \rangle, \quad \mathbf{B} \equiv \langle \mathbf{b} \rangle,$$

where the averaging procedure is defined by $\langle g(\mathbf{x}) \rangle \equiv \int d^3x' g(\mathbf{x} - \mathbf{x}') f(\mathbf{x}')$, and f is a weight function that is unit normalized, $\int d^3x' f(\mathbf{x}') = 1$ and decays over sufficiently large regions in

¹For example, the wave-length of visible light is about 600 nm, whilst the typical extension of a molecule is $\mathcal{O}(1 \text{ nm})$. Thus, ‘classical’ length scales of interest are about two to three orders of magnitude larger than the microscopic scales.

space. Since the averaging procedure commutes with taking derivatives w.r.t. both space and time, $\partial_{t,\mathbf{x}}\mathbf{E} = \langle \partial_{t,\mathbf{x}}\mathbf{e} \rangle$ (and the same for \mathbf{B}), the averaged Maxwell equations read as

$$\begin{aligned}\nabla \cdot \mathbf{E} &= 4\pi \langle \rho \rangle, \\ \nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{E} &= \frac{4\pi}{c} \langle \mathbf{j} \rangle, \\ \nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial}{\partial t} \mathbf{B} &= 0, \\ \nabla \cdot \mathbf{B} &= 0.\end{aligned}$$

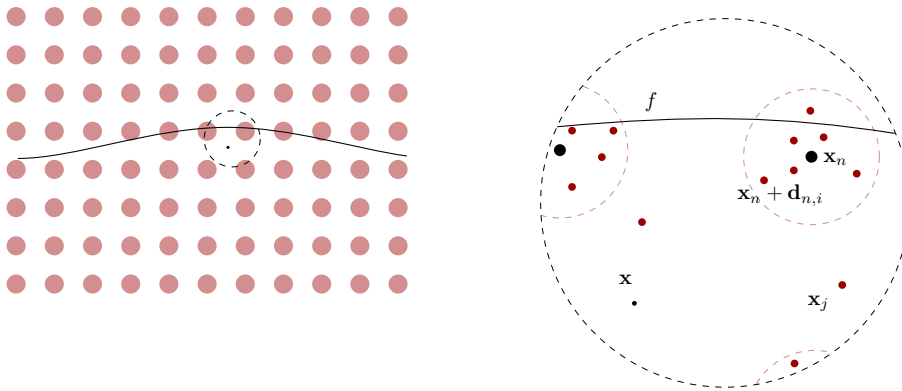


Figure 5.1: Schematic on the spatially averaged system of sources. A typical solid contains spatially localized molecules (indicated by the pattern of spheres) and (optionally) mobile electrons (the tiny inter-atomic dots.) The former can be polarized by external fields. Solid line: Cartoon of the averaging weight function f , centered around a reference point \mathbf{x} .

5.1.1 Averaged sources

To better understand the effect of the averaging on the sources, we need to take an (ever so superficial) look into the atomic structure of a typical solid. Generally, two different types of charges in solids have to be distinguished: first, there are charges — nuclei and valence electrons — that are by no more than a vector $\mathbf{a}_{m,j}$ off the center coordinate of a molecule (or atom for that matter) \mathbf{x}_m . Second, (in metallic systems) there are free conduction electrons which may abide at arbitrary coordinates $\mathbf{x}_i(t)$ in the system. Denoting the two contributions by ρ_b and ρ_f , respectively, we have $\rho_b(\mathbf{x}, t) = \sum_{m,j} q_{m,j} \delta(\mathbf{x} - \mathbf{x}_m(t) - \mathbf{a}_{m,j}(t))$ and $\rho_f = q_e \sum_i \delta(\mathbf{x} - \mathbf{x}_i(t))$, where $q_{m,j}$ is the charge of the i th molecular constituent and q_e the electron charge.

Averaged charge density and electric susceptibility

The density of the bound charges, averaged over macroscopic proportions, is given by

$$\begin{aligned}
 \langle \rho_b(\mathbf{x}) \rangle &= \int d\mathbf{x}' f(\mathbf{x}') \sum_{m,j} q_{m,j} \delta(\mathbf{x} - \mathbf{x}' - \mathbf{x}_m - \mathbf{a}_{m,j}) = \\
 &= \sum_{m,j} f(\mathbf{x} - \mathbf{x}_m - \mathbf{a}_{m,j}) q_{m,j} \simeq \\
 &\simeq \sum_m q_m f(\mathbf{x} - \mathbf{x}_m) - \nabla \sum_m f(\mathbf{x} - \mathbf{x}_m) \cdot \underbrace{\sum_j \mathbf{a}_{m,j} q_{m,j}}_{\mathbf{d}_m} = \\
 &= \sum_m q_m \langle \delta(\mathbf{x} - \mathbf{x}_m) \rangle - \nabla \cdot \mathbf{P}(\mathbf{x}).
 \end{aligned}$$

In the second line, we used the fact that the range over which the weight function f changes exceeds atomic extensions by far to Taylor expand to first order in the offsets \mathbf{a} . The zeroth order term of the expansion is oblivious to the molecular structure, i.e. it contains only the total charge $q_m = \sum_j q_{m,j}$ of the molecules and their center positions. The first order term contains the molecular dipole moments, $\mathbf{d}_m \equiv \sum_j \mathbf{a}_{m,j} q_{m,j}$. By the symbol

$$\mathbf{P}(\mathbf{x}, t) \equiv \left\langle \sum_m \delta(\mathbf{x} - \mathbf{x}_m(t)) \mathbf{d}_m(t) \right\rangle, \quad (5.1)$$

we denote the average **polarization of the medium**. Evidently, \mathbf{P} is a measure of the density of the dipole moments carried by the molecules in the system.

The average density of the mobile carriers in the system is given by $\langle \rho_f(\mathbf{x}, t) \rangle = \langle q_e \sum_i \delta(\mathbf{x} - \mathbf{x}_i(t)) \rangle$, so that we obtain the **average charge density** as

$$\langle \rho(\mathbf{x}, t) \rangle = \rho(\mathbf{x}, t) - \nabla \cdot \mathbf{P}(\mathbf{x}, t) \quad (5.2)$$

where we defined

$$\rho(\mathbf{x}, t) \equiv \left\langle \sum_m q_m \delta(\mathbf{x} - \mathbf{x}_m(t)) + q_e \sum_i \delta(\mathbf{x} - \mathbf{x}_i(t)) \right\rangle$$

to be the effective or macroscopic charge density in the system. Notice that the molecules/atoms present in the medium enter the quantity $\rho(\mathbf{x}, t)$ as point-like entities. Their finite polarizability is accounted for by the second term contributing to $\langle \rho \rangle$. Also notice that most solids are electrically neutral on average. I.e. in most cases, ρ will vanish, unless (a) external charges or charged impurities are present in the system, or (b) the charge distribution exhibits large scale collective oscillations.

Substituting Eq. (5.2) into the inhomogeneous Maxwell equations and rearranging terms, we obtain

$$\nabla \cdot (\mathbf{E} + 4\pi \mathbf{P}) = 4\pi \rho.$$

This equation suggests introducing a field

$$\boxed{\mathbf{D} \equiv \mathbf{E} + 4\pi\mathbf{P}.} \quad (5.3)$$

The sources of this **displacement field** is the net average charge density,

$$\nabla \cdot \mathbf{D} = 4\pi\rho, \quad (5.4)$$

(while the electric field has the spatial average of the full microscopic charge system as its source, $\nabla \cdot \mathbf{E} = 4\pi\langle\rho\rangle$.)

An obvious question to ask is where the medium polarization \mathbf{P} that makes the electric and the displacement field distinct actually comes from. The answer is that in most cases, a finite polarization is induced by the presence of an electric field, i.e. $\mathbf{P} = \mathbf{P}[\mathbf{E}]$. (Notice the feedback mechanism: an electric leads to polarization which in turn acts as a source for the electric field.)

▷ Info. On a microscopic level, at least two different **mechanisms causing field-induced polarization** need to be distinguished: Firstly, a finite electric field may cause a distortion of the electron clouds surrounding otherwise unpolarized atoms or molecules. The relative shift of the electron clouds against the nuclear centers causes a finite molecular dipole moment which integrates to a macroscopic polarization \mathbf{P} . Second, in many substances (mostly of organic chemical origin), the molecules (a) carry a permanent intrinsic dipole moment and (b) are free to move. A finite electric field will cause spatial alignment of the otherwise dis-oriented molecules. This leads to a polarization of the medium as a whole.

Now, external electric fields triggering a non-vanishing polarization are usually much weaker than the intrinsic microscopic fields keeping the constituents of a solid together. In practice, this means that the polarization may usually be approximated by a *linear* functional of the electric field. The most general form of a linear functional reads as²

$$\mathbf{P}(\mathbf{x}, t) = \int d^3x' \int dt' \chi(\mathbf{x}, t; \mathbf{x}', t') \mathbf{E}(\mathbf{x}', t'), \quad (5.5)$$

where the integral kernel χ is called the **electric susceptibility** of the medium. The susceptibility $\chi(\mathbf{x}, t; \mathbf{x}', t')$ describes how an electric field amplitude at \mathbf{x}' and $t' < t$ affects the polarization at (\mathbf{x}, t) .³ Notice that the polarization obtains by convolution of χ and \mathbf{E} . In Fourier space, the equation assumes the simpler form $\mathbf{P}(q) = \chi(q)\mathbf{E}(q)$, where $q = (\omega/c, \mathbf{q})$

²Yet more generally, one might allow for a non co-linear dependence of the polarization on the field vector:

$$P_i(\mathbf{x}, t) = \int d^3x' \int dt' \chi_{ij}(\mathbf{x}, t; \mathbf{x}', t') E_j(\mathbf{x}', t'),$$

where $\chi = \{\chi_{ij}\}$ is a 3×3 -matrix field.

³By causality, $\chi(\mathbf{x}, t; \mathbf{x}', t') \propto \Theta(t - t')$.

as usual. Accordingly, the electric field and the displacement field are connected by the linear relation

$$\mathbf{D}(q) = \epsilon(q)\mathbf{E}(q), \quad \epsilon(q) = 1 + 4\pi\chi(q), \quad (5.6)$$

where the function $\epsilon(q)$ is known as the **dielectric function**.

Information on the **frequency/momentum dependence of the dielectric function** has to be inferred from outside electrodynamics, typically from condensed matter physics, or plasma physics.⁴ In many cases of physical interest — e.g. in the physics of plasmas or the physics of metallic systems — the dependence of ϵ on both \mathbf{q} and ω has to be taken seriously. Sometimes, only the frequency dependence, or the dependence on the spatial components of momentum is of importance. However, the crudest approximation at all (often adopted in this course) is to approximate the function $\epsilon(q) \simeq \epsilon$ by a constant. (Which amounts to approximating $\chi(\mathbf{x}, t; \mathbf{x}', t') \propto \delta(\mathbf{x} - \mathbf{x}')\delta(t - t')$ by a function infinitely short ranged in space and time.) If not mentioned otherwise, we will thus assume

$$\mathbf{D} = \epsilon\mathbf{E},$$

where ϵ is a material constant.

Averaged current density and magnetic susceptibility

Averaging the current density proceeds in much the same way as the charge density procedure outlined above. As a result of a calculation (see the info block below) that is somewhat more laborious than the one above we obtain

$$\langle \mathbf{j} \rangle \simeq \mathbf{j} + \dot{\mathbf{P}} + c\nabla \times \mathbf{M}, \quad (5.7)$$

where

$$\mathbf{j} \equiv \left\langle q_e \sum \dot{\mathbf{x}}_i \delta(\mathbf{x} - \mathbf{x}_i) \right\rangle + \left\langle \sum_m q_m \dot{\mathbf{x}}_m \delta(\mathbf{x} - \mathbf{x}_m) \right\rangle$$

is the current carried by the free charge carriers and the point-like approximated molecules, respectively, and

$$\mathbf{M}(\mathbf{x}) = \left\langle \sum_m \delta(\mathbf{x} - \mathbf{x}_m) \frac{1}{2c} \sum_j q_{m,j} \mathbf{a}_{m,j} \times \dot{\mathbf{a}}_{m,j} \right\rangle \quad (5.8)$$

is the average **density of magnetic dipole moments** in the system.

▷ Info. To **compute the average of the microscopic current density**, we decompose $\mathbf{j} = \mathbf{j}_b + \mathbf{j}_f$ the current density into a bound and a free part. With $\mathbf{j}_b(x) = \sum_{m,j} (\dot{\mathbf{x}}_m + \dot{\mathbf{a}}_{m,j}) \delta(\mathbf{x} -$

⁴A **plasma** is a gas of charged particles.

$\mathbf{x}_m - \mathbf{a}_{m,j}$), the former is averaged as

$$\begin{aligned} \langle \mathbf{j}_b(\mathbf{x}) \rangle &= \int d\mathbf{x}' f(\mathbf{x}') \sum_{m,j} q_{m,j} (\dot{\mathbf{x}}_m + \dot{\mathbf{a}}_{m,j}) \delta(\mathbf{x} - \mathbf{x}' - \mathbf{x}_m - \mathbf{a}_{m,j}) = \\ &= \sum_{m,j} f(\mathbf{x} - \mathbf{x}_m - \mathbf{a}_{m,j}) (\dot{\mathbf{x}}_m + \dot{\mathbf{a}}_{m,j}) q_{m,j} \simeq \\ &\simeq \sum_{m,j} q_{m,j} \left[f(\mathbf{x} - \mathbf{x}_m) - \sum_m \nabla f(\mathbf{x} - \mathbf{x}_m) \cdot \mathbf{a}_{m,j} \right] (\dot{\mathbf{x}}_m + \dot{\mathbf{a}}_{m,j}). \end{aligned}$$

We next consider the different orders in \mathbf{a} contributing to this expansion in turn. At zeroth order, we obtain

$$\langle \mathbf{j}_b(\mathbf{x}) \rangle^{(0)} = \sum_m q_m f(\mathbf{x} - \mathbf{x}_m) \dot{\mathbf{x}}_m = \left\langle \sum_m q_m \dot{\mathbf{x}}_m \delta(\mathbf{x} - \mathbf{x}_m) \right\rangle,$$

i.e. the current carried by the molecules in a point-like approximation. The first order term is given by

$$\langle \mathbf{j}_b(\mathbf{x}) \rangle^{(1)} = \sum_m \left[f(\mathbf{x} - \mathbf{x}_m) \dot{\mathbf{d}}_m - (\nabla f(\mathbf{x} - \mathbf{x}_m) \cdot \mathbf{d}_m) \dot{\mathbf{x}}_m \right].$$

The form of this expression suggests to compare it with the time derivative of the polarization vector,

$$\dot{\mathbf{P}} = d_t \sum_m f(\mathbf{x} - \mathbf{x}_m) \mathbf{d}_m = \sum_m \left[f(\mathbf{x} - \mathbf{x}_m) \dot{\mathbf{d}}_m - (\nabla f(\mathbf{x} - \mathbf{x}_m) \cdot \dot{\mathbf{x}}_m) \mathbf{d}_m \right].$$

The difference of these two expressions is given by $\mathbf{X} \equiv \langle \mathbf{j}_b(\mathbf{x}) \rangle^{(1)} - \dot{\mathbf{P}} = \sum_m \nabla f(\mathbf{x} - \mathbf{x}_m) \times (\mathbf{d}_m \times \dot{\mathbf{x}}_m)$. By a dimensional argument, one may show that this quantity is negligibly small: the magnitude $|\dot{\mathbf{x}}_m| \sim v$ is of the order of the typical velocity of the atomic or electronic compounds inside the solid. We thus obtain the rough estimate $|\mathbf{X}(\mathbf{q}, \omega)| \sim v |\nabla \mathbf{P}(\mathbf{q}, \omega)| \sim v q |\mathbf{P}|$, where in the second step we generously neglected the differences in the vectorial structure of \mathbf{P} and \mathbf{X} , resp. However, $|(\dot{\mathbf{P}})(\mathbf{q}, \omega)| = \omega |\mathbf{P}(\mathbf{q}, \omega)|$. This means that $|\mathbf{X}|/|\dot{\mathbf{P}}| \sim v q / \omega \sim v/c$ is of the order of the ratio of typical velocities of non-relativistic matter and the speed of light. This ratio is so small that it (a) over-compensates the crudeness of our estimates above by far and (b) justifies to neglect the difference \mathbf{X} . We thus conclude that

$$\langle \mathbf{j}_b \rangle^{(1)} \simeq \dot{\mathbf{P}}.$$

The second order term contributing to the average current density is given by

$$\langle \mathbf{j}_b \rangle^{(2)} = - \sum_{m,j} q_{m,j} \mathbf{a}_{m,j} (\dot{\mathbf{a}}_{m,j} \cdot \nabla f(\mathbf{x} - \mathbf{x}_m))$$

This expression is related to the density of **magnetic dipole moments** carried by the molecules. The latter is given by (cf. Eq. (3.61) and Eq. (5.8))

$$\mathbf{M} = \frac{1}{2c} \sum_{m,j} q_{m,j} \mathbf{a}_{m,j} \times \dot{\mathbf{a}}_{m,j} f(\mathbf{x} - \mathbf{x}_m).$$

As a result of a straightforward calculation, we find that the curl of this expression is given by

$$\nabla \times \mathbf{M} = \frac{1}{c} \langle \mathbf{j}_b \rangle^{(2)} + d_t \frac{1}{2c} \sum_{m,j} q_{m,j} \mathbf{a}_{m,j} (\mathbf{a}_{m,j} \cdot \nabla f(\mathbf{x} - \mathbf{x}_m)).$$

The second term on the r.h.s. of this expression engages the time derivative of the electric *quadrupole* moments $\sim \sum q_a a_b$, $a, b = 1, 2, 3$ carried by the molecules. The fields generated by these terms are arguably very weak so that

$$\langle \dot{\mathbf{j}}_b \rangle^{(2)} \simeq c \nabla \times \mathbf{M}.$$

Adding to the results above the contribution of the free carriers $\langle \mathbf{j}_f(\mathbf{x}) \rangle = q_e \langle \sum_i \dot{\mathbf{x}}_i \delta(\mathbf{x} - \mathbf{x}_i) \rangle$, we arrive at Eq. (5.7).

Substituting Eq. (5.7) into the second inhomogeneous Maxwell equation, we obtain

$$\nabla \times (\mathbf{B} - 4\pi\mathbf{M}) - \frac{1}{c} \partial_t (\mathbf{E} + 4\pi\mathbf{P}) = \frac{4\pi}{c} \mathbf{j}.$$

The form of this equation motivates the definition of the **magnetic field**

$$\boxed{\mathbf{H} = \mathbf{B} - 4\pi\mathbf{M}.} \quad (5.9)$$

Expressed in terms of this quantity, the Maxwell equation assumes the form

$$\nabla \times \mathbf{H} - \frac{1}{c} \partial_t \mathbf{D} = \frac{4\pi}{c} \mathbf{j}. \quad (5.10)$$

According to this equation, the averaged (or macroscopic) current density acts as a source of the magnetic field. As with the electric sector of the theory, the magnetic fields created by externally imposed macroscopic current distributions (a) 'magnetize' solids, i.e. generate finite fields \mathbf{M} , and (b) are generally much weaker than the intrinsic microscopic fields inside a solid. This means that we can write $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}[\mathbf{H}]$, where $\mathbf{M}[\mathbf{H}]$ may be assumed to be linear in \mathbf{H} . In analogy to Eq. (5.5) we thus define

$$\mathbf{M}(\mathbf{x}, t) = \int d^3x' \int dt' \chi_m(\mathbf{x}, t; \mathbf{x}', t') \mathbf{H}(\mathbf{x}', t'), \quad (5.11)$$

where the function χ is called the **magnetic susceptibility** of the system. The magnetic susceptibility describes how a magnetic field at (\mathbf{x}', t') causes magnetization at (\mathbf{x}, t) . Everything we said above about the electric susceptibility applies similarly to the magnetic susceptibility. Specifically, we have the Fourier space relation $\mathbf{M}(q) = \chi_m(q) \mathbf{H}(q)$ and

$$\mathbf{B}(q) = \mu(q) \mathbf{H}(q), \quad \mu(q) = 1 + 4\pi\chi_m(q), \quad (5.12)$$

where $\mu(q)$ is called the **magnetic permeability**. In cases where the momentum dependence of the susceptibility is negligible, we obtain the simplified relation

$$\mathbf{B} = \mu \mathbf{H}, \quad (5.13)$$

where μ is a constant. As in the electric case, a number of different microscopic mechanisms causing deviations of μ off unity can be distinguished. Specifically, for molecules carrying no intrinsic magnetic moment, the presence of an external magnetic field may *induce* molecular

currents and, thus, a non-vanishing magnetization. This magnetization is generally directed opposite to the external field, i.e. $\mu < 1$. Substances of this type are called **diamagnetic**. (Even for the most diamagnetic substance known, bismuth, $1 - \mu = \mathcal{O}(10^{-4})$, i.e. diamagnetism is a very weak effect.) For molecules carrying a non-vanishing magnetic moment, an external field will cause alignment of the latter and, thus, an effective amplification of the field, $\mu > 1$. Materials of this type are called **paramagnetic**. Typical values of paramagnetic permeabilities are of the order of $\mu - 1 = \mathcal{O}(10^{-5}) - \mathcal{O}(10^{-2})$.

Before leaving this section, it is worthwhile to take a final look at the **general structure of the Maxwell equations in matter**. We note that

- ▷ The averaged charge densities and currents are sources of the electric displacement field \mathbf{D} and the magnetic field \mathbf{H} , respectively.
- ▷ These fields are related to the electric field \mathbf{E} and the magnetic induction \mathbf{B} by Eqs. (5.6) and (5.12), respectively.
- ▷ The actual fields one would measure in an experiment are \mathbf{E} and \mathbf{B} ; these fields determine the coupling (fields \rightsquigarrow matter) as described by the Lorentz force law.
- ▷ Similarly, the (matter un-related) homogeneous Maxwell equations contain the fields \mathbf{E} and \mathbf{B} .
- ▷ Notice a certain asymmetry in the definition of the electric and magnetic susceptibility, χ and χ_m , respectively. While the former is considered describes a linear relation $\mathbf{P} = \chi\mathbf{E}$ between polarization and the fundamental electric field, \mathbf{E} , the latter describes one $\mathbf{M} = \chi_m\mathbf{H}$ between the magnetic moment and the *external* field \mathbf{H} . These relations induce the likewise asymmetric ones, $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$. To the best of the author's knowledge, the reasons for this difference are historical.

5.2 Applications of macroscopic electrodynamics

In this sections, we will explore a number of phenomena caused by the joint presence of matter and electromagnetic fields. The organization of the section parallels that of the vacuum part of the course, i.e. we will begin by studying static electric phenomena, then advance to the static magnetic case and finally discuss a few applications in electrodynamics.

5.2.1 Electrostatics in the presence of matter

Generalities

Consider the macroscopic Maxwell equations of electrostatics

$$\begin{aligned}\nabla \cdot \mathbf{D} &= 4\pi\rho, \\ \nabla \times \mathbf{E} &= 0, \\ \mathbf{D} &= \epsilon\mathbf{E},\end{aligned}\tag{5.14}$$

where in the last equation we assumed a dielectric *constant* for simplicity. Now assume that we wish to explore the fields in the vicinity of a boundary separating two regions containing types of matter. In general, the dielectric constants characterizing the two domains will be different, i.e. we have $\mathbf{D} = \epsilon\mathbf{E}$ in region #1 and $\mathbf{D} = \epsilon'\mathbf{E}$ in region #2. (For $\epsilon' = 1$, we describe the limiting case of a matter/vacuum interface.) Optionally, the system boundary may carry a finite surface charge density η .

Arguing as in section 3.2.2, we find that the first and the second of the Maxwell equations imply, respectively,

$$\begin{aligned} (\mathbf{D}(\mathbf{x}) - \mathbf{D}'(\mathbf{x})) \cdot \mathbf{n}(\mathbf{x}) &= 4\pi\eta(\mathbf{x}), \\ (\mathbf{E}(\mathbf{x}) - \mathbf{E}'(\mathbf{x})) \times \mathbf{n}(\mathbf{x}) &= 0, \end{aligned} \quad (5.15)$$

where the unprimed/primed quantities refer to the fields in regions #1 and #2, respectively. Thus, the tangential component of the electric field is continuous, while the normal component of the displacement field jumps by an amount set by the charge density.

Example: Dielectric sphere in a homogeneous electric field

To illustrate the computation of electric fields in static environments with matter, consider the example of a massive sphere of radius R and dielectric constant ϵ placed in an homogeneous external displacement field \mathbf{D}_0 . We wish to compute the electric field \mathbf{E} in the entire medium.

Since there are no charges present, the electric potential inside and outside the sphere obeys the Laplace equation $\Delta\phi = 0$. Further, choosing polar coordinates such that the z -axis is aligned with the orientation of the external field, we expect the potential to be azimuthally symmetric, $\phi(r, \theta, \phi) = \phi(r, \theta)$. Expressed in terms of this potential, the boundary equations (5.15) assume the form

$$\epsilon\partial_r\phi|_{r=R-0} = \partial_r\phi|_{r=R+0}, \quad (5.16)$$

$$\partial_\theta\phi|_{r=R-0} = \partial_\theta\phi|_{r=R+0}, \quad (5.17)$$

$$\phi|_{r=R-0} = \phi|_{r=R+0}, \quad (5.18)$$

where $R\pm 0 \equiv \lim_{\eta \rightarrow 0} R \pm \eta$. To evaluate these conditions, we expand the potential inside and outside the sphere in a series of spherical harmonics, (3.43). Due to the azimuthal symmetry of the problem, only ϕ -independent functions $Y_{l,0} = \sqrt{\frac{2l+1}{4\pi}}P_l$ contribute, i.e.

$$\phi(r, \theta) = \sum_l P_l(\cos\theta) \times \begin{cases} A_l r^l & , r < R, \\ B_l r^l + C_l r^{-(l+1)} & , r \geq R, \end{cases},$$

where we have absorbed the normalization factors $(2l+1/4\pi)^{1/2}$ in the expansion coefficients A_l, B_l, C_l .

At spatial infinity, the potential must approach the potential $\phi = -Dz = -Dr \cos\theta = -DrP_1(\cos\theta)$ of the uniform external field $\mathbf{D}_0 = D\mathbf{e}_z$. Comparison with the series above

shows that $B_1 = -D$ and $B_{l \neq 1} = 0$. To determine the as yet unknown coefficients A_l and C_l , we consider the boundary conditions above. Specifically, Eq. (5.16) translates to the condition

$$\sum_l P_l(\cos \theta) (\epsilon l A_l R^{l-1} + (l+1)C_l R^{-(l+2)} + D\delta_{l,1}) = 0.$$

Now, the Legendre polynomials are a complete set of functions, i.e. the vanishing of the l.h.s. of the equation implies that all series coefficients must vanish individually:

$$\begin{aligned} C_0 &= 0, \\ \epsilon A_1 + 2R^{-3}C_1 + D &= 0, \\ \epsilon l A_l R^{l-1} + (l+1)C_l R^{-(l+2)} &= 0, \quad l > 1. \end{aligned}$$

A second set of equations is obtained from Eq. (5.17): $\partial_\theta \sum_l P_l(\cos \theta) ((A_l - B_l)R^l - C_l R^{-(l+1)}) = 0$ or $\sum_{l>0} P_l(\cos \theta) ((A_l - B_l)R^l - C_l R^{-(l+1)}) = \text{const.}$. The second condition implies (think why!) $(A_l - B_l)R^l - C_l R^{-(l+1)} = 0$ for all $l > 0$:

$$\begin{aligned} (A_1 - B_1)R + R^{-2}C_1 &= 0, \\ A_l R^l - C_l R^{-(l+1)} &= 0, \quad l > 1. \end{aligned}$$

The simultaneous solution of the two sets of $l > 1$ equations is possible only for $A_l = C_l = 0$, $l > 1$. The two $l = 1$ equations are solved by

$$C_1 = DR^3 \frac{\epsilon - 1}{\epsilon + 2}, \quad A_1 = -\frac{3}{\epsilon + 2}D.$$

Finally, the continuity of the potential requires $A_0 = 0$. Summarizing, the potential distorted by the presence of the sphere is given by

$$\phi(r, \theta) = -rD \cos \theta \times \begin{cases} \frac{3}{\epsilon + 2} & , r < R \\ 1 - \frac{\epsilon - 1}{\epsilon + 2} \left(\frac{R}{r}\right)^3 & , r \geq R, \end{cases} \quad (5.19)$$

The result (5.19) affords a very intuitive physical interpretation:

- ▷ Inside the sphere the electric field $\mathbf{E} = -\nabla\phi = \nabla Dz \frac{3}{\epsilon+2} = D \frac{3}{\epsilon+2} \mathbf{e}_z$ is parallel to the external electric field, but weaker in magnitude (as long as $\epsilon > 1$ which usually is the case.) This indicates the buildup of a polarization $\mathbf{P} = \frac{\epsilon-1}{4\pi} \mathbf{E} = \frac{3}{4\pi} \frac{\epsilon-1}{\epsilon+2} \mathbf{D}_0$.

Outside the sphere, the electric field is given by a superposition of the external field and the field generated by a dipole potential $\phi = \mathbf{d} \cdot \mathbf{x}/r^3 = d \cos \theta / r^2$, where the dipole moment is given by $\mathbf{d} = V\mathbf{P}$ and $V = 4\pi R^3/3$ is the volume of the sphere.

To understand the origin of this dipole moment, notice that in a polarizable medium, and in the absence of external charges, $0 = \nabla \cdot \mathbf{D} = \nabla \cdot \mathbf{E} + 4\pi \nabla \cdot \mathbf{P}$, while the microscopic

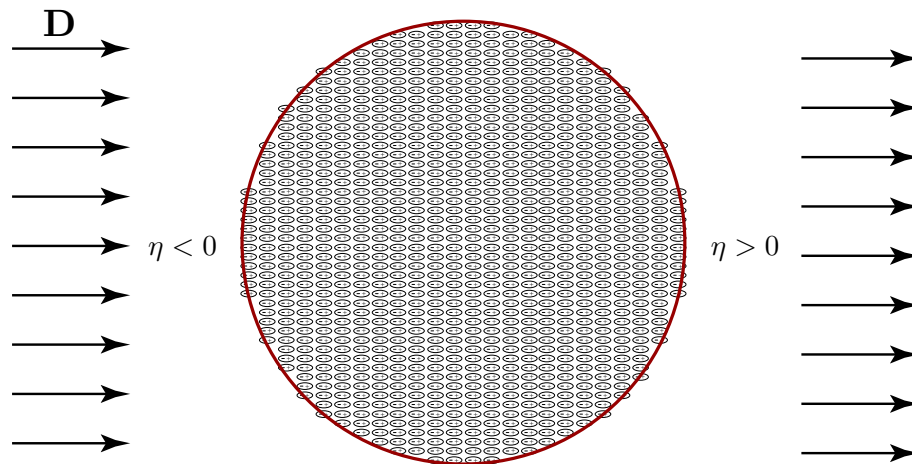


Figure 5.2: Polarization of a sphere by an external electric field.

charge density $\nabla \cdot \mathbf{E} = 4\pi\rho_{\text{mic}}$ need not necessarily vanish. In exceptional cases where ρ_{mic} does not vanish upon averaging, we denote $\langle \rho_{\text{mic}} \rangle = \rho_{\text{pol}}$ the **polarization charge**. The polarization charge is determined by the condition

$$\nabla \cdot \mathbf{P} = -\rho_{\text{pol}}.$$

Turning back to the sphere, the equation $0 = \nabla \cdot \mathbf{D} = \frac{\epsilon}{\chi} \nabla \cdot \mathbf{P}$ tells us that there is no polarization charge in the bulk of the sphere (nor, of course, outside the sphere.) However, arguing as in section 3.2.2, we find that the sphere (or any boundary between two media with different dielectric constants) carries a **surface polarization charge**

$$-\eta_{\text{pol}} = P_{\text{out}} - P_{\text{in}}.$$

Specifically, in our problem, $P_{\text{out}} = 0$ so that $\eta_{\text{in}} = P_{\text{in}} = \frac{3}{4\pi} \frac{\epsilon-1}{\epsilon+2} E_0 \cos \theta$. Qualitatively, the origin of this charge is easy enough to understand: while in the bulk of the sphere, the induced dipole moments cancel each other upon spatial averaging (see the figure), uncompensated charges remain at the surface of the medium. These surface charges induce a finite and quite macroscopic dipole moment which is aligned opposite to \mathbf{E}_0 and acts as a source of the *electric* field (but not of the displacement field).

5.2.2 Magnetostatics in the presence of matter

Generalities

The derivation of ‘magnetic boundary conditions’ in the presence of materials with a finite permeability largely parallels the electric case: the macroscopic equations of magnetostatics

read as

$$\begin{aligned}\nabla \times \mathbf{H} &= \frac{4\pi}{c} \mathbf{j}, \\ \nabla \cdot \mathbf{B} &= 0, \\ \mathbf{H} &= \mu^{-1} \mathbf{B}.\end{aligned}\tag{5.20}$$

Again, we wish to explore the behaviour of the fields in the vicinity of a boundary between two media with different permeabilities. Proceeding in the usual manner, i.e. by application of infinitesimal variants of Gauß' and Stokes law, respectively, we derive the equations

$$\begin{aligned}(\mathbf{B}(\mathbf{x}) - \mathbf{B}'(\mathbf{x})) \cdot \mathbf{n} &= 0, \\ (\mathbf{H}(\mathbf{x}) - \mathbf{H}'(\mathbf{x})) \times \mathbf{n} &= \frac{4\pi}{c} \mathbf{j}_s,\end{aligned}\tag{5.21}$$

where \mathbf{j}_s is an optional surface current.

Example: paramagnetic sphere in a homogeneous magnetic field

Consider a sphere of radius R and permeability $\mu > 1$ in the presence of an external magnetic field \mathbf{H} . We wish to compute the magnetic induction \mathbf{B} inside and outside the sphere. In principle, this can be done as in the analogous problem on the dielectric sphere, i.e. by Legendre polynomial series expansion. However, comparing the magnetic Maxwell equations to electric Maxwell equations considered earlier,

$$\begin{aligned}\nabla \cdot \mathbf{B} = 0 &\leftrightarrow \nabla \cdot \mathbf{D} = 0, \\ \nabla \times \mathbf{H} = 0 &\leftrightarrow \nabla \times \mathbf{E} = 0, \\ \mathbf{B} = \mu \mathbf{H} &\leftrightarrow \mathbf{D} = \epsilon \mathbf{E}, \\ \mathbf{P} = \frac{\epsilon - 1}{4\pi} \mathbf{E} &\leftrightarrow \mathbf{M} = \frac{\mu - 1}{4\pi} \mathbf{H},\end{aligned}$$

we notice that there is actually no need to do so: Formally identifying $\mathbf{B} \leftrightarrow \mathbf{D}$, $\mathbf{H} \leftrightarrow \mathbf{E}$, $\mu \leftrightarrow \epsilon$, $\mathbf{M} \leftrightarrow \mathbf{P}$, the two sets of equation become equivalent, and we may just copy the solution obtained above. Thus (i) the magnetic field outside the sphere is a superposition of the external field and a magnetic dipole field, where (ii) the dipole moment $\mathbf{M} = \frac{3}{4\pi} \frac{\mu - 1}{\mu + 2} \mathbf{H}$ is parallel to the external field. This magnetic moment is caused by the orientation of the intrinsic moments along the external field axis; consequently, the actually felt magnetic field (the magnetic induction) $\mathbf{B} = \mathbf{H} + 4\pi \mathbf{M}$ exceeds the external field.

5.3 Wave propagation in media

What happens if an electromagnetic wave impinges upon a medium characterized by a non-trivial dielectric function ϵ ?⁵ And how does the propagation behavior of waves relate to actual

⁵We emphasize dielectric function because, as shall be seen below, the magnetic permeability has a much lesser impact on electromagnetic wave propagation.

physical processes inside the medium? To prepare the discussion of these questions, we will first introduce a physically motivated model for the dependence of the dielectric function on its arguments. This modeling will establish a concrete link between the amplitude of the electromagnetic waves and the dynamics of the charge carriers inside the medium.

5.3.1 Model dielectric function

In Fourier space, the dielectric function, $\epsilon(\mathbf{q}, \omega)$ is a function of both wave vector and frequency. It owes its frequency dependence to the fact that a wave in matter may prompt the creation of excitations of the molecular degrees of freedom. The feedback of these excitations to the electromagnetic wave is described by the frequency dependence of the function ϵ . To get an idea of the relevant frequency scales, notice that typical excitation energies in solids are of the order $\hbar\omega < 1 \text{ eV}$. Noting that Planck's constant $\hbar \simeq 0.6 \times 10^{-15} \text{ eVs}$, we find that the characteristic frequencies at which solids dominantly respond to electromagnetic waves are of the order $\omega < 10^{15} \text{ s}^{-1}$. However, the wave lengths corresponding to these frequencies, $\lambda = 2\pi/|\mathbf{q}| \sim 2\pi c/\omega \sim 10^{-6} \text{ m}$ are much larger than the range of a few interatomic spacings over which we expect the non-local feedback of the external field into the polarization to be 'screened' (the range of the susceptibility χ .) This means (think why!) that the function $\epsilon(\mathbf{q}, \omega) \simeq \epsilon(\omega)$ is largely momentum-independent at the momentum scales of physical relevance.

To obtain a crude model for the function $\epsilon(\omega)$, we imagine that the electrons surrounding the molecules inside the solid are harmonically bound to the nuclear center coordinates. An individual electron at spatial coordinate \mathbf{a} (measured w.r.t. a nuclear position) will then be subject to the equation of motion

$$m(d_t^2 + \omega_0^2 + \gamma d_t) \mathbf{a}(t) = e\mathbf{E}(t), \quad (5.22)$$

where ω_0 is the characteristic frequency of the oscillator motion of the electron, γ a relaxation constant, and the variation of the external field over the extent of the molecule has been neglected (cf. the discussion above.) Fourier transforming this equation, we obtain $\mathbf{d} \equiv e\mathbf{a}(\omega) = e^2\mathbf{E}(\omega)m^{-1}/(\omega_0^2 - \omega^2 - i\gamma\omega)$. The dipole moment $d_m(\omega)$ of an entire molecule with Z electrons f_i of which oscillate with frequency ω_i and damping rate γ_i ($\sum_i f_i = Z$) is given by $\mathbf{d}_m(\omega) = \sum_i f_i \mathbf{d}_{(\omega_0 \rightarrow \omega_i, \gamma \rightarrow \gamma_i)}$. Denoting the density of molecules in the by n , we thus obtain (cf. Eqs. (5.1), (5.3), and (5.6))

$$\epsilon(\omega) = 1 + \frac{4\pi n e^2}{m} \sum_i \frac{f_i}{\omega_i^2 - \omega^2 - i\gamma_i \omega}. \quad (5.23)$$

With suitable quantum mechanical definitions of the material constants f_i, ω_i and γ_i , the **model dielectric function** (5.23) provides an accurate description of the molecular contribution to the dielectric behaviour of solids. A schematic of the typical behaviour of real and imaginary part of the dielectric function is shown in Fig. 5.3. A few remarks on the profile of these functions:

- ▷ The material constant ϵ employed in the ‘static’ sections of this chapter is given by $\epsilon = \epsilon(0)$. The imaginary part of the zero frequency dielectric function is negligible.
- ▷ For reasons to become clear later on, we call the function

$$\alpha(\omega) = \frac{2\omega}{c} \text{Im} \sqrt{\mu\epsilon(\omega)} \quad (5.24)$$

the (frequency dependent) **absorption coefficient** and

$$n(\omega) = \text{Re} \sqrt{\mu\epsilon(\omega)} \quad (5.25)$$

the (frequency dependent) **refraction coefficient** of the system.

- ▷ For most frequencies, i.e. everywhere except for the immediate vicinity of a resonant frequency ω_i , the absorption coefficient is negligibly small. In these regions, $d_\omega \text{Re} \epsilon(\omega) > 0$. The positivity of this derivative defines a region of **normal dispersion**.
- ▷ In the immediate vicinity of a resonant frequency, $|\omega - \omega_i| < \gamma_i/2$, the absorption coefficient shoots up while the derivative $d_\omega \text{Re} \epsilon(\omega) < 0$ changes sign. As we shall discuss momentarily, the physics of these frequency windows of **anomalous dispersion** is governed by resonant absorption processes.

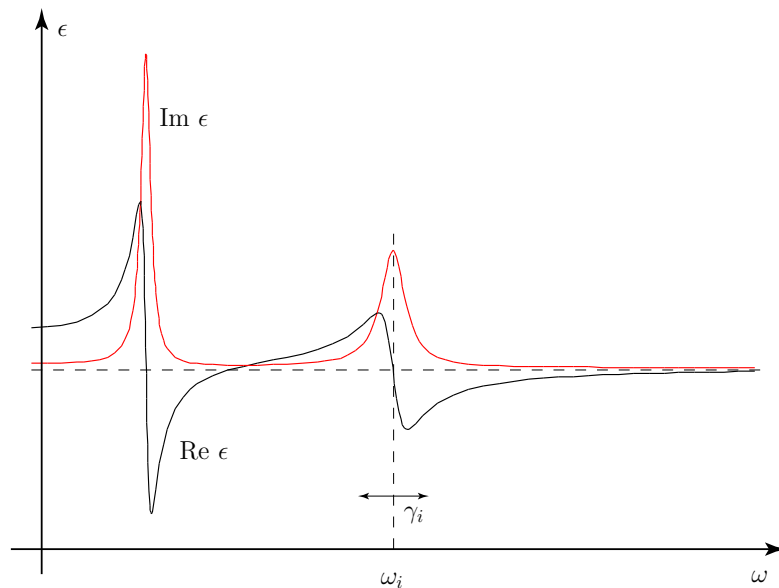


Figure 5.3: Schematic of the functional profile of the function $\epsilon(\omega)$. Each resonance frequency ω_i is center of a peak in the imaginary part of ϵ and of a resonant swing of in the real part. The width of these structures is determined by the damping rate γ_i .

5.3.2 Plane waves in matter

To study the impact of the non-uniform dielectric function on the behaviour of electromagnetic waves in matter, we consider the spatio-temporal Fourier transform of the Maxwell equations in a medium free of extraneous sources, $\rho = 0$, $\mathbf{j} = 0$. Using that $\mathbf{D}(k) = \epsilon(k)\mathbf{E}(k)$ and $\mathbf{H}(k) = \mu^{-1}(k)\mathbf{B}(k)$, where $k = (\omega/c, \mathbf{k})$, we have

$$\begin{aligned} \mathbf{k} \cdot (\epsilon \mathbf{E}) &= 0, \\ \mathbf{k} \times (\mu^{-1} \mathbf{B}) + \frac{\omega}{c} (\epsilon \mathbf{E}) &= 0, \\ \mathbf{k} \times \mathbf{E} - \frac{\omega}{c} \mathbf{B} &= 0, \\ \mathbf{k} \cdot \mathbf{B} &= 0, \end{aligned}$$

where we omitted momentum arguments for notational clarity. We next compute $\mathbf{k} \times$ (third equation) + $(\omega\mu/c) \times$ (second equation) and $(\omega/c) \times$ (third equation) - $\mu\mathbf{k} \times$ (second equation) to obtain the **Fourier representation of the wave equation**

$$\left(k^2 - \frac{\omega^2}{c^2(\omega)} \right) \times \begin{cases} \mathbf{E}(\mathbf{k}, \omega) \\ \mathbf{B}(\mathbf{k}, \omega) \end{cases} = 0, \quad (5.26)$$

where

$$\boxed{c(\omega) = \frac{c}{\sqrt{\mu\epsilon(\omega)}}}, \quad (5.27)$$

is the **effective velocity of light in matter**.

Comparing with our discussion of section 4.3, we conclude that a **plane electric wave in a matter** is mathematically described by the function

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_0 \exp(i(\mathbf{k} \cdot \mathbf{x} - \omega t)),$$

where $\mathbf{k} = k\mathbf{n}$ and $k = \omega/c(\omega)$ is a (generally complex) function of the wave frequency. Specifically, for $\text{Im}(\mathbf{k} \cdot \mathbf{x}) > 0$, the wave will be exponentially damped. Splitting the wave number k into its real and imaginary part, we have

$$k = \text{Re } k + i \text{Im } k = \frac{\omega}{c} n(\omega) + \frac{i}{2} \alpha(\omega),$$

where n and α are the refraction and absorption coefficient, respectively. To better understand the meaning of the absorption coefficient, let us compute the Poynting vector of the electromagnetic wave. As in section 4.3, the Maxwell equation $\nabla \cdot \mathbf{D} = \epsilon \nabla \cdot \mathbf{E} \propto \mathbf{k} \cdot \mathbf{E}_0$ implies transversality of the electric field. (Here we rely on the approximate \mathbf{x} -independence of the dielectric function.) Choosing coordinates such that $\mathbf{n} = \mathbf{e}_3$ and assuming planar polarization for simplicity, we set $\mathbf{E}_0 = E_0 \mathbf{e}_1$ with a real coefficient E_0 . The — physically relevant — real part of the electric field is thus given by

$$\text{Re } \mathbf{E} = E_0 \mathbf{e}_1 \cos(\omega(zn/c - t)) e^{-\alpha z/2}.$$

From the Maxwell equations $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{E} + c^{-1} \partial_t \mathbf{B}$, we further obtain (exercise) $\text{Re } \mathbf{B} \simeq E_0 \frac{c}{\text{Re } c(\omega)} \mathbf{e}_2 \cos(\omega(zn/c - t)) e^{-\alpha z/2}$. Assuming that $c(\omega)$ is approximately real⁶ and that the permeability is close to unity, $\mu \simeq 1$ or $\mathbf{B} \simeq \mathbf{H}$, we obtain for the magnitude of the Poynting vector

$$|\mathbf{S}| = \frac{c}{4\pi} |\mathbf{E} \times \mathbf{H}| \sim \cos^2(\omega(zn/c - t)) e^{-\alpha z} \xrightarrow{\langle \dots \rangle_t} e^{-\alpha z},$$

where in the last step we have averaged over several intervals of the oscillation period $2\pi/\omega$. According to this result,

The electromagnetic energy current inside a medium decays at a rate set by the absorption coefficient.

This phenomenon is easy enough to interpret: according to our semi-phenomenological model of the dielectric function above, the absorption coefficient is non-vanishing in the immediate vicinity of a resonant frequency ω_i , i.e. a frequency where molecular degrees of freedom oscillate. The energy stored in an electromagnetic wave of this frequency may thus get converted into mechanical oscillator energy. This goes along with a loss of field intensity, i.e. a diminishing of the Poynting vector or, equivalently, a loss of electromagnetic energy density, w .

5.3.3 Dispersion

In the previous section we have focused on the imaginary part of the dielectric function and on its attenuating impact on individual plane waves propagating in matter. We next turn to the discussion of the — equally important — role played by the real part. To introduce the relevant physical principles in a maximally simple environment, we will focus on a one-dimensional model of wave propagation throughout. Consider, thus, the one-dimensional variant of the wave equations (5.26),

$$\left(k^2 - \frac{\omega^2}{c^2(\omega)} \right) \psi(k, \omega) = 0, \quad (5.28)$$

where k is a one-dimensional wave 'vector' and $c(\omega) = c/\sqrt{\epsilon(\omega)}$ as before. (For simplicity, we set $\mu = 1$ from the outset.) Plane wave solutions to this equation are given by $\psi(x, t) = \psi_0 \exp(ik(\omega)x - i\omega t)$, where $k(\omega) = \omega/c(\omega)$.⁷ Notice that an alternative representation of the plane wave configurations reads as $\psi(x, t) = \psi_0 \exp(ikx - i\omega(k)t)$, where $\omega(k)$ is defined implicitly, viz. as a solution of the equation $\omega/c(\omega) = k$.

⁶If this assumption is violated, \mathbf{E} and \mathbf{B} are no longer in phase, i.e. the \cos -time dependence of \mathbf{B} picks up a phase relative to that of \mathbf{E} . However, this generalization does not change the essence of the argument.

⁷There is also the 'left moving' solution, $\psi(x, t) = \psi_0 \exp(ik(\omega)x + i\omega t)$, but for our purposes it will be sufficient to consider right moving waves.

To understand the consequences of frequency dependent variations in the real part of the dielectric function, we need, however, to go beyond the level of isolated plane wave. Rather, let us consider a superposition of plane waves (cf. Eq. (4.13)),

$$\psi(x, t) = \int dk \psi_0(k) e^{+i(kx - \omega(k)t)}, \quad (5.29)$$

where $\psi_0(k)$ is an arbitrary function. (Exercise: check that this superposition solves the wave equations.)

Specifically, let us choose the function $\psi_0(k)$ such that, at initial time $t = 0$, the distribution $\psi(x, t = 0)$ is localized in a finite region in space. Configurations of this type are called **wave packets**. While there is a lot of freedom in choosing such spatially localizing envelope functions, a convenient (and for our purposes sufficiently general) choice of the weight function ψ_0 is a Gaussian,

$$\psi_0(k) = \psi_0 \exp\left(- (k - k_0)^2 \frac{\xi^2}{4}\right),$$

where $\psi_0 \in \mathbb{C}$ is a amplitude coefficient fixed and ξ a coefficient of dimensionality 'length' that determines the spatial extent of the wave package (at time $t = 0$.) To check that latter assertion, let us compute the spatial profile $\psi(x, 0)$ of the wave package at the initial time:

$$\begin{aligned} \psi(x, 0) &= \psi_0 \int dk e^{-(k-k_0)^2 \frac{\xi^2}{4} + ikx} = \psi_0 e^{ik_0 x} \int dk e^{-k^2 \frac{\xi^2}{4} + ikx} = \\ &= \psi_0 e^{ik_0 x} \int dk e^{-k^2 \frac{\xi^2}{4} - (x/\xi)^2} = \frac{2\sqrt{\pi}}{\xi} \psi_0 e^{ik_0 x} e^{-(x/\xi)^2}. \end{aligned}$$

The function $\psi(x, 0)$ is concentrated in a volume of extension ξ ; it describes the profile of a small wave 'packet' at initial time $t = 0$. What will happen to this wave packet as time goes on?

To answer this question, let us assume that the scales over which the function $\omega(k)$ varies are much larger than the extension of the wave package in wave number space, k . It is, then, a good idea to expand $\omega(k)$ around k_0 :

$$\omega(k) = \omega_0 + v_g(k - k_0) + \frac{a}{2}(k - k_0)^2 + \dots,$$

where we introduced the abbreviations $\omega_0 \equiv \omega(k_0)$, $v_g \equiv \omega'(k_0)$ and $a \equiv \omega''(k_0)$. Substituting this expansion into Eq. (5.29) and doing the Gaussian integral over wave numbers we obtain

$$\begin{aligned} \psi(x, t) &= \psi_0 \int dk e^{-(k-k_0)^2 \xi^2/4 + i(kx - \omega(k)t)} \simeq \\ &\simeq \psi_0 e^{ik_0(x - v_p t)} \int dk e^{-(\xi^2 + 2iat)(k-k_0)^2/4 + i(k-k_0)(x - v_g t)} = \\ &\simeq \frac{2\sqrt{\pi}\psi_0}{\xi(t)} e^{ik_0(x - v_p t)} e^{-((x - v_g t)/\xi(t))^2}, \end{aligned}$$

where we introduced the function $\xi(t) = \sqrt{\xi^2 + 2iat}$, and the abbreviation $v_p \equiv \omega(k_0)/k_0$

Let us try to understand the main characteristics of this result:

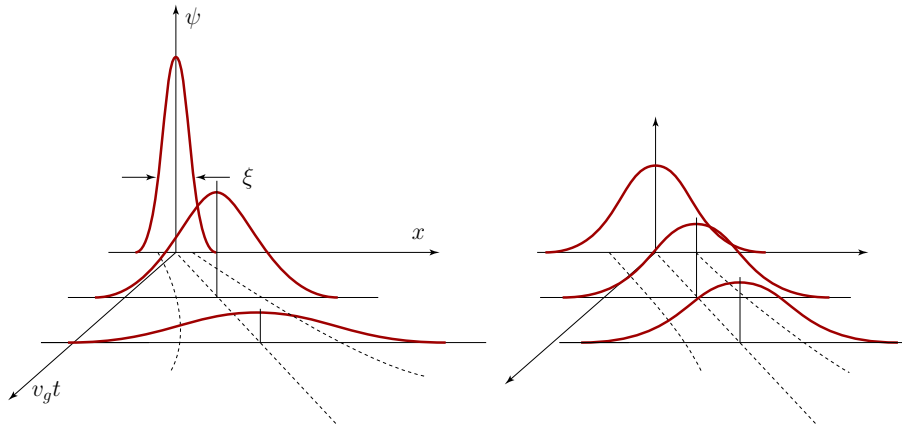


Figure 5.4: Left: Dispersive propagation of an initially sharply focused wave package. Right: A more shallow wave package suffers less drastically from dispersive deformation.

- ▷ The center of the wave package moves with a characteristic velocity $v_g = \partial_k \omega(k_0)$ to the right. In as much as v_g determines the effective center velocity of a superposition of a continuum or ‘group’ of plain waves, v_g is called the **group velocity** of the system. Only in vacuum where $c(\omega) = c$ is independent of frequency we have $\omega = ck$ and the group velocity $v_g = \partial_k \omega = c$ coincides with the vacuum velocity of light.

A customary yet far less useful notion is that of a **phase velocity**: an individual plane wave behaves as $\sim \exp(i(kx - \omega(k)t)) = \exp(ik(x - (\omega(k)/k)t))$. In view of the structure of the exponent, one may call $v_p \equiv \omega(k)/k$ the ‘velocity’ of the wave. Since, however, the wave extends over the entire system anyway, the phase velocity is not of direct physical relevance. Notice that the phase velocity $v_p = \omega(k)/k = c(k) = c/n(\omega)$, where the refraction coefficient has been introduced in (5.25). For most frequencies and in almost all substances, $n > 1 \Rightarrow v_p < c$. In principle, however, the phase velocity may become larger than the speed of light.⁸ The group velocity $v_g = \partial_k \omega(k) = (\partial_\omega k(\omega))^{-1} = (\partial_\omega \omega n(\omega))^{-1} c = (n(\omega) + \omega \partial_\omega n(\omega))^{-1} c = v_p (1 + n^{-1} \omega \partial_\omega n)^{-1}$. In regions of normal dispersion, $\partial_\omega n > 0$ implying that $v_g < v_p$. For the discussion of anomalous cases, where v_g may exceed the phase velocity or even the vacuum velocity of light, see [1].

- ▷ The **width of the wave package**, Δx changes in time. Inspection of the Gaussian factor controlling the envelope of the packet shows that $\text{Re } \psi \sim e^{(x-v_g t)^2 (\xi^{-2} + \xi^{*-2})/2}$, where the symbol ‘ \sim ’ indicates that only the exponential dependence of the package is taken into account. Thus, the width of the package is given by $\xi(t) = \xi(1 + (2at/\xi^2)^2)^{1/2}$.

According to this formula, the rate at which $\xi(t)$ increases (the wave package flows apart) is the larger, the sharper the spatial concentration of the initial wave package

⁸However, there is no need to worry that such anomalies conflict with Einsteins principle of relativity to be discussed in chapter 6 below: the dynamics of a uniform wave train is not linked to the transport of energy or other observable physical quantities, i.e. there is no actual physical entity that is transported at a velocity larger than the speed of light.

was (cf. Fig. 5.4). For large times, $t \gg \xi^2/a$, the width of the wave package increases as $\Delta x \sim t$. The disintegration of the wave package is caused by a phenomenon called **dispersion**: The form of the package at $t = 0$ is obtained by (Fourier) superposition of a large number of plane waves. If all these plane waves propagated with the same phase velocity (i.e. if $\omega(k)$ was k -independent), the form of the wave package would remain un-changed. However, due to the fact that in a medium plane waves of different wave number k generally propagate at different velocities, the Fourier spectrum of the package at non-zero times differs from that at time zero. Which in turn means that the integrity of the form of the wave package gets gradually lost.

The dispersive spreading of electromagnetic wave packages is a phenomenon of immense applied importance. For example, dispersive deformation is one of the main factors limiting the information load that can be pushed through **fiber optical cables**. (For too high loads, the 'wave packets' constituting the bit stream through such fibers begin to overlap thus losing their identity. The construction of ever more sophisticated counter measures optimizing the data capacity of optical fibers represents a major stream of applied research.

5.3.4 Electric conductivity

As a last phenomenon relating to macroscopic electrodynamics, we discuss the electric conduction properties of metallic systems.

Empirically, we know that in a metal, a finite electric field will cause current flow. In its most general form, the relation between field and current assumes a form similar to that between field and polarization discussed above:

$$j_i(\mathbf{x}, t) = \int d^3x' \int_{-\infty}^t dt' \sigma_{ij}(\mathbf{x} - \mathbf{x}', t - t') E_j(\mathbf{x}', t'), \quad (5.30)$$

where $\sigma = \{\sigma_{ij}(\mathbf{x}, t)\}$ is called the **conductivity tensor**. This equation states that a field may cause current flow at different spatial locations in time. Equally, a field may cause current flow in *directions* different from the field vector. (For example, in a system subject to a magnetic field in z -direction, an electric field in x -direction will cause current flow in both x - and y -direction, why?) As with the electric susceptibility, the non-local space dependence of the conductivity may often be neglected, $\sigma(\mathbf{x}, t) \propto \delta(\mathbf{x})$, or $\sigma(\mathbf{q}, \omega) = \sigma(\omega)$ in Fourier space. However, except for very low frequencies, the ω -dependence is generally important. Generally, one calls the finite-frequency conductivity 'AC conductivity', where 'AC' stands for 'alternating current'. For $\omega \rightarrow 0$, the conductivity crosses over to the 'DC conductivity', where 'DC' stands for 'directed current'. In the DC limit, and for an isotropic medium, the field-current relation assumes the form of **Ohm's law**

$$\mathbf{j} = \sigma \mathbf{E}. \quad (5.31)$$

In the following, we wish to understand how the phenomenon of electrical conduction can be explained from our previous considerations.

There are two different ways to specify the dynamical response of a metal to external fields: We may either postulate a current–field relation in the spirit of Ohms law. Or we may determine a dielectric function which (in a more microscopic way) describes the response of the mobile charge carriers in the system to the presence of a field. However, since that response generally implies current flow, the simultaneous specification of both a frequency dependent⁹ dielectric function and a current–field relation would over–determine the problem. In the following, we discuss the two alternatives in more detail.

The phenomenological route

Let us impose Eq. (5.31) as a condition extraneous to the Maxwell equations; We simply postulate that an electric field drives a current whose temporal and spatial profile is rigidly locked to the electric field. As indicated above, this postulate largely determines the dynamical response of the system to the external field, i.e. the dielectric function $\epsilon(\omega) = \epsilon$ has to be chosen constant. Substitution of this ansatz into the temporal Fourier transform of the Maxwell equation (2.2) obtains

$$\nabla \times \mathbf{H}(\mathbf{x}, \omega) = \frac{1}{c} (-i\epsilon\omega + 4\pi\sigma) \mathbf{E}(\mathbf{x}, \omega), \quad (5.32)$$

For the moment, we leave this result as it is and turn to

The microscopic route

We want to describe the electromagnetic response of a metal in terms of a model dielectric function. The dielectric function (5.23) constructed above describes the response of charge carriers harmonically bound to molecular center coordinates by a harmonic potential $\sim m_i\omega_i^2/2$ and subject to an effective damping or friction mechanism. Now, the conduction electrons of a metal may be thought of as charge carriers whose confining potential is infinitely weak, $\omega_i = 0$, so that their motion is not tied to a reference coordinate. Still, the conduction electrons will be subject to friction mechanisms. (E.g., scattering off atomic imperfections will impede their ballistic motion through the solid.) We thus model the dielectric function of a metal as

$$\epsilon(\omega) = 1 - \frac{4\pi n_e e^2}{m\omega(\omega + \frac{i}{\tau})} + \frac{4\pi n_e e^2}{m} \sum_i \frac{f_i}{\omega_i^2 - \omega^2 - i\gamma_i\omega} \stackrel{\omega \ll \omega_i}{\simeq} \epsilon - \frac{4\pi n_e e^2}{m\omega(\omega + \frac{i}{\tau})}, \quad (5.33)$$

where $n_e \equiv Nf_0$ is a measure of the concentration of conduction electrons, the friction coefficient of the electrons has been denoted by τ^{-1} ,¹⁰ and in the last step we noted that for frequencies well below the oscillator frequencies ω_i of the valence electrons, the frequency dependence of the second contribution to the dielectric function may be neglected; We thus lump that contribution into an effective material constant ϵ . Keep in mind that the free electron

⁹As we shall see below, a dielectric *constant* does not lead to current flow.

¹⁰... alluding to the fact that the attenuation of the electrons is due to collisions of impurities which take place at a rate denoted by τ .

contribution to the dielectric function exhaustingly describes the acceleration of charge carriers by the field (the onset of current), i.e. no extraneous current–field relations must be imposed.

Substitution of Eq. (5.33) into the Maxwell equation (2.2) obtains the result

$$\nabla \times \mathbf{H}(\mathbf{x}, \omega) = -\frac{i\omega\epsilon(\omega)}{c} \mathbf{E}(\mathbf{x}, \omega) = \frac{1}{c} \left(-i\omega\epsilon + \frac{4\pi n_e e^2}{m(-i\omega + \frac{1}{\tau})} \right) \mathbf{E}(\mathbf{x}, \omega), \quad (5.34)$$

Comparison to the phenomenological result (5.32) yields the identification

$$\sigma(\omega) = \frac{n_e e^2}{m(-i\omega + \frac{1}{\tau})}, \quad (5.35)$$

i.e. a formula for the **AC conductivity** in terms of electron density and mass, and the impurity collision rate.

Eq. (5.35) affords a very intuitive physical interpretations:

- ▷ For **high frequencies**, $\omega \gg \tau^{-1}$, we may approximate $\sigma \sim (i\omega)^{-1}$, or $\omega \mathbf{j} \sim \mathbf{E}$. In time space, this means $\partial_t \mathbf{j} \sim \rho \dot{\mathbf{v}} \sim \mathbf{E}$. This formula describes the ballistic acceleration of electrons by an electric field: On time scales $t \ll \tau$, which, in a Fourier sense, correspond to large frequencies $\omega \gg \tau^{-1}$ the motion of electrons is not hindered by impurity collisions and they accelerate as free particles.
- ▷ For **low frequencies**, $\omega \ll \tau^{-1}$, the conductivity may be approximated by a constant, $\sigma \sim \tau$. In this regime, the motion of the electrons is impeded by repeated impurity collisions. As a result they diffuse with a constant drift induced by the electric field.

6.0 Relativistic invariance

6.1 Introduction

6.1.1 Galilei invariance and its limitations

The laws of mechanics are the same in two coordinate systems K and K' moving at a constant velocity relative to one another. Within classical Newtonian mechanics, the space time coordinates (t, \mathbf{x}) and (t', \mathbf{x}') of two such systems are related by a **Galilei transformation**,

$$\begin{aligned}t' &= t \\ \mathbf{x}' &= \mathbf{x} - \mathbf{v}t.\end{aligned}$$

Substituting this transformation into Newton's equations of motion as formulated in system K ,

$$K : \quad \frac{d^2 \mathbf{x}_i}{dt^2} = -\nabla_{\mathbf{x}_i} \sum_j V_{ij}(\mathbf{x}_i - \mathbf{x}_j)$$

(\mathbf{x}_i are particle coordinates in system K and V_{ij} is a pair potential which, crucially, depends only on the differences between particle coordinates and not, say, on any distinguished reference point in universe.) we find that

$$K' : \quad \frac{d^2 \mathbf{x}'_i}{dt'^2} = -\nabla_{\mathbf{x}'_i} \sum_j V_{ij}(\mathbf{x}'_i - \mathbf{x}'_j),$$

i.e. the equations remain form invariant. This is what is meant when we say that classical mechanics is Galilei invariant.¹

Now, suppose a certain physical phenomenon in K (the dynamics of a water surface, say) is effectively described by a wave equation,

$$K : \quad \left(\Delta - \frac{1}{c^2} \partial_t^2 \right) f(\mathbf{x}, t) = 0.$$

¹More precisely, Galilei invariance implies invariance under all transformations of the Galilei group, the uniform motions above, rotations of space, and translations of space.

Now apply the chain rule

$$\begin{aligned}\partial_{x^i} &= \frac{\partial x^{j'}}{\partial x^i} \partial_{x^{j'}} + \frac{\partial t'}{\partial x^i} \partial_{t'} = \partial_{x^{i'}}, \\ \partial_t &= \frac{\partial x^{i'}}{\partial t} \partial_{x^{i'}} + \frac{\partial t'}{\partial t} \partial_{t'} = v^i \partial_{x^{i'}} + \partial_{t'},\end{aligned}$$

to obtain the transformed equation

$$K' : \quad \left(\Delta' - \frac{1}{c^2} \partial_{t'}^2 - \frac{2}{c^2} \mathbf{v} \cdot \nabla' \partial_{t'} - \frac{1}{c^2} (\mathbf{v} \cdot \nabla')^2 \right) f(\mathbf{x}', t') = 0,$$

where $f(\mathbf{x}', t') = f(\mathbf{x}(\mathbf{x}', t'), t)$. The equation has changed form. At first sight this result may look worrisome. After all, many waves (water waves, sound waves, etc.) are of pure mechanical origin which means that wave propagation ought to be a Galilei invariant phenomenon. The resolution to this problem lies with the fact, that matter waves generally propagate in a host medium (water, say.) While in K this medium is at rest, it isn't in K' . But a wave propagating in a non-stationary medium must be controlled by a different equation of motion, i.e. there is no a priori contradiction to the principle of Galilei invariance. Equivalently, one may observe that a wave emitted by a point source stationary in K will propagate with velocity c (in K .) However, an observer in K' will observe wave fronts propagating with different velocities.² Mathematically, this distortion of the wave pattern is described by K' 's 'wave equation'.

But what about electromagnetic waves? So far, we have never talked about a 'host medium' supporting the propagation of electromagnetic radiation. The lack of Galilean invariance of Maxwells equations then leaves us with three possibilities:

1. Nature is Galilean invariant, but Maxwells equations are incorrect (or, to put it more mildly, incomplete.)
2. An electromagnetic host medium (let's call it the **ether**) does, indeed, exist.
3. Nature is invariant under a group of space-time transformations different from Galilei. If so, Newton's equations of motion are in need of modification.

These questions became pressing in the second half of the 19th century, after Maxwell had brought the theory of electromagnetism to completion, and the exploration of all kinds of electromagnetic wave equation represented a focal point of theoretical and experimental research. In view of the sensational success of Maxwells theory the first answer was not really considered an option.

Since Newtonian mechanics wasn't a concept to be sacrificed lightly either, the existence of an ether was postulated. However, it was soon realized that the ether medium would have to have quite eccentric physical properties. Since no one had ever actually observed an

²The actually measured velocity depends on the distance of the observation point to the point of origin of the wave.

ether, this medium would have to be infinitely light and fully non-interacting with matter. (Its presumed etheric properties actually gave it its name.) Equally problematic, electromagnetic waves would uniformly propagate with the speed of light, c , only in the distinguished ether rest frame. I.e., the ether postulate was tantamount to the abandonment of the idea of 'no preferential inertial frames in nature' a concept close at heart to the development of modern physics. Towards the end of the nineteenth century, experimental evidence was mounting that the vacuum velocity of light was, independent of the observation frame, given by c ; the idea of an ether became increasingly difficult to sustain.

6.1.2 Einsteins postulates

In essence, this summarizes the state of affairs before Einstein entered the debate. In view of the arguments outlined above, Einstein decided that 3. had to be the correct option. Specifically, he put forward two fundamental postulates

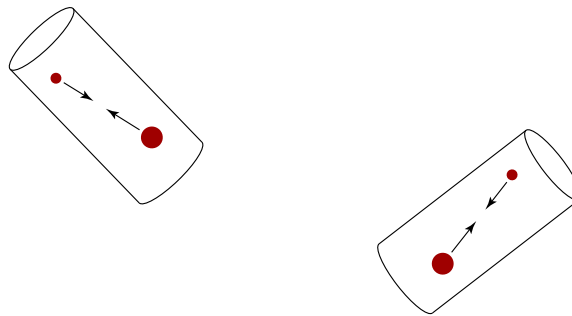
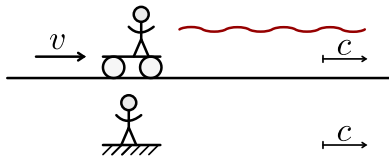


Figure 6.1: Measurement of Newton's law of gravitational attraction (any other law would be just as good) in two frames that are relatively inertial. The equations describing the force between masses will be the same in both frames

- ▷ The **postulate of relativity**: This postulate is best formulated in a negative form: there is no such thing like 'absolute rest' or an 'absolute velocity', or an 'absolute point in space'. No absolute frame exists in which the laws of nature assume a distinguished form.

To formulate this principle in a positive form, we define two coordinate frames to be **inertial** with respect to each other if one is related to the other by a translation in space and/or time, a constant rotation, a uniform motion, or a combination of these operations. The postulate then states that the laws of nature will assume the same form (be expressed by identical fundamental equations) in both systems.

As a corollary, this implies that physical laws must never make reference to absolute coordinates, times, angles, etc.

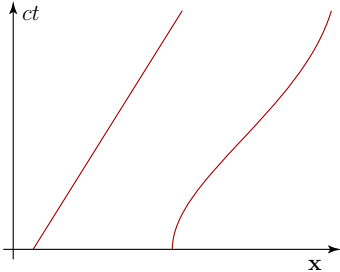


- ▷ The **postulate of the constancy of the speed of light**: the speed of light is independent of the motion of its source.

What makes this postulate — which superficially might seem to be no more than an innocent tribute to experimental observation³ — so revolutionary is that it implies the abandonment of ‘absolute time’. For example, two events that are simultaneous in one inertial frame will in general no longer be simultaneous in other inertial frames. While, more than one hundred years later, we have grown used to its implications the revolutionary character of Einsteins second postulate cannot be exaggerated.

6.1.3 First consequences

As we shall see below, the notion of an ‘absolute time’ — which makes perfect sense in Newtonian mechanics⁴ — becomes meaningless in Einsteins theory of relativity. When we monitor physical processes we must be careful to assign to each physical event its own space *and* time coordinate $x = (ct, \mathbf{x})$ (where the factor of c has been included for later convenience.) An event is canonical in that it can be observed in all frames (no matter whether they are relatively inertial or not). However, both its space and time coordinates are non-canonical, i.e. when observed from a different frame K' it will have coordinates $x' = (ct', \mathbf{x}')$, where t' may be different from t .



When we talk of a ‘body’ or a ‘particle’ in relativity, what we actually mean is the continuous sequence of events $x = (ct, \mathbf{x})$ defined by its instantaneous position \mathbf{x} at time t (both monitored in a specific frame). The assembly of these events obtains the **world line** of the body, a (directed) curve in a space–time coordinate frame (cf. the figure for a schematic of the (1+1)–dimensional world lines of a body in uniform motion (left) and an accelerated body (right).)

The most important characteristic of the relative motion of inertial frames is that no acceleration is involved. This implies, in particular, that a uniform motion in one frame will stay uniform in the other; straight world lines transform into straight world lines. Parameterizing a general uniform motion in system K by $x = (ct, \mathbf{w}t + \mathbf{b})$, where \mathbf{w} is the velocity and \mathbf{b} an offset vector, we must require that the coordinate representation x' in any inertial system K' is again a straight line. Now, the most general family of transformations mapping straight lines onto straight lines are the **affine transformations**

$$x' = \Lambda x + a, \quad (6.1)$$

³... although the constancy of the speed of light had not yet been fully established experimentally when Einstein made his postulate.

⁴Two events that occur simultaneously in one frame will remain simultaneous under Galilei transformations.

where $\Lambda : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1}$ is an arbitrary invertible linear map from $(d+1)$ -dimensional⁵ space time into itself and $a \in \mathbb{R}^{d+1}$ a vector describing the displacement of the origin of coordinate systems in space and or time. For $a = 0$, we speak of a homogeneous coordinate transformation. Since any transformation may be represented as a succession of a homogeneous transformation and a trivial translation, we will focus on the former class throughout.

What conditions will the transformation matrix Λ have to fulfill in order to qualify as a transformation between inertial frames? Referring for a rigorous discussion of this question to Ref. [2], we here merely note that symmetry conditions implied by the principle of relativity nearly but not completely specify the class of permissible transformations. However, they *do* exclude the Galilei transformations featuring as the relevant transformation class in classical mechanics.

To see why, consider **Einstein's second postulate**: Let K and K' be two inertial frames related to each other by a homogeneous coordinate transformation. At time $t = 0$ the origins of the two systems coalesce. Let us assume, however, the K' moves with relatively to K at some constant velocity \mathbf{v} . Now consider the event: at time $t = 0$ a light source at $\mathbf{x} = 0$ emits a signal. In K , the wave fronts of the light signal then trace out world lines propagating at the vacuum speed of light, i.e. the spatial coordinates \mathbf{x} of a wave front obey the relation

$$\mathbf{x}^2 - c^2 t^2 = 0.$$

The crucial point now is that in spite of its motion relative to the point of emanation of the light signal, an observer in K' , too, will observe a light signal moving with velocity c in all directions. (Notice the difference to, say, a sound wave. From K' 's point of view, fronts moving in parallel to \mathbf{v} would be slowed down while those moving in the opposite direction would propagate at higher speed, $\mathbf{v}'_{\text{sound}} = \mathbf{v}_{\text{sound}} - \mathbf{v}$.) This means that the light front coordinates in \mathbf{x}' obey the same relation

$$\mathbf{x}'^2 - c^2 t'^2 = 0.$$

This additional condition unambiguously fixes the set of permissible transformations. To formulate it in the language of linear coordinate transformations (which, in view of the linearity of transformations between inertial frames is the adequate one), let us define the matrix

$$g = \{g_{\mu\nu}\} \equiv \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}, \quad (6.2)$$

where all non-diagonal entries are zero. The constancy of the speed of light may then be expressed in concise form as $x^T g x = 0 \Rightarrow x'^T g x' = 0$. This condition suggests to focus on

⁵Although our primary interest is in $d = 3$, it is occasionally useful to generalize to other dimensions. Specifically, space-time diagrams are easiest to draw/imagine in $(1+1)$ -dimensions. Why do we require Λ to be invertible?

coordinate transformations for which the bilinear form $x^T g x$ is conserved,⁶

$$x^T g x = x'^T g x'. \quad (6.3)$$

Substituting $x' = \Lambda x$, we find that Λ must obey the condition

$$\Lambda^T g \Lambda = g \quad (6.4)$$

Linear coordinate transformations satisfying this condition are called **Lorentz transformations**. We thus postulate that

All laws of nature must be invariant under affine coordinate transformations (6.1) where the homogeneous part of the transformation obeys the Lorentz condition (6.4).

While this statement 'implicitly' characterizes the set of legitimate transformations, our discussion below will require a much more explicit description. The purely mathematical task of obtaining this description will be tackled in the next section.

6.2 The mathematics of special relativity I: Lorentz group

In this section, we will develop the mathematical framework required to describe transformations between inertial frames. We will show that these transformations form a continuous group, the Lorentz group.

6.2.1 Background

The bilinear form introduced above defines a scalar product of \mathbb{R}^4 :⁷

$$g : \mathbb{R}^4 \times \mathbb{R}^4 \rightarrow \mathbb{R}, \quad (6.5)$$

$$(x, y) \mapsto x^T g y \equiv x^\mu g_{\mu\nu} y^\nu. \quad (6.6)$$

A crucial feature of this scalar product is that it is not positive definite. For reasons to be discussed below, we call vectors whose norm is positive, $x^T g x = (ct)^2 - \mathbf{x}^2 > 0$, **time like** vectors. Vectors with negative norm will be called **space like** and vectors of vanishing norm **light like**.

⁶Notice the gap in the logics of the argument. Einsteins second postulate requires Eq. (6.3) only for those space time vectors for which $x^T g x = 0$ vanishes; We now suggest to declare it as a universal condition, i.e. one that holds irrespective of the value of $x^T g x$. To show that the weaker condition implies the stronger, one may first prove (cf. Ref. [2]) that relatively straightforward symmetry considerations suffice to fix the class of allowed transformations up to one undetermined scalar constant. The value of that constant is then fixed by the 'weak condition' above. Once it has been fixed, one finds that Eq. (6.3) holds in general.

⁷We here use covariant notation in which vector components are labeled with superscript indices, x^μ , $\mu = 0, 1, 2, 3$. For further discussion of this notation, see section 6.4.1.

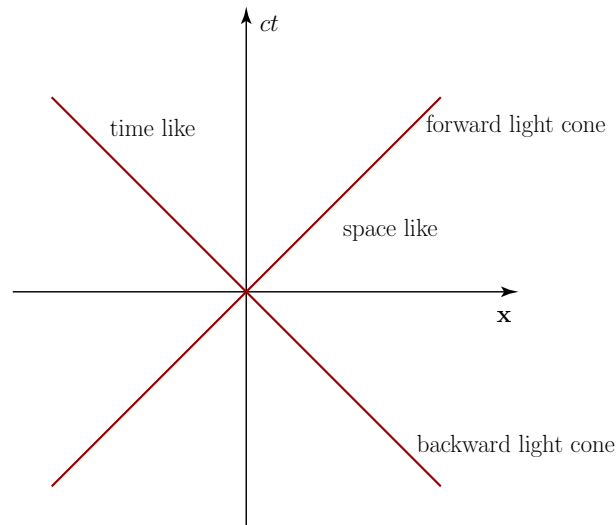


Figure 6.2: Schematic of the decomposition of space time into a time like region and a space like region. The two regions are separated by a 'light cone' of light-like vectors. (A 'cone' because in space dimensions $d > 1$ it acquires a conical structure.) The forward/backward part of the light cone extends to positive/negative times

We are interested in Lorentz transformations, i.e. linear transformations $\Lambda : \mathbb{R}^4 \rightarrow \mathbb{R}^4, x \mapsto \Lambda x$ that leave the scalar product g invariant, $\Lambda^T g \Lambda = g$. (Similarly to, say, the orthogonal transformations which leave the standard scalar product invariant.) Before exploring the structure of these transformations in detail, let us observe a number of general properties. First notice that the set of Lorentz transformations forms a group, the **Lorentz group**, L . For if Λ and Λ' are Lorentz transformations, so is the composition $\Lambda \Lambda'$. The identity transformation obviously obeys the Lorentz condition. Finally, the equation $\Lambda^T g \Lambda = g$ implies that Λ is non-singular (why?), i.e. it possesses an inverse, Λ^{-1} . We have thus shown that the Lorentz transformations form a group.

Global considerations on the Lorentz group

What more can be said about this group? Taking the determinant of the invariance relation, we find that $\det(\Lambda^T g \Lambda) = \det(\Lambda)^2 \det(g) = \det(g)$, i.e. $\det(\Lambda)^2 = 1$ which means that $\det(\Lambda) \in \{-1, 1\}$. (Λ is a real matrix, i.e. $\det(\Lambda) \in \mathbb{R}$.) We may further observe that the absolute value of the component Λ^0_0 is always larger than unity. This is shown by inspection of the 00-element of the invariance relation: $1 = g_{00} = (\Lambda^T g \Lambda)_{00} = (\Lambda^0_0)^2 - \sum_i (\Lambda^0_i)^2$. Since the subtracted term is positive (or zero), we have $|\Lambda^0_0| \geq 1$. We thus conclude that L contains four components defined by

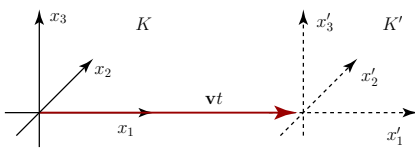
$$\begin{aligned}
 L_+^\uparrow &: \det \Lambda = +1, \quad \Lambda^0_0 \geq +1, \\
 L_+^\downarrow &: \det \Lambda = +1, \quad \Lambda^0_0 \leq -1, \\
 L_-^\uparrow &: \det \Lambda = -1, \quad \Lambda^0_0 > +1, \\
 L_-^\downarrow &: \det \Lambda = -1, \quad \Lambda^0_0 \leq -1.
 \end{aligned} \tag{6.7}$$

Transformations belonging to any one of these subsets cannot be continuously deformed into a transformation of a different subset (why?), i.e. the subsets are truly disjoint. In the following, we introduce a few distinguished transformations belonging to the different components of the Lorentz group.

Space reflection, or parity, $\mathcal{P} : (ct, \mathbf{x}) \rightarrow (ct, -\mathbf{x})$ is a Lorentz transformation. It belongs to the component L_-^\uparrow . **Time inversion**, $\mathcal{T} : (ct, \mathbf{x}) \rightarrow (-ct, \mathbf{x})$ belongs to the component L_-^\downarrow . The product of space reflection and time inversion, $\mathcal{PT} : (ct, \mathbf{x}) \rightarrow (-ct, -\mathbf{x})$ belongs to the component L_+^\downarrow . Generally, the product of two transformations belonging to a given component no longer belongs to that component, i.e. the subsets do not form subgroups of the Lorentz group. However, the component L_+^\uparrow is exceptional in that it is a subgroup. It is called the **proper orthochronous Lorentz group**, or **restricted Lorentz group**. The attribute 'proper' indicates that it does not contain exceptional transformations (such as parity and time inversion) which cannot be continuously deformed back to the identity transform. It is called 'orthochronous' because it maps the positive light cone into itself, i.e. it respects the 'direction' of time. To see that L_+^\uparrow is a group, notice that it contains the identity transformation. Further, its elements can be continuously contracted to unity; however, if this property holds for two elements $\Lambda, \Lambda' \in L_+^\uparrow$, then so for the product $\Lambda\Lambda'$, i.e. the product is again in L_+^\uparrow .

▷ Info. The (Lorentz) group of transformations respecting the Minkowski metric as $\Lambda^T \eta \Lambda = \eta$ is also denoted as $O(1, 3)$. The notation underlines the analogy to the orthogonal group $O(4)$ whose elements satisfy, $O^T \mathbb{1}_4 O = \mathbb{1}_4$, and the fact that the Minkowski metric has *signature* $(1, 3)$, i.e. one positive and three negative eigenvalues. (Discuss why the signature of a metric can't change under basis transformations.) Much as $SO(4)$ is the subgroup of the orthogonal group containing matrices of unit determinant (and hence the unit transformation), the proper Lorentz group is denoted as $SO^+(1, 3)$, where the $+$ -superscript indicates the $\Lambda^0_0 > 1$ condition.

The restricted Lorentz group

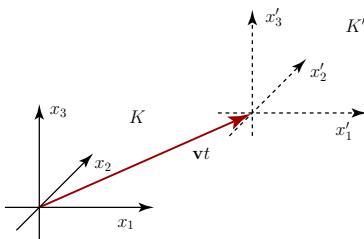


The restricted Lorentz group contains transformations which we associate with coordinate transformations between inertial frames. These transformations fall into two large categories. The first are rotations of space,

$$\Lambda_R \equiv \begin{pmatrix} 1 & \\ & R \end{pmatrix},$$

where R is a three-dimensional rotation matrix. The second we associate with uniform motions between inertial frames.

To represent these in terms of Lorentz transformations, consider two inertial frames K and K' whose origins at time $t = 0$ coalesce (which means that the affine coordinate transformation (6.1)



will be homogeneous, $a = 0$. We assume that K' moves at constant velocity \mathbf{v} relative to K . Lorentz transformations of this type, $\Lambda_{\mathbf{v}}$, are called **special Lorentz transformations**. Physically, they describe what is sometimes called a **Lorentz boost**, i.e. the passage from a stationary to a moving frame. (Although we are talking about a mere coordinate transformation, i.e. no accelerated physical process is described.) We wish to explore the mathematical and physical properties of these transformations.

Without loss of generality, we assume that the vector \mathbf{v} describing the motion is parallel to the e_1 direction of K (and hence to the e'_1 -direction of K' : Any general velocity vector \mathbf{v} may be rotated to $\mathbf{v} \parallel e_1$ by a space like rotation R . This means that a generic special Lorentz transformation $\Lambda_{\mathbf{v}}$ may be obtained from the prototypical one Λ_{ve_1} by a spatial rotation, $\Lambda_{\mathbf{v}} = \Lambda_R \Lambda_{ve_1} \Lambda_R^{-1}$ (cf. the figure.)

A special transformation in e_1 -direction will leave the coordinates $x_2 = x'_2$ and $x_3 = x'_3$ unchanged.⁸ The sought for transformation thus assumes the form

$$\Lambda_{ve_1} = \begin{pmatrix} A_v & & & \\ & 1 & & \\ & & & \\ & & & 1 \end{pmatrix},$$

where the 2×2 matrix A_v operates in (ct, x_1) -space.

There are many ways to determine the matrix A_v . We here use a group theory inspired method which may come across as somewhat abstract but has the advantage of straightforward applicability to many other physical problems. Being element of a continuous group of transformations (a Lie group), the matrix $A_v \equiv \exp(\sum_i \lambda_i T_i)$ can be written in an exponential parameterization. Here, λ_i are real parameters and T_i two-dimensional fixed 'generator matrices' (Cf. with the conventional rotation group; the T_i 's play the role of angular momentum generator matrices, and the λ_i 's are generalized real-valued 'angles'.) To determine the group generators, we consider the transformation A_v for infinitesimal angles λ_i . Substituting the matrix A_v into the defining equation of the Lorentz group, using that the restriction of the metric to (ct, x) space assumes the form $g \rightarrow \sigma_3 \equiv \text{diag}(1, -1)$, and expanding to first order in λ_i , we obtain

$$A_v^T \sigma_3 A_v \simeq (1 + \sum_i \lambda_i T_i^T) \sigma_3 (1 + \sum_i \lambda_i T_i) \simeq \sigma_3 + \sum_i \lambda_i (T_i^T \sigma_3 + \sigma_3 T_i) \stackrel{!}{=} \sigma_3.$$

This relation must hold regardless of the value of λ_i , i.e. the matrices T_i must obey the condition $T_i^T \sigma_3 = -\sigma_3 T_i$. Up to normalization, there is only one matrix that meets this condition, viz $T_i = \sigma_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The most general form of the matrix A_v is thus given by

$$A_v = \exp \left(\lambda \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) = \begin{pmatrix} \cosh \lambda & \sinh \lambda \\ \sinh \lambda & \cosh \lambda \end{pmatrix},$$

⁸To actually prove this, notice that from the point of view of an observer in K' we consider a special transformation with velocity $-ve'_1$. Then apply symmetry arguments (notably the condition of parity invariance; a mirrored universe should not differ in an observable manner in its transformation behavior from our one.) to show that any change in these coordinates would lead to a contradiction.

where we denoted the single remaining parameter by λ and in the second equality expanded the exponential in a power series. We now must determine the parameter $\lambda = \lambda(v)$ in such a way that it actually describes our v -dependent transformation. Substituting A_v into the transformation law $x' = \Lambda_v x$, or

$$\begin{pmatrix} x^{0'} \\ x^{1'} \end{pmatrix} = A_v \begin{pmatrix} x^0 \\ x^1 \end{pmatrix}.$$

Now, the origin of K' (having coordinate $x^{1'} = 0$) is at coordinate $x^1 = vt$. Substituting this condition into the equation above, we obtain the condition

$$\tanh \lambda = -v/c.$$

Using this result and introducing the (standard) notation

$$\beta = |v|/c, \quad \gamma \equiv (1 - \beta^2)^{-1/2}, \quad (6.8)$$

the special Lorentz transformation assumes the form

$$\Lambda_{ve_1} = \begin{pmatrix} \gamma & -\gamma\beta & & \\ -\gamma\beta & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \quad (6.9)$$

▷ EXERCISE. Repeat the argument above for the full transformation group, i.e. not just the set of special Lorentz transformations along a certain axis. To this end, introduce an exponential representation $\Lambda = \exp(\sum_i \lambda_i T_i)$ for general elements of the restricted Lorentz group. Use the defining condition of the group to identify six linearly independent matrix generators.⁹ Among the matrices T_i , identify three as the generators J_i of the spatial rotation group, $T_i = J_i$. Three others generate special Lorentz transformations along the three coordinate axes of K . (Linear combinations of these generators can be used to describe arbitrary special Lorentz transformations.)

A few more remarks on the transformations of the restricted Lorentz group:

- ▷ In the limit of small β , the transformation Λ_{ve_1} asymptotes to the Galilei transformation, $t' = t$ and $x'_1 = x_1 - vt$.
- ▷ We repeat that a general special Lorentz transformation can be obtained from the transformation along e_1 by a space-like rotation, $\Lambda_{\mathbf{v}} = \Lambda_R \Lambda_{ve_1} \Lambda_R^{-1}$, where R is a rotation mapping e_1 onto the unit vector in \mathbf{v} -direction.
- ▷ Without proof (However, the proof is by and large implied by the exercise above.) we mention that a general restricted Lorentz transformation can always be represented as

$$\Lambda = \Lambda_R \Lambda_{\mathbf{v}},$$

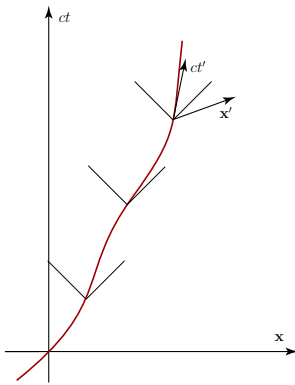
i.e. as the product of a rotation and a special Lorentz transformation.

⁹Matrices are elements of a certain vector space (the space of linear transformations of a given vector space.) As a set of matrices A_1, \dots, A_k is linearly independent if no linear combination exists such that $\sum_i x_i A_i = 0$ vanishes.

6.3 Aspects of relativistic dynamics

In this section we explore a few of the physical consequences deriving from Einsteins postulates. The discussion remains on a superficial level and key topics of the theory of relativity are not mentioned at all. Instead, the prime objective of this section is to provide the background required to discuss the relativistic invariance of electrodynamics below.

6.3.1 Proper time



Consider the world line of a particle moving in space in a (not necessarily uniform manner.) At any instance of time, t , the particle will be at rest in an inertial system K' whose origin coincides with $\mathbf{x}(t)$ and whose instantaneous velocity w.r.t. K is given by $d_t\mathbf{x}(t)$. At this point, it is important to avoid confusion. Of course, the coordinate system whose origin globally coincides with $\mathbf{x}(t)$ is *not* inertial w.r.t. K (unless the motion is uniform.) Put differently, the origin of the inertial system K' defined by the instantaneous position $\mathbf{x}(t)$ and the instantaneous velocity $d_t\mathbf{x}(t)$ coincides with the particles position only at the very instant t . The t' axis of K' is parallel to the vector $d_t\mathbf{x}(t)$. Also, we know that the t' axis cannot be tilted w.r.t. the t axis by more than an angle $\pi/4$, i.e. it must stay within the instantaneous light cone attached to the reference coordinate \mathbf{x} .

Now, imagine an observer traveling along the world line $\mathbf{x}(t)$ and carrying a watch. We assume that at $t = t' = 0$, the observer passes through the origin of the system K (see the figure). We wish to identify the time coordinates of the event 'watch in the moving system K' shows (infinitesimal) time $d\tau'$ '. In K' the answer to this question is formulated easily enough, the event has coordinates $(cd\tau, 0)$, where we noted that, due to the orthogonality of the space-like coordinate axes to the world line, the watch will remain at coordinate $\mathbf{x}' = 0$. In K the event will take place at an as yet undetermined time dt . From K 's perspective, the watch is in motion and at dt will be at the spatial coordinate $dt\mathbf{v}$, where \mathbf{v} is the velocity of K' relative to K at the time of passage and we assumed that this velocity changes only imperceptibly during the small time window dt . Comparing the coordinate representations in K and K' , we conclude that $(cd\tau, 0)$ in K' and $(cdt, \mathbf{v}dt)$ in K describe the same event, and hence must be related by a Lorentz transformation. Specifically, $(cd\tau)^2 = (cdt)^2 - (\mathbf{v}dt)^2$, or

$$d\tau = dt\sqrt{1 - \mathbf{v}^2/c^2}.$$

The time between events of finite duration may be obtained by integration,

$$\tau = \int dt\sqrt{1 - \mathbf{v}^2/c^2}.$$

This equation expresses the famous phenomenon of the relativity of time: For a moving observer time proceeds slower than for an observer at rest. The time measured by the propagating watch, τ , is called **proper time**. The attribute ‘proper’ indicates that τ is defined with reference to a coordinate system (the instantaneous rest frame) that can be canonically defined. I.e. while the time t associated to a given point on the world line $x(t) = (ct, \mathbf{x}(t))$ will change from system to system (as we just saw), the proper time remains the same; the proper time can be used to parameterize the space–time curve in an invariant manner as $x(\tau) = (c\tau, \mathbf{x}(\tau))$.

▷ Info. The relativity of time implies the distortion of other metric quantities, notably the **contraction of length**. An easy way to see this is by the following Gedanken experiment. Suppose you have an object of length L in your own rest frame. How long will this object be in a frame moving with velocity v parallel to the extension L ? From the perspective of an observer in the moving frame, the object appears to be moving with velocity $-v$. One thing the two observers will agree upon is the magnitude v of their relative velocity. From the perspective of the rest frame observer, the moving one will need time $t = L/v$ to progress from one end of the object to the other. However, during that time, only $t' = t/\gamma$ has passed in the moving frame. So, the object appears to have length $L' = vt' < L$ in K' .

6.3.2 Relativistic mechanics

Relativistic energy momentum relation

In section 6.1.2 we argued that Einstein solved the puzzle posed by the Lorentz invariance of electrodynamics by postulating that Newtonian mechanics was in need of generalization. We will begin our discussion of the physical ramifications of Lorentz invariance by developing this extended picture. Again, we start from postulates motivated by physical reasoning:

- ▷ The extended variant of Newtonian mechanics (lets call it relativistic mechanics) ought to be invariant under Lorentz transformations, i.e. the generalization of Newtons equations should assume the same form in all inertial frames.
- ▷ In the rest frame of a particle subject to mechanical forces, or in frames moving at low velocity $v/c \ll 1$ relatively to the rest frame, the equations must asymptote to Newtons equations.

Newton's equations in the rest frame K are of course given by

$$m \frac{d^2}{d\tau^2} \mathbf{x} = \mathbf{f},$$

where m is the particle's mass (in its rest frame — as we shall see, the mass is not an invariant quantity. We thus better speak of a particle **rest mass**), and \mathbf{f} is the force. As always in relativity, a spatial vector (such as \mathbf{x} or \mathbf{f}) must be interpreted as the space–like component

of a four–vector. The zeroth component of x is given by $x^0 = c\tau$, so that the rest frame generalization of the Newton equation reads as $md_\tau^2 x^\mu = f^\mu$, where $f = (0, \mathbf{f})$ and we noted that for the zeroth component $md_\tau^2 x^0 = 0$, so that $f^0 = 0$. It is useful to define the **four–momentum** of the particle by $p^\mu = md_\tau x^\mu$. (Notice that both τ and the rest mass m are Lorentz invariant, i.e. the transformation behaviour of p is determined by that of x .) The zeroth component of the momentum, $md_\tau x^0 = mc$ carries the dimension [energy]/[velocity]. This suggests to define¹⁰

$$p = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix},$$

where E is some characteristic energy. (In the instantaneous rest frame of the particle, $E = mc^2$ and $\mathbf{p} = 0$.)

Expressed in terms of the four–momentum, the Newton equation assumes the simple form

$$d_\tau p = f.$$

Let us explore what happens to the four momentum as we boost the particle from its rest frame to a frame moving with velocity vector $\mathbf{v} = v\mathbf{e}_1$. Qualitatively, one will expect that in the moving frame the particle *does* carry a finite space–like momentum (which will be proportional to the velocity of the boost, v), and some renormalization of its energy. (From the point of view of the moving frame, the particle has acquired kinetic energy.)

Focusing on the 0 and 1 component of the momentum (the others won't change), Eq. (6.9) implies

$$\begin{pmatrix} mc^2/c \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} E'/c \\ p'_1 \end{pmatrix} \equiv \begin{pmatrix} \gamma mc^2/c \\ -(\gamma m)v \end{pmatrix}.$$

This equation may be interpreted in two different ways Substituting the explicit formula $E = mc^2$, we find $E' = (\gamma m)c^2$ and $p'_1 = -(\gamma m)v$. In Newtonian mechanics, we would have expected that a particle at rest in K carries momentum $p'_1 = -mv$ in K' . The difference to the relativistic result is the renormalization of the mass factor: A particle that has mass m in its rest frame appears to carry a velocity dependent mass

$$m(\mathbf{v}) = \gamma m = \frac{m}{\sqrt{1 - (v/c)^2}}$$

in K' . For $v \rightarrow c$, it becomes infinitely heavy and its further acceleration will become progressively more difficult:

Particles of finite rest frame mass cannot move faster than with the speed of light.

The energy in K' is given by $(m\gamma)c^2$, and is again affected by mass renormalization. However, a more useful representation is obtained by noting that $p^T g p$ is a conserved quantity. In

¹⁰Momentum frequently appears with the coordinate vector in the Lorentz invariant combination $x_\mu p^\mu = x^0 p^0 - \mathbf{x} \cdot \mathbf{p}$. Since $x^0 = ct$ contains time with a factor c scaled out, it is natural to define $p^0 = E/c$ with a factor c^{-1} scaled out. Energy, E , and time, t , then appear as partners much as momentum \mathbf{p} and coordinates, \mathbf{x} .

K , $p^T g p = (E/c)^2 = (mc)^2$. Comparing with the bilinear form in K' , we find $(mc)^2 = (E'/c)^2 - \mathbf{p}'^2$ or

$$\boxed{E' = \sqrt{(mc^2)^2 + (pc)^2}}, \quad (6.10)$$

where we used the standard notation $p = |\mathbf{p}'|$ for the modulus of the momentum.¹¹ This **relativistic energy–momentum relation** determines the relation between energy and momentum of a free particle.

Its most important implication is that even a free particle at rest carries energy $E = mc^2$, Einsteins world-famous result. However, this energy is usually not observable (unless it gets released in nuclear processes of mass fragmentation, with all the known consequences.) What we do observe in daily life are the changes in energy as we observe the particle in different frames. This suggests to define the **kinetic energy** of a particle by

$$T \equiv E' - mc^2 = \sqrt{(mc^2)^2 + (pc)^2} - mc^2,$$

For velocities $v \ll c$, the kinetic energy $T \simeq p^2/2m$ indeed reduces to the familiar non-relativistic result. Deviations from this relation become sizeable at velocities comparable to the speed of light.

Particle subject to a Lorentz force

We next turn back to the discussion of the relativistically invariant Newton equation $d_\tau p^\mu = f^\mu$. Both, p^μ and the components of the force, f^μ , transform as vectors under Lorentz transformations. Specifically, we wish to explore this transformation behavior on an example of relevance to electrodynamics, the Newton equation of a particle of velocity \mathbf{v} subject to an electric and a magnetic field. Unlike in much of our previous discussion our observer frame K is not the rest frame of the particle. The four-momentum of the particle is thus given by $p = (m\gamma c, m\gamma \mathbf{v})$. A stationary observer will find that the spatial components of this momentum change according to the Lorentz force, $d_t(m\gamma \mathbf{v}) = q(\mathbf{E} + (\mathbf{v}/c) \times \mathbf{B})$. The differentiation in the time, t , measured by the stationary observer is related via $d_t = \gamma^{-1} d_\tau$ to one in the proper time of the moving particle, so that $d_\tau(m\gamma \mathbf{v}) = \gamma q(\mathbf{E} + (\mathbf{v}/c) \times \mathbf{B})$. Comparison with $d_\tau \mathbf{p} = \mathbf{f}$ shows that the space like components of force acting on the particle are given by, $\mathbf{f} = q\gamma(\mathbf{E} + (\mathbf{v}/c) \times \mathbf{B})$.

To complete our derivation of the Newton equation in K , we need to identify the zeroth component of the force, f_0 . The zeroth component of the left hand side of the Newton equation is given by $c^{-1} d_\tau E = \gamma c^{-1} d_t E$, i.e. $(\gamma/c) \times$ the rate at which the kinetic energy of the particle changes. However, we have seen before that this rate of *potential* energy change is given by $d_t U = -\mathbf{f} \cdot \mathbf{v} = -q\mathbf{E} \cdot \mathbf{v}$. Energy conservation implies that this energy gets converted into the kinetic energy, $d_t T = d_t E = +q\mathbf{E} \cdot \mathbf{v}$. This implies the identification $f^0 = (\gamma q/c)\mathbf{E} \cdot \mathbf{v}$. Finally, using that $p^0 = E/c = m\gamma c$, we obtain the generalized form of Newtons equations

$$m d_\tau \begin{pmatrix} \gamma c \\ \gamma \mathbf{v} \end{pmatrix} = \frac{q}{c} \begin{pmatrix} \mathbf{E} \cdot (\gamma \mathbf{v}) \\ \gamma c \mathbf{E} + (\gamma \mathbf{v}) \times \mathbf{B} \end{pmatrix}.$$

¹¹This is not to be confused with the four-momentum. Also note, that p will be the kinematic momentum of the moving particle observed from the perspective of an observer at rest in the moving frame.

The form of these equations suggests to introduce the **four-velocity** vector

$$v \equiv \begin{pmatrix} \gamma c \\ \gamma \mathbf{v} \end{pmatrix}$$

(in terms of which the four-momentum is given by $p = mv$, i.e. by multiplication by the rest mass.) The equations of motion can then be expressed as

$$m d_{\tau} v^{\mu} = \frac{q}{c} F^{\mu\nu} v_{\nu}, \quad (6.11)$$

where we used the index raising and lowering convention originally introduced in chapter 4, i.e. $v^0 = v_0$ and $v^i = -v_i$, and the matrix F is defined by

$$F = \{F^{\mu\nu}\} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{pmatrix}. \quad (6.12)$$

The significance of Eq. (6.11) goes much beyond that of a mere reformulation of Newton's equations: The electromagnetic field enters the equation through a *matrix*, the so-called **field strength tensor**. This signals that the transformation behavior of the electromagnetic field will be different from that of vectors. Throughout much of the course, we treated the electric field as if it was a (space-like) vector. Naively, one might have expected that in the theory of relativity, this field gets augmented by a fourth component to become a four-vector. However, a moment's thought shows that this picture cannot be correct. Consider, say, a charged particle at rest in K . This particle will create a Coulomb electric field. However, in a frame K' moving relative to K , the charge will be in motion. Thus, an observer in K' will see an electric current plus the corresponding magnetic field. This tells us that under inertial transformations, electric and magnetic fields get transformed into one another. We thus need to find a relativistically invariant object accommodating the six components of the electric and magnetic field 'vectors'. Eq. (6.12) provides a tentative answer to that problem. The idea is that the fields enter the theory as a matrix and, therefore, transform in a way fundamentally different from that of vectors.

6.4 Mathematics of special relativity II: Co- and Contravariance

So far, we have focused on the Lorentz transformation behaviour of vectors in \mathbb{R}^4 . However, our observations made in the end of the previous section suggest to include other objects (such as matrices) into our discussion. We will begin by providing some essential mathematical background and then, finally, turn to the discussion of Lorentz invariance of electrodynamics.

6.4.1 Covariant and Contravariant vectors

Generalities

Let $x = \{x^\mu\}$ be a four component object that transforms under homogeneous Lorentz transformations as $x \rightarrow \Lambda x$ or $x^\mu \rightarrow x'^\mu = \Lambda^\mu_\nu x^\nu$ in components. (By Λ^μ_ν , we denote the components of the Lorentz transformation matrix. For the up/down arrangement of indices, see the discussion below.) Quantities of this transformation behaviour are called **contravariant vectors** (or contravariant tensors of first rank). In this notation, the Lorentz invariance condition assumes the form $x'^\mu g_{\mu\nu} y'^\nu = (\Lambda^\mu_\rho x^\rho) g_{\mu\nu} (\Lambda^\nu_\sigma y^\sigma) = x^\rho g_{\rho\sigma} y^\sigma$, or $\Lambda^\mu_\rho g_{\mu\nu} \Lambda^\nu_\sigma = g_{\rho\sigma}$. Multiplication by the matrix Λ^{-1} brings this into the form¹²

$$g_{\mu\nu} \Lambda^\nu_\sigma = (\Lambda^{-1})^\tau_\mu g_{\tau\sigma}. \quad (6.13)$$

Now, let us define another four-component object, $x_\mu \equiv g_{\mu\nu} x^\nu$. Under a Lorentz transformation, $x_\mu \rightarrow g_{\mu\nu} \Lambda^\nu_\sigma x^\sigma = (\Lambda^{-1})^\tau_\mu g_{\tau\sigma} x^\sigma = x_\tau (\Lambda^{-1})^\tau_\mu$, where we used Eq. (6.13). Quantities of this transformation behavior will be denoted as **covariant vectors** (or covariant vectors of first rank). We now upgrade co- and contravariance to a general principle by postulating

$$\begin{aligned} \text{Contravariant index:} \quad & X^{\dots\mu} \xrightarrow{\Lambda} \Lambda^\mu_\nu X^{\dots\nu}, \\ \text{Covariant index:} \quad & X_{\dots\mu} \xrightarrow{\Lambda} X_{\dots\nu} (\Lambda^{-1})^\nu_\mu, \end{aligned} \quad (6.14)$$

where the ...'s refer to arbitrary other indices. In this way, co- and contravariance are *defined* via their behavior under transformations. Notice that the placement of the matrix elements Λ^μ_ν and $(\Lambda^{-1})^\nu_\mu$ (numbers) to the left or right of the transformed objects is irrelevant. The summation over a contravariant and a covariant index is called a **contraction**. For example, $v^\mu w_\mu \equiv \phi$ is the contraction of a contravariant and a covariant vector. It yields a Lorentz scalar, i.e. an object that does not change under Lorentz transformations. More generally, the number of co- and contravariant indices carried by an object defines its degree as a tensor, ($R^\mu_{\nu\rho\sigma}$ is contravariant of first and covariant of third degree), and the contraction of two indices lowers the index, or 'tensor' degree of an expression by two. For example the contraction of a mixed tensor of degree two and a contravariant tensor of degree one, $A^\mu_\nu v^\nu \equiv w^\mu$ yields a contravariant tensor of degree one, etc. This is because the definitions above are designed so as to leave contractions $X^{\dots\mu} Y_{\dots\mu} \rightarrow X^{\dots\nu} (\Lambda^{-1})^\nu_\mu \Lambda^\mu_\rho Y^{\dots\rho} = X^{\dots\mu} Y_{\dots\mu}$ invariant under transformations.

For example, we define $g^{\mu\nu}$ to be the inverse metric. This object is naturally contravariant. In this way, the product $g^{\mu\nu} g_{\nu\rho} = \delta^\mu_\rho$ contains an invariant contraction and yields the Kronecker- δ . Given contravariant $X^{\dots\mu}$, a covariant index may be produced as $X_{\dots\mu} \equiv g_{\mu\nu} X^{\dots\nu}$, conversely, a contravariant $X^{\dots\mu}$ is obtained as $X^{\dots\mu} = g^{\mu\nu} X_{\dots\nu}$. These operations are called the lowering and raising of indices, respectively.

¹²Although this looks more clumsy than the component free relation $\Lambda^T g \Lambda = g \Rightarrow g \Lambda = (\Lambda^T)^{-1} g$, working with the matrix transpose, T , often leads to confusion in the context of relativity. It is generally safer to use explicit index contractions as in Eq. (6.13).

Critical readers may find these 'definitions' unsatisfactory. Formulated in a given basis, one may actually wonder whether they are definitions at all. For a discussion of the basis invariant meaning behind the notions of co- and contravariance, we refer to the info block below. However, we here proceed in a pragmatic way and continue to explore the consequences of the definitions above.

▷ **Info.** Consider a general real vector space V . Recall that the dual space V^* is the linear space of all mappings $\tilde{\mathbf{v}} : V \rightarrow \mathbb{R}, \mathbf{v} \mapsto \tilde{\mathbf{v}}(\mathbf{v})$. (This space is a vector space by itself which is why we denote its elements by vector-like symbols, $\tilde{\mathbf{v}}$.) For a given basis $\{\mathbf{e}_\mu\}$ of V , a dual basis $\{\tilde{\mathbf{e}}_\mu\}$ of V^* may be defined by the condition $\tilde{\mathbf{e}}_\mu(\mathbf{e}_\nu) = \delta_{\mu\nu}$. With the expansions $\mathbf{v} = \sum_\mu v^\mu \mathbf{e}_\mu$ and $\tilde{\mathbf{v}} = \sum_\mu \tilde{v}_\mu \tilde{\mathbf{e}}_\mu$, respectively, we have $\tilde{\mathbf{v}}(\mathbf{v}) = \tilde{v}_\mu v^\mu$. (Notice that components of objects in dual space will be indexed by superscripts throughout.) In the literature of relativity, elements of the vector space — then to be identified with space-time, see below — are usually called **contravariant vectors** while their partners in dual space are called **covariant vectors**.

▷ **EXERCISE.** Let $A : V \rightarrow V$ be a linear map defining a basis change in V , i.e. $\mathbf{e}_\mu = A_\mu^\nu \mathbf{e}'_\nu$, where $\{A_\mu^\nu\}$ are the matrix elements of A . Show that

- ▷ The components of a contravariant vector \mathbf{v} transform as $v^\mu \rightarrow v'^\mu = (A^T)^\mu_\nu v^\nu$.
- ▷ The components of a covariant vector, $\tilde{\mathbf{w}}$ transform as $\tilde{w}_\mu \rightarrow \tilde{w}'_\mu = (A^{-1})^\nu_\mu \tilde{w}_\nu$.
- ▷ The action of $\tilde{\mathbf{w}}$ on \mathbf{v} remains invariant, i.e. $\tilde{\mathbf{w}}(\mathbf{v}) = \tilde{w}_\mu v^\mu = \tilde{w}'_\mu v'^\mu$ does not change. (Of course, the action of the linear map $\tilde{\mathbf{w}}$ on vectors must not depend on the choice of a particular basis.)

In physics it is widespread (if somewhat tautological) practice to *define* co- or contravariant vectors by their transformation behaviour. For example, a set of d -components $\{w_\mu\}$ is denoted a covariant vector if these components change under linear transformations as $w_\mu \rightarrow A_\mu^\nu w_\nu$.

Now, let us assume that V is a vector space with scalar product $g : V \times V \rightarrow \mathbb{R}, (\mathbf{v}, \mathbf{w}) \mapsto v^T g w \equiv \langle v, w \rangle$. A special feature of vector spaces with scalar product is the existence of a *canonical* mapping $V \rightarrow V^*, \mathbf{v} \mapsto \tilde{\mathbf{v}}$, i.e. a mapping that to all vectors \mathbf{v} canonically (without reference to a specific basis) assigns a dual element $\tilde{\mathbf{v}}$. The vector $\tilde{\mathbf{v}}$ is implicitly defined by the condition $\forall \mathbf{w} \in V, \tilde{\mathbf{v}}(\mathbf{w}) \stackrel{!}{=} \langle \mathbf{v}, \mathbf{w} \rangle$. For a given basis $\{\mathbf{e}_\mu\}$ of V the components of $\tilde{\mathbf{v}} = \sum_\mu \tilde{v}_\mu \tilde{\mathbf{e}}_\mu$ may be obtained as $\tilde{v}_\mu = \tilde{\mathbf{v}}(\mathbf{e}_\mu) = \langle \mathbf{v}, \mathbf{e}_\mu \rangle = v^\nu g_{\nu\mu} = g_{\mu\nu} v^\nu$, where in the last step we used the symmetry of g . With the so-called **index lowering convention**

$$v_\mu = g_{\mu\nu} v^\nu, \quad (6.15)$$

we are led to the identification $\tilde{v}_\mu = v_\mu$ and $\tilde{\mathbf{v}}(\mathbf{w}) = v_\mu w^\mu$.

Consider a map $A : V \rightarrow V, \mathbf{v} \mapsto A\mathbf{v}$. In general the action of the canonical covariant vector $\tilde{A}\mathbf{v}$ on transformed vectors $A\mathbf{w}$ need not equal the original value $\tilde{\mathbf{v}}(\mathbf{w})$. However, it is natural to focus on those transformations that are compatible with the canonical assignment, i.e. $\tilde{A}\mathbf{v}(A\mathbf{w}) = \tilde{\mathbf{v}}(\mathbf{w})$. In a component language, this condition translates to

$$(A^T)_\mu^\sigma g_{\sigma\rho} A^\rho_\nu = g_{\mu\nu}$$

or $A^T g A = g$. I.e. transformations compatible with the canonical identification (vector space) \leftrightarrow (dual space.) have to respect the metric. E.g., in the case of the Minkowski metric, the 'good' transformations belong to the Lorentz group.

Summarizing, we have seen that the invariant meaning behind contra- and covariant vectors is that of vectors and dual vectors respectively. Under Lorentz transformations these objects behave as is required by the component definitions given in the main text. A general tensor of degree (n, m) is an element of $(\otimes_1^n V) \otimes (\otimes_1^m V^*)$.

The number of indices carried by a certain object defines its *degree* as a tensor. For example, a two component quantity $W^{\mu\nu}$ is called a contravariant **tensor of second degree** and it transforms under Lorentz transformations as $W^{\mu\nu} \rightarrow \Lambda^\mu_{\mu'} \Lambda^\nu_{\nu'} W^{\mu'\nu'}$. Similarly, a covariant tensor of second degree transforms as $W_{\mu\nu} \rightarrow (\Lambda^{-1})^{\mu'}_{\mu} (\Lambda^{-1})^{\nu'}_{\nu} W_{\mu'\nu'}$. Covariant and contravariant tensors are related to each other by index raising/lowering, e.g. $W_{\mu\nu} = g_{\mu\mu'} g_{\nu\nu'} W^{\mu'\nu'}$. A mixed second rank tensor $W_\mu{}^\nu$ transforms as $W_\mu{}^\nu \rightarrow (\Lambda^{-1})^{\mu'}_{\mu} \Lambda^\nu_{\nu'} W_{\mu'}{}^{\nu'}$. The generalization of these definitions to tensors of higher degree should be obvious.

Finally, we define a contravariant **vector field** (as opposed to a fixed vector) as a field $v^\mu(x)$ that transforms under Lorentz transformations as $v^\mu(x) \rightarrow v'^\mu(x') = \Lambda^\mu_{\nu} v^\nu(x)$. Here, $x = x(x')$, i.e. on the r.h.s. the 'old' coordinate vector x is to be expressed as $x = \Lambda^{-1} x'$ as a function of the 'new' coordinate vector x' . A **Lorentz scalar** (field), or tensor of degree zero is one that does not actively transform under Lorentz transformations: $\phi(x) \rightarrow \phi'(x') = \phi(x)$. Covariant vector fields, tensor fields, etc. are defined in an analogous manner.

6.5 Relativistic covariance of electrodynamics

We now have everything in store to prove the relativistic covariance of electrodynamics, i.e. the form-invariance of its basic equations under Lorentz transformations. Basically, what we need to do is assign a definite transformation behavior scalar, co-/contravariant tensor, etc. to the building blocks of electrodynamics.

6.5.1 Covariant formulation of fields and currents

The coordinate vector x^μ is a contravariant vector. In electrodynamics we frequently differentiate w.r.t. the components x^μ , i.e. the next question to ask is whether the four-dimensional gradient $\{\partial_{x^\mu}\}$ is co- or contravariant. According to the chain rule,

$$\frac{\partial}{\partial x'^{\mu'}} = \frac{\partial x^\mu}{\partial x'^{\mu'}} \frac{\partial}{\partial x^\mu}.$$

Using that $x^\mu = (\Lambda^{-1})^\mu_{\mu'} x'^{\mu'}$, we find that

$$\frac{\partial}{\partial x'^{\mu'}} = (\Lambda^{-1})^\mu_{\mu'} \frac{\partial}{\partial x^\mu},$$

i.e. $\partial_{x^\mu} \equiv \partial_\mu$ transforms as a covariant vector. In components,

$$\partial_\mu = (c^{-1}\partial_t, \nabla), \quad \partial^\mu = (c^{-1}\partial_t, -\nabla).$$

The four-current vector

We begin by showing that the four-component object $j = (c\rho, \mathbf{j})$ is a Lorentz vector. Consider the current density carried by a point particle at coordinate $\mathbf{x}(t)$ in some coordinate system K . The four-current density carried by the particle is given by $j(x) = q(c, d_t\mathbf{x}(t))\delta(\mathbf{x} - \mathbf{x}(t))$, or $j^\mu(x) = qd_t x^\mu(t)\delta(\mathbf{x} - \mathbf{x}(t))$, where $x^\mu(t) = (ct, \mathbf{x}(t))$. To show that this is a contravariant vector, we introduce a dummy integration,

$$j^\mu(x) = q \int d\tilde{t} \frac{dx^\mu(\tilde{t})}{d\tilde{t}} \delta(\mathbf{x} - \mathbf{x}(\tilde{t}))\delta(t - \tilde{t}) = qc \int d\tau \frac{dx^\mu}{d\tau} \delta^4(x - x(\tau)),$$

where in the last step we switched to an integration over the proper time $\tau = \tau(\tilde{t})$ uniquely assigned to the world line parameterization $x(\tilde{t}) = (c\tilde{t}, \mathbf{x}(\tilde{t}))$ in K . Now, the four-component δ -distribution, $\delta^4(x) = \delta(\mathbf{x})\delta(ct)$ is a Lorentz scalar (why?). The proper time, τ , also is a Lorentz scalar. Thus, the transformation behaviour of j^μ is dictated by that of x^μ , i.e. j^μ defines a contravariant vector.

As an important corollary we note that the **continuity equation** — the contraction of the covariant vector ∂_μ and the contravariant vector j^μ is Lorentz invariant, $\partial_\mu j^\mu = 0$ is a Lorentz scalar.

Electric field strength tensor and vector potential

In section 6.3.2 we obtained Eq. (6.11) for the Lorentz invariant generalization of Newton's equations. Since v_μ and v^ν transform a covariant and contravariant vectors, respectively, the matrix $F^{\mu\nu}$ must transform as a contravariant tensor of rank two. Now, let us define the four **vector potential** as $\{A^\mu\} = (\phi, \mathbf{A})$. It is then a straightforward exercise to verify that the field strength tensor is obtained from the vector potential as

$$\boxed{F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu.} \quad (6.16)$$

(Just work out the antisymmetric combination of derivatives and compare to the definitions $\mathbf{B} = \nabla \times \mathbf{A}$, $\mathbf{E} = -\nabla\phi - c^{-1}\partial_t\mathbf{A}$.) This implies that the four-component object A^μ indeed transforms as a contravariant vector.

Now, we have seen in chapter 4.2 that the **Lorentz condition** can be expressed as $\partial_\mu A^\mu = 0$, i.e. in a Lorentz invariant form. In the Lorenz gauge, the inhomogeneous wave equations assume the form

$$\square A^\mu = \frac{4\pi}{c} j^\mu, \quad (6.17)$$

where $\square = \partial_\mu \partial^\mu$ is the Lorentz invariant wave operator. Formally, this completes the proof of the Lorentz invariance of the theory. The combination Lorenz condition/wave equations, which we saw carries the same information as the Maxwell equations, has been proven to be invariant. However, to stay in closer contact to the original formulation of the theory, we next express Maxwells equations themselves in a manifestly invariant form.

6.5.2 Invariant formulation of Maxwells equations

One may verify by direct inspection that the two inhomogeneous Maxwell equations can be formulated in a covariant manner as

$$\partial_\mu F^{\mu\nu} = \frac{4\pi}{c} j^\nu. \quad (6.18)$$

To obtain the covariant formulation of the homogeneous equations a bit of preparatory work is required: Let us define the fourth rank antisymmetric tensor as

$$\epsilon^{\mu\nu\rho\sigma} \equiv \begin{cases} 1, & (\mu, \nu, \rho, \sigma) = (0, 1, 2, 3) \text{ or an even permutation} \\ -1 & \text{for an odd permutation} \\ 0 & \text{else.} \end{cases} \quad (6.19)$$

One may show (do it!) that $\epsilon^{\mu\nu\rho\sigma}$ transforms as a contravariant tensor of rank four (under the transformations of the unit-determinant subgroup L_+ .) Contracting this object with the covariant field strength tensor $F_{\mu\nu}$, we obtain a contravariant tensor of rank two,

$$\mathcal{F}^{\mu\nu} \equiv \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma},$$

known as the **dual field strength tensor**. Using (6.20), it is straightforward to verify that

$$\mathcal{F} = \{\mathcal{F}^{\mu\nu}\} = \begin{pmatrix} 0 & -B_1 & -B_2 & -B_3 \\ B_1 & 0 & E_3 & -E_2 \\ B_2 & -E_3 & 0 & E_1 \\ B_3 & E_2 & -E_1 & 0 \end{pmatrix}, \quad (6.20)$$

i.e. that \mathcal{F} is obtained from F by replacement $\mathbf{E} \rightarrow \mathbf{B}$, $\mathbf{B} \rightarrow -\mathbf{E}$. One also verifies that the homogeneous equations assume the covariant form

$$\partial_\mu \mathcal{F}^{\mu\nu} = 0. \quad (6.21)$$

This completes the invariance proof. Critical readers may find the definition of the dual tensor somewhat un-motivated. Also, the excessive use of indices — something one normally tries to avoid in physics — does not help to make the structure of the theory transparent. Indeed, there exists a much more satisfactory, coordinate independent formulation of the theory in terms of differential forms. However, as we do not assume familiarity of the reader with the theory of forms, all we can do is encourage him/her to learn it and to study the truly invariant formulation of electrodynamics not discussed in this text.

6.5.3 Electrodynamics from a variational principle

In classical mechanics, we learn that its equations of motion afford an interpretation as variational equations whose solution extremizes a certain functional — the Lagrangian or Hamiltonian action. Maxwell's equations likewise are 'equations of motion', namely those describing

the infinitely-many-degrees-of-freedom system 'electromagnetic field'. Are they, too, variational equations of a functional? And if yes, how can we find it?

It turns out that the answer to the first question is affirmative, and that to the second surprising. Up to constants which can be absorbed in changes of units, the 'action' of electromagnetism is fixed entirely by criteria of mathematical consistency. In other words, Maxwell's equations follow from mathematics, and not just need mathematics to write them down.

In classical mechanics, we start our discussion of actions with a choices of generalized coordinates, in terms of which we define a Lagrangian function. Turning to electromagnetism, we have seen that a similar role is played by the components of the vector potential, A_μ . At each point in space-time, x , they define four degrees of freedom $A_\mu(x)$, from which the fields themselves may be obtained by differentiation. The A_μ -coordinates are designed so that the fields satisfy the constraint imposed by the homogeneous equations. So, all is set, and the action we need is one for A_μ and its derivatives $\partial_\nu A_\mu$.

We also know that the equations of motion are linear in A_μ , and they are of second order in derivatives. This implies (think about it. Compare to the situation with the harmonic oscillator in classical mechanics), that our action must be quadratic in A , and contain two derivatives acting on these fields. Turning to the r.h.s. of the equations of motion, they are of zeroth order in A_μ and of first order in the currents j^μ . Accordingly, the action must contain a piece of first order in both A and j . Finally, the action must have the same invariance properties at the theory itself, i.e. it must be Lorentz invariant, and gauge invariant.

We now reason, that there is a unique local functional satisfying these properties (up to constants). The unique¹³ Lorentz invariant term of second order in A and derivatives satisfying these criteria is given by

$$F_{\mu\nu}F^{\mu\nu} = (\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu). \quad (6.22)$$

Similarly, the unique invariant term linear in both A and j is $A_\mu j^\mu$. We thus propose the action

$$S[A] = \int dx \mathcal{L} = - \int dx \left(\frac{1}{16\pi} F_{\mu\nu}F^{\mu\nu} + \frac{1}{c} A_\mu j^\mu \right) \quad (6.23)$$

where we use the shorthand notation $\int dx \equiv \int d^4x$ and we have anticipated a choice of constants that will lead to the correct equations of motion. Here, the **Lagrangian density** of the theory is defined by the second equation. It is a 'density' in the sense that integration over three-dimensional space produces a time dependent function $L \equiv \int d^3x \mathcal{L}$ which is analogous to a classical Lagrangian function. (In practice, we never work with L , the density itself is a more useful object to consider.)

To check that this action does the job, let's compute its variation under infinitesimal changes of the field, $A_\mu \rightarrow A_\mu + \epsilon_\mu$. A short calculation using the antisymmetry $F^{\mu\nu} = -F^{\nu\mu}$. Shows that, to first order $F_{\mu\nu}F^{\mu\nu} \rightarrow F_{\mu\nu}F^{\mu\nu} + 4\partial_\mu \epsilon_\nu F^{\mu\nu}$, and $A_\nu j^\nu \rightarrow A_\nu j^\nu + \epsilon_\nu j^\nu$. So we

¹³Actually, this statement is not quite correct, the term $\epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}F_{\rho\sigma}$ likewise satisfies these criteria. However, its variation in A_μ yields only a total derivative which we ignore.

obtain,

$$S[A + \epsilon] = S[A] + \frac{1}{4\pi} \int dx \left(\partial_\mu \epsilon_\nu F^{\mu\nu} + \frac{4\pi}{c} \epsilon_\nu j^\nu \right) = S[A] - \frac{1}{4\pi} \int dx \epsilon_\nu \left(\partial_\mu F^{\mu\nu} - \frac{4\pi}{c} j^\nu \right).$$

The variation vanishes if $\partial_\mu F^{\mu\nu} - \frac{4\pi}{c} j^\nu = 0$, i.e. if the Maxwell equations hold. In this way, we have established Eq. (6.23) as the **action of the electromagnetic field**.

More precisely, Eq. (6.23) defines the action in the Lagrangian formulation of the theory: a functional containing second order time derivatives of the ‘generalized coordinates $A^\mu(x) = A^\mu(\mathbf{x}, t)$ ’. Is there a matching Hamiltonian formulation? To find out, we try to proceed as in classical mechanics, and begin with the identification of a generalized momentum, defined as the derivative of the Lagrangian in the time-derivative $\partial_0 A^\mu$ of the generalized coordinates. Noting that $\partial_0 A^\mu = \partial^0 A^\mu$, we define

$$\pi_\mu = \frac{\partial \mathcal{L}}{\partial(\partial^0 A^\mu)}, \quad \mathcal{L} = -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu} - \frac{1}{c} A_\mu j^\mu,$$

From Eq. (6.22) we then obtain $\pi_\mu = \frac{1}{4\pi}(\partial_\mu A_0 - \partial_0 A_\mu)$, or

$$\pi_0 = 0, \quad \pi_i = \frac{1}{4\pi} \left(\frac{1}{c} \partial_t A_i + \partial_i A_0 \right) = -\frac{1}{4\pi} E^i.$$

Surprisingly, the canonical momentum associated to the spatial components of the vector potential coordinates is the electric field!

Let’s take this observation seriously, and pass from the Lagrangian to the Hamiltonian formulation of the theory. More precisely, we wish to consider the Hamiltonian of the pure electromagnetic field, i.e. we consider $j^\mu = 0$. Judging from classical mechanics, we associate ‘energy’ to the concept of Hamiltonians, i.e. we expect to find an energy-like expression. Continuing with our analogy building to mechanics, we start with a Legendre transform, i.e. we define the **Hamiltonian density**

$$\mathcal{H} = \pi_\mu \partial^0 A^\mu - \mathcal{L},$$

To compute this in more explicit terms, note that

$$F_{\mu\nu} F^{\mu\nu} = 2(\partial_0 A_i - \partial_i A_0)(\partial^0 A^i - \partial^i A^0) + (\partial_i A_j - \partial_j A_i)(\partial^i A^j - \partial^j A^i) = -2\mathbf{E} \cdot \mathbf{E} + 2\mathbf{B} \cdot \mathbf{B},$$

and $\pi_i \partial^0 A^i = -\frac{1}{4\pi} \mathbf{E} \cdot c^{-1} \partial_t \mathbf{A}$. With this, we obtain

$$\mathcal{H} = \frac{1}{8\pi} \left(-2\mathbf{E} \cdot \frac{1}{c} \partial_t \mathbf{A} - \mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B} \right).$$

This expression looks unfinished. However, using that we are considering a source free situation, $\nabla \cdot \mathbf{E} = 0$, in infinite space, we may smuggle in the surface term $-\frac{1}{4\pi} \nabla \cdot (\mathbf{E}\phi) = -\frac{1}{4\pi} \mathbf{E} \cdot \nabla \phi$, where ϕ is the scalar potential. Since the Hamiltonian density exclusively appears

under an integral, the modification $\mathcal{H} \rightarrow \mathcal{H} - \frac{1}{4\pi} \nabla \cdot (\mathbf{E}\phi)$ is physically inconsequential, and we obtain

$$\mathcal{H} \rightarrow \frac{1}{8\pi} \left(2\mathbf{E} \cdot \underbrace{(-c^{-1}\partial_t\mathbf{A} - \nabla\phi)}_{\mathbf{E}} - \mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B} \right) = \frac{1}{8\pi} (\mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B}).$$

This confirms the expectation that the Hamiltonian density of the electromagnetic field equals its energy density, as obtained earlier in a first-hand construction.

6.5.4 Towards Quantum Electrodynamics

Classical electrodynamics is intrinsically consistent, but limited in scope. For example, the energy density \mathcal{H} of the electric generated by a point charge diverges in the charge center, indicating a lack of regularization at short distance scales. By analogy with classical mechanics, we indeed suspect that quantum effects should begin to play a role at sufficiently small scales.

To see how this may happen, further the analogies between electrodynamics and classical mechanics, a discipline where we do have some understanding of how quantum generalizations. We begin by considering the field action (6.23) in the absence of matter, $j = 0$. We impose the *Coulomb gauge*, $\nabla \cdot \mathbf{A} = 0$ and $\phi = 0$.¹⁴ It is straightforward to show that in this gauge the field action assumes the form (do it!)

$$\begin{aligned} S[A] &= \frac{1}{8\pi} \int dt \int d^3x (c^{-2}\partial_t\mathbf{A} \cdot \partial_t\mathbf{A} - (\nabla \times \mathbf{A}) \cdot (\nabla \times \mathbf{A})) = \\ &= \frac{1}{8\pi} \int dt \int d^3x (c^{-2}\partial_t\mathbf{A} \cdot \partial_t\mathbf{A} + \mathbf{A} \cdot \Delta\mathbf{A}), \end{aligned} \quad (6.24)$$

where in the second equality we integrated by parts ($\nabla \times$ does not catch a minus sign under integration by parts) and used that $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \Delta\mathbf{A} = -\Delta\mathbf{A}$.

The action (6.24) will be our gateway to understanding the emergence of quantum effects in the theory. To make this discussion more concrete, we consider a rectangular wave guide as physical reference system. An **electromagnetic waveguide** is a quasi-one-dimensional cavity with metallic boundaries (see fig. 6.3). The practical use of waveguides is that they are good at confining EM waves. At large frequencies, where the wavelengths are of order meters or less, radiation loss in conventional conductors is high. In this frequency domain, hollow conductors provide the only practical way of transmitting radiation.

Suppose we wanted to find nontrivial solutions of an electromagnetic field propagating along the longitudinal direction of the wave guide. Remembering our earlier discussion of vacuum electromagnetic waves we might anticipate the existence of plane-wave like solutions with modifications due to the presence of confining metallic boundaries. To find them we

¹⁴In our present discussion, relativistic invariance will not be an issue, so we don't lose anything by imposing a non-relativistically invariant gauge. However, keep in mind that once it has been fixed, we cannot expect our theory to be fully gauge invariant anymore.

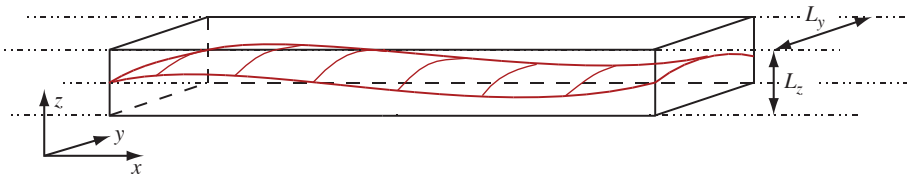


Figure 6.3: A wave guide with rectangular cross-section $L_y \times L_z$ and longitudinal extension $L_x \gg L_{y,z}$. Discussion, see text.

would solve Maxwell's equations, subject to the appropriate boundary conditions. Conceptually, this solution of 'effective equations of motion' is similar to the solution of Newton's equations in classical mechanics. Presently, we will take a somewhat different approach, which is conceptually akin to the Lagrangian approach. We take the action principle seriously and reason that *any* extremal configuration of the action functional represents a physical solution of the problem. Thinking of \mathbf{A} as 'effective coordinates' we turn the presence of boundary conditions to an advantage and first restrict the set of arguments fields $\mathbf{A}(\mathbf{x}, t)$ (we are not talking about physical solution fields yet) to those respecting all boundary conditions. These will define our good coordinates. We then seek for extremal configurations of \mathbf{A} , thereby identifying the physical fields \mathbf{E} and \mathbf{B} . Technically, this is just an alternative, and arguably more efficient approach to a direct solution of Maxwell's equations. However, as a crucially important byproduct, it will contain the decisive hint as to where quantum mechanics should enter the stage.

Now embarking on the first part of the program, think of the components of $A(\mathbf{x}, t)$ as time dependent functions on the support set $V = [0, L_x] \times [0, L_y] \times [0, L_z]$ of the guide. (We will eventually consider the limit $L_x \rightarrow \infty$, however, not just yet.) These functions play a role similar to the curves $q(t)$ entering as arguments in action functionals of classical mechanics. Similarly to the presence of boundary conditions in mechanics, here, too, we need to consider a number of constraints. First, the assumption that our guide is made of perfect conductors means $\mathbf{E}_{\parallel} = 0$ inside the metallic plates confining it: a field component parallel to the guide boundaries would lead to an infinite current, which is unphysical. Second, $\mathbf{B}_{\perp} = 0$ at the boundaries: by assumption we have $\mathbf{B} = 0$ immediately outside the box, so a finite \mathbf{B}_{\perp} at the boundary would be in violation of $\nabla \cdot \mathbf{B} = 0$ (discuss this point.) All \mathbf{A} we consider, must be such that the fields \mathbf{E} and \mathbf{B} obtained from them by differentiation respect these conditions. We wish to describe these functions in maximally economical ways.

At this point, we should turn our attention to the Laplace operator Δ featuring in the action. Since this operator acts on our \mathbf{A} 's, it is tempting to expand the latter in a basis of eigenfunctions of Δ .¹⁵ Denoting these functions by $\mathbf{A}_{\mathbf{k}}(\mathbf{x})$, and their eigenvalues by $\lambda_{\mathbf{k}}$, i.e.

¹⁵Carefully think about this point. You have seen such situations before. Laplace operators are hermitian operators and hence possess 'complete sets of eigenfunctions'. It is generally a good idea to use them as basis functions. The simplest example of these are the exponential functions $\exp(ikx)$, interpreted as eigenfunctions of the 'Laplace operator' d_x^2 and defining the Fourier basis.

$\Delta A_{\mathbf{k}}^i = \lambda_{\mathbf{k}} A_{\mathbf{k}}^i$, for each of the three vector component functions, we thus consider the ansatz

$$\mathbf{A}(\mathbf{x}, t) = \sum_{\mathbf{k}} \alpha_{\mathbf{k}}(t) \mathbf{A}_{\mathbf{k}}(\mathbf{x}),$$

where the precise identity of the vectors \mathbf{k} is yet to be determined. It turns out that the Laplace eigenfunctions relevant to the problem at hand are the configurations

$$\mathbf{A}_{\mathbf{k}}(\mathbf{x}) = \begin{pmatrix} c_1 \cos(k_x x) \sin(k_y y) \sin(k_z z) \\ c_2 \sin(k_x x) \cos(k_y y) \sin(k_z z) \\ c_3 \sin(k_x x) \sin(k_y y) \cos(k_z z) \end{pmatrix},$$

where $k_i = n_i \pi / L_i$ with positive integer n_i , and the constraint $c_1 k_x + c_2 k_y + c_3 k_z = 0$ on the three real coefficients $c_{1,2,3}$ secures the Coulomb gauge condition $\nabla \cdot \mathbf{A} = 0$. It is straightforwardly verified (do it) that these functions satisfy the eigenfunction property with eigenvalues $\lambda_{\mathbf{k}}^2 = k^2 \equiv k_x^2 + k_y^2 + k_z^2$. Finally, we choose the coefficients c_i such that $\int_{x,y,z} \mathbf{A}_{\mathbf{k}} \cdot \mathbf{A}_{\mathbf{k}'} = \delta_{\mathbf{k},\mathbf{k}'}$ holds. (Fill in the details.)

It is straightforward to check verify the functions $\mathbf{A}_{\mathbf{k}}$ are eigenfunctions of the Laplace operator, with eigenvalues $\lambda_{\mathbf{k}} = -k^2$:

$$\Delta \mathbf{A}_{\mathbf{k}}(\mathbf{x}) = -k^2 \mathbf{A}_{\mathbf{k}}(\mathbf{x}), \quad k^2 = k_x^2 + k_y^2 + k_z^2. \quad (6.25)$$

A little more thought is required to verify the boundary conditions $\mathbf{B}_{\perp} = \mathbf{E}_{\parallel} = 0$. Concerning the electric condition, note that in the Coulomb gauge, $\mathbf{E} = -\nabla\phi - \frac{1}{c}\partial_t\mathbf{A} = -\frac{1}{c}\partial_t\mathbf{A}$. While we have not yet talked about time dependence, the vanishing of components of \mathbf{A} at the boundary is sufficient to make the corresponding components of \mathbf{E} vanish. On this basis, check that all boundary conditions are satisfied.

After this preparation, the further solution of the problem is straightforward. We substitute the 'mode decomposition' Eq. (6.25) into the action, and use the mode normalization to obtain

$$S[\alpha] = \frac{1}{8\pi} \sum_{\mathbf{k}} \int dt \left(\frac{1}{c^2} \dot{\alpha}_{\mathbf{k}}^2 - k^2 \alpha_{\mathbf{k}}^2 \right).$$

This action reminds us of that of the classical harmonic oscillator

$$S[q] = \int dt \left(\frac{m}{2} \dot{q}^2 - \frac{m\omega^2}{2} q^2 \right).$$

More precisely, it equals the sum over \mathbf{k} of harmonic oscillator actions of particles with mass $m = 1/(4\pi c^2)$, and spring constant $\omega^2 = c^2 k^2$. We notice another advantage of the Lagrangian approach: it reveals connections between physical systems which are less easy to see in different ways. In hindsight, the harmonic oscillator analogy is rather natural. For each value of \mathbf{k} we select a specific oscillation mode of the wave guide. The dynamics of that mode is oscillatory, and $\alpha_{\mathbf{k}}$ plays the role of its oscillation coordinate, as described by the corresponding action. We may make this quantitative by varying the action to obtain the equations

of motion $\ddot{\alpha}_{\mathbf{k}} = -\omega^2 \alpha_{\mathbf{k}}$, whose solutions show harmonic time dependence with frequency ω as we should expect from an electromagnetic wave solution.

As with the full action of the electromagnetic field, we may pass from the Lagrangian action $S[\alpha]$ to a Hamiltonian formulation: define the canonical momentum $\pi_{\mathbf{k}} \equiv \delta_{\dot{\alpha}_{\mathbf{k}}} L = m\dot{\alpha}_{\mathbf{k}}$ to obtain the Hamiltonian action

$$S[\alpha, \pi] = \sum_{\mathbf{k}} \int dt \left(\pi_{\mathbf{k}} \dot{\alpha}_{\mathbf{k}} - \frac{1}{2m} \pi_{\mathbf{k}}^2 - \frac{m\omega^2}{2} \alpha_{\mathbf{k}}^2 \right),$$

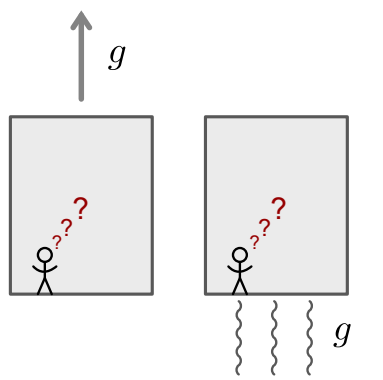
In terms of ‘coordinates and momenta’. At this point, we take a bold next step and *quantize* the theory. As in classical mechanics, this is done by promoting the scalar coordinates $\alpha_{\mathbf{k}}, \pi_{\mathbf{k}}$ to hermitian operators with proclaimed commutation relations $[\hat{\alpha}_{\mathbf{k}}, \hat{\pi}_{\mathbf{k}}] = i\hbar$. This will get us from a system of classical to one of quantum harmonic oscillators, with energy spectra $\hbar\omega(n + 1/2), n = 0, 1, \dots$. While the detailed discussion of the emerging **quantum field** is beyond the scope of this text and the subject of **quantum electrodynamics**, we here just mention two striking observations. First, already the ground state energy of the electromagnetic field $\hbar \sum_{\mathbf{k}} \frac{\omega_{\mathbf{k}}}{2}$ is non-vanishing. This is the famous zero point energy of the electromagnetic field, which is physically observable. Second, excitation on top of the ground state have signatures of quantum particles, presently called **photons**. Their emergence has the striking consequence that on action scales resolving the finiteness of \hbar , the electromagnetic field becomes ‘granular’ and particle like.

6.6 A glimpse of general relativity

We conclude this chapter with a brief outlook into general relativity. This extension *generalizes* special relativity in that the focus moves beyond inertial frames. An inertial frame is one in which bodies remain at rest or in states of uniform motion once initialized in such states.

Of course, there is no such thing as a genuine inertial frame in nature. Inertia is a state that can be reached approximately, if gravitational bodies — in relativity, the focus is on gravitational forces — are sufficiently far away. Newtonian mechanics posits that masses generate an instantaneous force, which then acts upon bodies via acceleration, removing inertia.

6.6.1 Strong equivalence principle

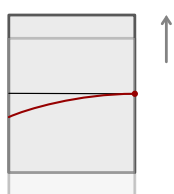


Einstein's path towards understanding the physics of non-inertial frames arguably started with the attempt to remove 'force' as a middle agent between gravitation and acceleration. While we are accustomed to thinking of 'force' as the cause of acceleration, the concept of switching frames, practiced in relativity for inertial frames, suggests alternative views. The links between gravity and acceleration emerging in this way are subsumed under the name **equivalence principles**. The so-called *weak* equivalence principle states the equivalence of inertial and gravitational mass: In Newton's equation $m\ddot{x} = -mg$,

the same ' m ' appears on the two sides of the equation. Masses accelerate in a gravitational field with the same rate. We are so accustomed to this fact, that its non-triviality is easily forgotten: why doesn't a heavy rock accelerate differently from a light pebble? The strong(er) equivalence principle posits that there is no way by which an observer in a uniformly accelerated elevator cabin (no windows) can tell the difference to a cabin at rest in a uniform gravitational field. Similarly, an observer inside a cabin falling in a gravitational field will find that she is in an inertial frame: objects moving at constant velocity relative to her will remain in this state.

There are two classes of conclusions to be drawn from the strong equivalence principle. The first is that effects commonly attributed to the action of forces can be equally described by changes of the system of reference. (This concept is closely related to that of 'fictitious forces'.) We come back to this point momentarily. The other are concrete, and observable predictions concerning the physics of light.

6.6.2 The bending of light and gravitational redshift



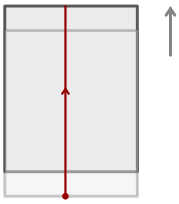
Imagine a light ray emitted in transverse direction inside a uniformly accelerated cabin. During the time the light needs to travel between the cabin walls, the latter will have moved upwards, i.e. the ray hits the other side at a point lower than the point of emission. If we take the equivalence principle seriously, we have to conclude that light gets attracted towards gravitational centers. The effect was observed in 1919 in experiments

probing light propagating close to the outer rim of Sun.

A related but somewhat more involved effect is gravitational redshift. Recall the Doppler effect, which is the shift in frequency by $\Delta\omega = -\omega v/c$ experienced by a receiver moving with velocity v relative to an emitter of optical radiation of frequency ω .

▷ **Info.** To understand the **Doppler effect**, imagine a situation where a wave crest has just passed the moving observer. What will be the time t_o when the next crest comes through? Light travels with velocity c . During the time t_o , the observer will have moved by a distance vt_o . So the light ray must actually travel a distance $\lambda + t_ov$. It does so at speed c , which leads to the equation $\lambda + t_ov = ct_o$, which we solve for $t_o = \lambda/(c - v) \equiv 2\pi/\omega_o$. With $c/\lambda = \omega$, we conclude

$\omega_o = \omega(1 - v/c)$, which is the above result.



Assume that a light ray has been emitted from the bottom of the accelerated cabin towards the ceiling. Denoting the height of the cabin by h , it will take a time $t \simeq h/c$ to reach the ceiling. We write ' \simeq ' because the actual time is somewhat longer, in time t the cabin ceiling will have moved a little away from the source. At that time, the upper wall will have reached a velocity $v = gt = gh/c$, and hence we expect a Doppler shift $\Delta\omega = -\omega v/c = -\omega gh/c^2$. Being inversely proportional to the square of the speed of light, gravitational redshift is a small effect, which however has been observed in astronomic experiments. Also notice the proportionality to gh , a factor otherwise entering the potential energy of heavy objects.

6.6.3 The idea behind general relativity

The phenomena discussed above rely on the equivalence principle and the constancy of light. Otherwise, they do not go beyond the level of Newtonian mechanics. However, beyond these predictions, the equivalence principle triggered a revolutionary development: it suggested a conceptually novel way of thinking about motion under the influence of gravity. The equivalence principle addresses motion in different 'frames'. In inertial frames far detached from gravitational fields, particles move along straight lines, which are lines of shortest distance, or straight *geodesics*. If we move closer to gravitational centers or, equivalently, change into an accelerated frame, the motion becomes curved. What if particles continue to move along geodesics, only that these are no longer straight but curved, because gravity has changed the intrinsic properties of ambient space? Metaphorically thinking of empty space as an infinite sheet of rubber, and gravitational centers as masses immersed into the sheet to induce *curvature*, the observable trajectories would be lines of shortest distance in this geometrically distorted environment. What we commonly attribute to the acceleration due to gravitational forces would be the departures from straight geodesics.

Much as Newton's theory, such a picture cannot be proven from earlier principles. However, it can be *tested* via the successful prediction of phenomena Newton mechanics cannot explain. The actual execution of this geometric approach to gravity then has two parts to it: deriving quantitative results for the curvature of space in the presence of mass or, equivalently, *energy*¹⁶. The second is the prediction of kinematic motion for a given curvature profile.

The first part of this program was by far the hardest, and it took Einstein more than ten years to execute it. In short, Einstein proposed to describe the geometric properties of space-time in terms of a local metric $g_{\mu\nu}(x)$, differing from the Minkowski metric $\eta_{\mu\nu}$ in regions close to gravitational centers, but sharing with it the *signature*: three eigenvalues are negative, one positive. In this framework, $g_{\mu\nu}(x)$ is a tensor valued field, whose profile must be obtained

¹⁶The relation $E = mc^2$ of special relativity implies that mass and energy stand more or less on the same footing concerning the above argument.

from field equations having a conceptual status similar to the Maxwell equations of electromagnetism. These **Einstein equations** contain mass and energy as sources. Otherwise, they are of second order in space-time derivatives and nonlinear. This second feature makes them much harder to solve than Maxwells equations, and it implies that there is no superposition principle in gravity. (Only for weak sources can the equations be linearized, in which limit they predict wave like solutions. These *gravitational waves* have been a center of much excitement recently.)

The discussion of Einstein's equations is at the heart of general relativity and a subject beyond the scope of this text. Instead, we will restrict ourselves to a brief discussion of their simplest nontrivial solution the **Schwarzschild metric**¹⁷.

The Schwarzschild metric is an exact solution of Einstein's equations, solving it outside a spherical mass. (In this regard, its status is not so different from that of spherical wave solutions of Maxwell's equations outside rotationally symmetric sources.) In spherical coordinates (r, θ, ϕ) of three-dimensional space, it is given by the invariant line element

$$c^2 d\tau^2 = \left(1 - \frac{r_s}{r}\right) c^2 dt^2 - \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 - r^2 d\Omega^2, \quad (6.26)$$

where $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ is the metric of a unit sphere. In this way, the tensor $g_{\mu\nu}$ specifying the metric is defined as $g_{00} = 1 - r_s/r$, $g_{rr} = -(1 - r_s/r)^{-1}$, $g_{\theta\theta} = -r^2$ and $g_{\phi\phi} = -r^2 \sin^2 \theta$, with all other elements vanishing.

The metric has a singularity at the origin $r = 0$ which corresponds to a center of infinite intrinsic curvature. This singularity upon approaching the center of a mass distribution should not be too surprising. However, there also is a second singularity at $r = r_s$, the **Schwarzschild radius**. Notice how weird this singularity is. Due to the vanishing of the prefactor in $\left(1 - \frac{r}{r_s}\right) c^2 dt^2$ a distant observer at $r \rightarrow \infty$ will conclude that an in-falling object can never reach the **event horizon** at $r \rightarrow r_s$. Conversely, no object can escape from there. It takes infinite time to overcome the vanishing metric prefactor. This is one way of reasoning that from the perspective of the outside world, the inside $r < r_s$ is barren. However, one can show that there actually is no physical singularity at the event horizon, the singularity of the metric above is due to a system of coordinates ill-suited to describe the vicinity of it. As a case in point, consider the spherical metric $\sin^2 \theta d\phi^2 + d\theta^2$ close to the north pole. It *seems* to be singular, because infinitely many ϕ -coordinate lines are converging. However, a person actually walking across the north-pole will find no singularity there. Similarly, an in-falling observer will sense no singularity at the event horizon, and effortlessly pass it in finite time. Trouble starts only, when he or she approaches the origin $r \rightarrow 0$.

The Schwarzschild metric describes spherically symmetric mass distributions. This includes **black holes** for which the event horizon marks a point of no return. For more benign mass distributions the Schwarzschild radius is buried deep within the physical extension of the mass density. For example, the Schwarzschild radius of Earth is at $r_s \simeq 9$ mm.

Before discussing this metric further, let us investigate how it will affect the motion of bodies passing through it.

¹⁷Named in honor of Karl Schwarzschild who discovered it in 1915, a month after Einstein had published his field equations, and shortly before his death due to an illness contracted in World War I.

6.6.4 Geodesic motion

In this section, we discuss how a metric distorted away from the Minkowski limit affects the motion of bodies. The working postulate is that physical motion will be along the geodesics, i.e. curves of shortest length of a metric.

Geodesics are space-time curves $x = x(\tau)$ extremizing the length functional

$$L[x] = \int d\tau (d_\tau x^\mu(\tau) g_{\mu\nu}(x(\tau)) d_\tau x^\nu(\tau))^{1/2}. \quad (6.27)$$

Notice that in the case of a trivial metric $g_{\mu\nu}$, this expression reduces to the length of a curve in an Euclidean setting. Also notice that in order to compute the length, we need to pick a representation $x^\mu(\tau)$. However, the value of the functional is invariant with respect to this choice (why?).

For all what follows, it will be convenient to parameterize the curve by its own length. To this end, pick an arbitrary parameterization $x^\mu(t)$, and define

$$c\tau(t) = \int_0^t ds (d_s x^\mu(s) g_{\mu\nu}(x(s)) d_s x^\nu(s))^{1/2}.$$

We will not compute this function explicitly. All what matters is that it exists, and that it satisfies the property (why?) $d_t \tau(t) = c^{-1} (d_t x^\mu g_{\mu\nu} d_t x^\nu)^{1/2}$, where we simplified the notation by omitting arguments. Also notice that in the length representation

$$d_\tau x^\mu g_{\mu\nu} d_\tau x^\nu = d_t x^\mu g_{\mu\nu} d_t x^\nu \left(\frac{dt}{d\tau} \right)^2 = c^2,$$

by the chain rule, i.e. the curve is traversed with uniform 'velocity'. Finally, for infinitesimal δ_t we can write $\delta_t d_t x^\mu = \delta x^\mu$, leading to $\delta\tau^2 = \delta x^\mu g_{\mu\nu} \delta x^\nu$. This construction tells us that geodesic 'length increments' equal increments in the invariant metric distance $\delta x^\mu g_{\mu\nu} \delta x^\nu$. Specifically, in a frame co-moving with the curve, $\delta x^i = 0$, the parameter τ becomes the proper time of an observer moving along it. (All these statements are worth thinking about.)

Using the compactified notation $d_\tau x^\mu(\tau) = \dot{x}^\mu$, we vary the action to obtain

$$\begin{aligned} \delta L[x] &= L[x + \epsilon] - L[x] = \int d\tau [(\dot{x} + \dot{\epsilon})^\mu g_{\mu\nu}(x + \epsilon) (\dot{x} + \dot{\epsilon})^\nu]^{1/2} - (\dot{x}^\mu g_{\mu\nu}(x) \dot{x}^\nu)^{1/2} = \\ &= \int d\tau \frac{1}{2(\dot{x}^\mu g_{\mu\nu} \dot{x}^\nu)^{1/2}} (2\dot{\epsilon}^\mu g_{\mu\nu} \dot{x}^\nu + \dot{x}^\mu \partial_\rho g_{\mu\nu} \epsilon^\rho \dot{x}^\nu) + \mathcal{O}(\epsilon^2). \end{aligned}$$

At this point, the advantage of using the length parameterization becomes evident: the denominator of the fraction upfront equals unity. Proceeding as in analogous calculations in classical mechanics, we integrate by parts to remove the time derivative from ϵ and use the chain rule $d_\tau g_{\mu\nu} = \partial_\rho g_{\mu\nu} \dot{x}^\rho$ to obtain

$$\delta L[x] = \int d\tau \epsilon^\mu \left(-g_{\mu\nu} \ddot{x}^\nu - \partial_\rho g_{\mu\nu} \dot{x}^\rho \dot{x}^\nu + \frac{1}{2} \partial_\mu g_{\rho\nu} \dot{x}^\rho \dot{x}^\nu \right) + \mathcal{O}(\epsilon^2).$$

Stationarity requires the vanishing of the term in parentheses. Multiplying it with $g^{\lambda\mu}$ to remove the metric multiplying the double time derivative: $g^{\lambda\mu}g_{\mu\nu}\ddot{x}^\nu = \ddot{x}^\lambda$ and symmetrizing the center term, we obtain the **geodesic equation** as

$$\ddot{x}^\lambda + \Gamma_{\rho\nu}^\lambda \dot{x}^\rho \dot{x}^\nu = 0, \quad \Gamma_{\rho\nu}^\lambda = \frac{1}{2}g^{\lambda\mu} (\partial_\rho g_{\mu\nu} + \partial_\nu g_{\mu\rho} - \partial_\mu g_{\rho\nu}) \quad (6.28)$$

where the functions $\Gamma_{\rho\nu}^\lambda$ are known as the **Christoffel symbols**. These are important equations. They describe geodesic motion, i.e. motion along connections of extremal length for arbitrary geometries.

▷ EXERCISE. Interpret the equations for an Euclidean metric $g_{\mu\nu} = \delta_{\mu\nu}$. Why are the solutions of these equations geometrically straight curves?

▷ EXERCISE. Consider the metric of the sphere: $g_{\theta\theta} = 1$, $g_{\phi\phi} = \sin^2 \theta$ in spherical coordinates. Show that in this case the geodesic equations assume the form:

$$\ddot{\theta} = \sin \theta \cos \theta \dot{\phi}^2, \quad \ddot{\phi} = 2 \frac{\cos \theta}{\sin \theta} \dot{\phi} \dot{\theta}.$$

These are the equations one needs to solve if one wants to describe, e.g. optimal flight routes on a globe with boundary conditions specified as $(\theta, \phi)(\tau_{\text{initial}})$ and $(\theta, \phi)(\tau_{\text{final}})$. They simplify greatly for specific choices of initial conditions. For example, it is evident that uniform ($\dot{\theta} = \text{const.}$) motion along a circle of constant polar angle $\phi = \text{const.}$ is a solution.

6.6.5 Newtonian limit of gravitational dynamics

In this section, to touch base with the familiar Newton dynamics, we consider motion in the case where a metric is almost, but not quite Minkowski. This should give us a clue as to the general workings of the theory. For concreteness, we consider the Schwarzschild metric at a point $r \gg r_s$ far away from the Schwarzschild radius. We aim for solutions describing motion at velocities $v \ll c$. This limit implies that parametric variations of the coordinates $dx^0 = cdt$ will be way larger than those of spatial coordinates dx^i . We can also use that $dx^0 = cdt \approx cd\tau$, or $\frac{dt}{d\tau} \approx 1$: proper time and coordinate time are approximately equal. Similarly, in the line element $c^2 d\tau^2 = c^2 dt^2 g_{00} + dx^i dx^i g_{ii}$ only deviations away from the Minkowski metric in the component g_{00} are significant.

Using these simplifications, we write $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, where the weak distortion follows from Eq. (6.26) as $h_{\mu\nu} \approx -\delta_{\mu 0} \delta_{\nu 0} (r_s/r)$. Now consider the geodesic equation Eq. (6.28), and let's be maximally economical. On the r.h.s. we have coordinate derivatives of the distortion $h_{\mu\nu}$. Since we are working to first order in h , we may approximate $g^{\mu\nu} = \eta^{\mu\nu}$. We also approximate $\dot{x}^\rho = \delta^{\rho 0} c\dot{t} \approx c\delta^{\rho 0}$. On this basis we obtain

$$\Gamma_{00}^\lambda \approx \frac{1}{2}\eta^{\lambda\mu} (0 + 0 + \partial_\mu h_{00}),$$

where the first two 0 are due to the time independence of the metric $\partial_0 g_{\mu\nu} = 0$. For the same reason, $\partial_0 h_{00} = 0$ and $\Gamma_{00}^0 = 0$. The first of our geodesic equations thus becomes $\ddot{x}^0 = 0$ consistent with $x^0 = ct$. The spatial equations become

$$\ddot{x}^i \approx \frac{r_s c^2}{2} \partial_i \frac{1}{r}.$$

This is consistent with the Newton equation in a gravitational potential $V = -MG/r$, if we identify the Schwarzschild radius as $r_s = 2MG/c^2$. In this case, the equation for a test body of mass m assumes the form $m\ddot{x}^i = -\partial_i mV = F_i$.

Notice what we have got: far away from gravitational centers, the geodesic equations reduce to the Newton equations in a 'fictitious' force. The force is fictitious in the sense that it represents an intrinsic geometric property of space (the metric). Also, the equivalence principle is hardwired into our discussion, i.e. acceleration is independent of the mass, m , of a test body. Finally, it is straightforward to include next to leading order corrections in powers of c^{-1} into the analysis. As a result one obtains deviations from Newtonian dynamics which naturally are strongest close to gravitational centers. The first prediction of this kind has been a deviation of Mercur's (the planet closest to Sun) orbit away from a Kepler ellipse. While gravitational redshift and the distortion of light rays were relatively straightforward consequences of the equivalence principle, this one required general relativity in its full form. The successful comparison with experimental data (some going back to the mid 19th century) marked the ultimate victory of this theory.

7.0 Hydrodynamics

We commonly distinguish three forms of macroscopic matter: gaseous, liquid, and solid. Informally, a solid resists external deformation forces, a gas doesn't, and it will fill any volume available to it. The intermediate form of a liquid does not resist deformation forces, but does not expand into available volumes either. One can be more educated concerning these definitions, but for our purposes they are sufficient.

The subject of this chapter is a nutshell introduction into liquid- or hydrodynamics, the theory describing the dynamics of liquids. Hydrodynamics is a prime example of a classical field theory: it starts by lumping large numbers of molecular constituents of a liquid into effective degrees of freedom. Following the general paradigm of field theory, hydrodynamics then translates Newton's equations for the liquid's microscopic components to equations of motion of these effective degrees of freedom. Many of the concepts applied in this construction will be familiar from our earlier discussion of electromagnetic fields. However, there is one crucial difference: hydrodynamics is not a linear theory, which makes its equations of motion way harder to solve than Maxwell's equations. We will put the emphasis in this chapter on the construction of the fundamental equations of this theory, and for a discussion of the amazingly rich phenomenology of hydrodynamics refer to specialized courses.

7.1 The concept of a fluid particle

Fluid particles are the minimal entities resolved by hydrodynamics. Think of them as assemblies of a macroscopically large number of fluid molecules, but small enough that all effective properties of the fluid — density, average velocity, effective forces, etc. — are uniform across it. Importantly, fluid particles are large enough to have a surface (in view of their smallness, the detailed form of the surface is of no relevance) which makes them geometric objects different from, say, the point particles often considered in abstractions of classical mechanics.

There are two different approaches to describing the dynamics of fluid particles: the ambitious **Lagrangian view** considers all fluid particles present inside a macroscopic volume V_0 at time t_0 . A particle initially at time $\mathbf{x} \in V_0$ will move to $\mathbf{x}(t)$ in time t , implying that the whole volume will turn into

$$V_t \equiv \{\mathbf{x}(t) | \mathbf{x}(0) \in V_0\}.$$

Knowledge of all maps $V_0 \rightarrow V_t$ is equivalent to a complete characterization of the fluid's dynamics. However, it requires the tracking of all fluid particle trajectories, which in most

cases of interest is excessive. The alternative and more practical **Eulerian view** reasons that the state of the fluid is described by a *field* $\mathbf{v}(\mathbf{x}, t) \equiv \mathbf{v}(x)$ giving the instantaneous velocity of fluid particles at space-time coordinate $(\mathbf{x}, t) \equiv x$. Intuitively, this vector field describes the current profile of the liquid, and its identification can be considered a solution of the problem. We will adopt the Eulerian view throughout.

7.2 Evolution of integral quantities carried by a fluid

Assume that the vector field $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}(x)$ was known and consider a physical property $f(\mathbf{x}, t)$ attached to a fluid element at \mathbf{x} . Examples include its mass, energy, or momentum, for the time being, we need not be specific. We allow for explicit time dependence, e.g. $f = T(t)$ a temperature profile explicitly depending on time. A volume V_t at considered at some time t will then include the integral amount of f ,

$$I_t \equiv \int_{V_t} dV f(\mathbf{x}, t),$$

where dV is the three-dimensional volume element. How will this quantity vary during an infinitesimal instance of time, δt ? Several things are time dependent here: a fluid element will move from $\mathbf{x} \rightarrow \mathbf{x} + \mathbf{v}\delta t$, where we write $\mathbf{v}(\mathbf{x}, t) = \mathbf{v}$ for simplicity. We thus have $f(\mathbf{x}, t) \rightarrow f(\mathbf{x} + \mathbf{v}\delta t, t + \delta t)$. However, the volume also changes to $V_{t+\delta t} \equiv \{\mathbf{y} = \mathbf{x} + \mathbf{v}\delta t | \mathbf{x} \in V_t\}$. We thus have

$$I_{t+\delta t} = \int_{V_{t+\delta t}} dV f(\mathbf{x} + \mathbf{v}\delta t, t + \delta t).$$

A good way to obtain a quantitative expression for the difference $I_{t+\delta t} - I_t \equiv \delta t d_t I + \mathcal{O}(\delta t^2)$ is to think of the assignment $\mathbf{y} : V_t \rightarrow V_{t+\delta t}$, $\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x}) \equiv \mathbf{x} + \delta t \mathbf{v}$ as an invertible and smooth (why?) map between V_t and the time evolved $V_{t+\delta t}$. The theorem for change of variables in higher-dimensional integration then states that

$$I_{t+\delta t} = \int_{V_{t+\delta t}} d^3 y f(\mathbf{y}, t + \delta t) = \int_{V_t} d^3 x \left| \frac{\partial y}{\partial x} \right| f(\mathbf{y}(\mathbf{x}), t + \delta t),$$

where $J \equiv |\partial y / \partial x|$ is the Jacobi determinant of our map. The difference between $I_{\delta t}$ and I_0 is that in the latter we are integrating the original $f(\mathbf{x}, t)$ over V_t , instead of the time evolved $f(\mathbf{x}(\mathbf{y}), t + \delta t)$. Thus

$$I_{t+\delta t} - I_t = \int_{V_t} d^3 r \left(\left| \frac{\partial y}{\partial x} \right| f(\mathbf{y}(\mathbf{x}), t + \delta t) - f(\mathbf{x}, t) \right).$$

To obtain a more concrete expression, we approximate $f(\mathbf{y}(\mathbf{x}), t + \delta t) = f(\mathbf{x} + \mathbf{v}\delta t, t + \delta t) \approx (\partial_t + v^i \partial_i) f(\mathbf{x}, t) \delta t$. This combination of derivatives occurs so frequently in hydrodynamics that it is given its own name: the **material derivative** of a function is defined as

$$\frac{Df(\mathbf{x}, t)}{Dt} \equiv (\partial_t + v^i \partial_i) f(\mathbf{x}, t).$$

Note that the material derivative makes reference to a velocity vector field. We finally compute the Jacobian of our map as

$$J = \det \left(\frac{\partial y}{\partial x} \right) = \exp \operatorname{tr} \ln \left(\left\{ \frac{\partial x^i + v^i \delta t}{\partial x^j} \right\} \right) = \exp \operatorname{tr} \ln (\{\delta_j^i + \delta t \partial_j v^i\}) \approx \exp(\delta t \partial_i v^i) \\ \approx 1 + \delta t \nabla \cdot \mathbf{v},$$

where we encounter the divergence of the vector field $\mathbf{v}(\mathbf{x})$. Combining everything, we obtain

$$\boxed{d_t I = \int_{V_t} dV \left(f(x) \nabla \cdot \mathbf{v}(x) + \frac{Df(x)}{Dt} \right)} \quad (7.1)$$

a result that will play an important role in our further analysis of the liquid's dynamics. Eq. 7.1 simplifies for the case of **incompressible liquids**. A liquid is incompressible if the volume enclosed by it does not change in time. Many liquids satisfy this condition to a very good approximation. To understand the consequences, consider the case $f = 1$, in which case $I_t = V_t$ is the enclosed volume. The equation

$$d_t V_t = \int_{V_t} dV \nabla \cdot \mathbf{v}(x) = 0,$$

must hold for all V_t , requiring that

$$\nabla \cdot \mathbf{v} = 0.$$

Intuitively the absence of sources in the velocity field of an incompressible liquid means that fluid elements neither expand nor contract.

7.3 Conservation laws

Let us apply the identity derived above to explore a number of conservation laws constraining the motion of a liquid.

7.3.1 Mass conservation

One of the most basic properties of a liquid is its mass (distribution). We define $\rho(x)$ to be a mass density such that $\rho(x)dV$ is the mass contained in a volume element dV . The mass contained in a macroscopic volume V_t then is

$$M_t \equiv \int_{V_t} dV \rho(x),$$

and mass conservation requires that it does not change in the course of time: $d_t M_t = 0$. Application of our auxiliary relation to the case $f = \rho$ then yields

$$0 = \int_{V_t} d^3x \left(\rho(x) \nabla \cdot \mathbf{v}(x) + \frac{D\rho(x)}{Dt} \right)$$

Holding for arbitrary V_t , this requires the vanishing of the integrand (cf. the same rationale applied in passing from the integral to differential laws of electromagnetism), leading to the condition

$$0 = \rho(\nabla \cdot \mathbf{v}) + \frac{D\rho}{Dt} = \rho(\nabla \cdot \mathbf{v}) + \mathbf{v} \cdot \nabla \rho + \partial_t \rho = \nabla \cdot (\rho(x)\mathbf{v}(x)) + \partial_t \rho(x),$$

where in the last step, we used the product rule of differentiation. We have thus obtained the important **mass continuity equation**,

$$\boxed{\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0.} \quad (7.2)$$

Comparison with the generic form of continuity equations: $\partial_t(\text{density}) + \nabla \cdot (\text{current}) = 0$, leads to the identification of the **mass current**

$$\mathbf{j} = \rho \mathbf{v}.$$

We may integrate Eq. (7.2) over an arbitrary volume V and apply Stokes theorem to obtain a relation for the (generally time dependent) mass M contained in V :

$$d_t M = d_t \int_V dV \rho = \int_V dV \partial_t \rho = - \int_V dV \nabla \cdot \mathbf{j} = - \int_S dS \cdot \mathbf{j},$$

i.e. the mass change is due to the mass current flow through V 's surface, as one would expect. Finally, note that incompressibility ($\nabla \cdot \mathbf{v} = 0$) does not necessarily imply constancy of the mass density ($\nabla \rho = 0$). For an incompressible liquid, the mass continuity relation becomes $\partial_t \rho + \mathbf{v} \cdot \nabla \rho = 0$. We may consider situations where the mass density of an incompressible liquid changes in time, $\partial_t \rho \neq 0$, e.g., due to changes in temperature. Such situations require both, a non-vanishing gradient $\nabla \rho \neq 0$, and liquid flow, $\mathbf{v} \neq 0$, to be consistent with mass conservation. (Think about this point.) Similarly, in equilibrium, $\partial_t \rho = 0$, an incompressible liquid may admit density changes, $\nabla \rho \neq 0$, due to, e.g., spatial changes in the chemical composition of a liquid, provided their gradients are perpendicular to velocity $\nabla \rho \cdot \mathbf{v} \neq 0$.

As a corollary, we note that numerous physical properties are tied to mass. For example, the momentum carried by a fluid element with volume δV is $\delta V \rho \mathbf{v}$, making $\rho \mathbf{v}$ a momentum density. Let $f = \rho \psi$ be any such quantity. Its change in relation to a volume V_t is given by Eq. (7.1). However, by the product rule, the material derivative featuring in this equation assumes the form

$$\frac{Df}{Dt} = \frac{D\rho}{Dt} \psi + \rho \frac{D\psi}{Dt}.$$

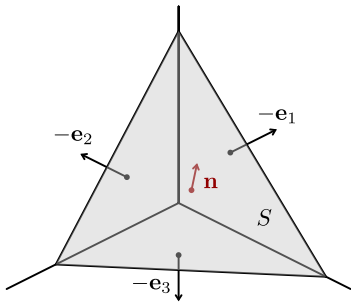
Rewriting the mass continuity relation as $\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = \frac{D\rho}{Dt} + \rho(\nabla \cdot \mathbf{v}) = 0$, we obtain the nice result

$$d_t I = \int_{V_t} d^3x \rho \frac{D\psi}{Dt}. \quad (7.3)$$

7.3.2 Stress tensor

The next entry on our list of conserved quantities is momentum. We know that changes in momentum are due to the action of forces, so we must take a closer look at the forces acting on fluid elements. Consider a fluid element of small extension δ volume $\delta V \sim \delta^3$ and surface $\delta S \sim \delta^2$. We may anticipate the presence of two types of forces: the first are 'volumic forces', defined as external forces coupling to the masses inside our fluid element. The gravitational force is an example of this type. The net volumic force is proportional to mass which in turn is proportional to volume, i.e. $\sim \delta^3$. In addition, the entirety of other fluid elements will exert a force to our fluid element in the form of pressure and shear. Pressure is a force perpendicular to the elements surface, and shear is the drag tangential to it arising when fluid elements are in relative motion to each other. What both have in common is that they are proportional to surface area and hence $\propto \delta^2$.

The dynamics of our fluid element will be governed by an equation of motion of the structure (mass) \times (acceleration) = (forces). Since (mass) $\propto \delta V \sim \delta^3$ the only way to make this an equation in the limit of small δ is the vanishing of the total surface force, i.e. the terms of $\mathcal{O}(\delta^2)$.



To explore this condition in more detail, consider a volume of tetrahedron geometry, as indicated in the figure. The force acting on each of its surface elements, will depend on the orientation of the surface, as described by its normal vector \mathbf{n} . It will also be proportional to the surface area. Presently, the total surface force $\mathbf{F}_{\text{tot}} = \sum_{a=1}^3 \mathbf{F}^a + \mathbf{F}$ is the sum of three forces \mathbf{F}^a acting on the three areas inside the coordinate planes, and the force \mathbf{F} acting on the surface S . Defining $\mathbf{n} = n^i \mathbf{e}_i$

to be its normal vector, an elementary geometric construction shows that the surfaces S_j of the coordinate plane areas are given by $S_i = n^i S$. We need to make sure that $|\mathbf{F}_{\text{tot}}| = \mathcal{O}(\delta^3)$, i.e. that the forces proportional to areas mutually cancel out. To establish this property, we postulate the existence of a **stress tensor** $\sigma = \{\sigma_{ij}(x)\}$ defined such that the i th component of the force on an area element normal vector \mathbf{n} is given by

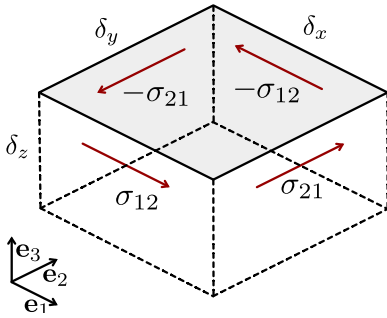
$$F_i = \sigma_{ij} n^j S.$$

In other words, σ encodes the vectorial orientation of the force per unit area in dependence of the orientation of the surface element. With this ansatz, the cancellation property holds. To see this, note that the normal vectors to S_i are $-\mathbf{e}_i$, leading to

$$F_{\text{tot},i} = \sum_j F_i^j + F_i = -\sigma_{ij} S_j + \sigma_{ij} n^j S = -\sigma_{ij} n^j S + \sigma_{ij} n^j S = 0.$$

Mathematically, the stress tensor is a covariant tensor of second degree. This is to say that it requires a vector (a contravariant tensor of first degree) to produce a force, which is covariant, first degree. The action of this force on an infinitesimal displacement vector, s^i then produces

the work per unit area $s^i \sigma_{ij} n^j$ associated to the displacement of a surface element. Besides these formalities, what else can we say about the stress tensor?



Isotropy of pressure: A fluid at rest does not support any shear forces, i.e. forces tangential to surface elements. (If there were any, it would respond with a shear motion, contradicting the condition of equilibrium.) This means that the work associated to displacement in the direction of any vector perpendicular to \mathbf{n} , i.e. $\mathbf{n} \times \mathbf{v}$ for arbitrary \mathbf{v} vanishes: $(\mathbf{n} \times \mathbf{v})^i \sigma_{ij} n^j = 0$. Choosing $\mathbf{v} = \mathbf{e}_l$ a unit vector, we obtain the condition $\epsilon_{ikl} n^k \sigma_{ij} n^j = 0$.

This must be zero, regardless of the orientation of \mathbf{n} , which requires $\sigma_{ij} \propto \delta_{ij}$. We conclude that for a static fluid the stress tensor reduces to the unit-tensor. Its diagonal elements $\sigma_{ij} = -\delta_{ij} P$ define the magnitude of the **pressure**, P , acting on a fluid element: the force per unit area is given by $\sigma_{ij} n^j = -n_i P$, i.e. it is perpendicular to the normal vector, and inward directed.

Absence of surface torque: The introduction of the stress tensor was based on the condition that the volumic and surface forces acting on a fluid element separately cancel out. Now let us apply a similar consideration to the torque acting on a fluid element, as shown in the figure. Volume and surface torque again scale with different powers of δ . Torque causes changes in angular momentum, which is of the same order as the volume torque (why?). We conclude that the total surface torque must vanish. For definiteness, let us consider the three component of the torques, $N_3 = (r^1 F_2 - r^2 F_1)$ acting on points (r^1, r^2, r^3) relative to an arbitrary point inside the fluid element. The tangential forces acting on the vertical surfaces indicated in the figure yield the torque as (think why) $N_3 = (\delta_x \sigma_{21} (\delta_y \delta_z) - \delta_y \sigma_{12} (\delta_x \delta_z)) = (\sigma_{12} - \sigma_{21}) \delta V$. The vanishing of this expression requires $\sigma_{12} = \sigma_{21}$. Application of the same construction to the other components of the angular momentum shows the **symmetry of the stress tensor**,

$$\sigma_{ij} = \sigma_{ji}.$$

7.3.3 Momentum conservation

Consider the total momentum carried by a co-moving volume, $\mathbf{P}_t = \int_{V_t} dV \rho \mathbf{v}$. Thinking of this quantity as the sum of microscopic momenta, each of which is governed by Newton's law, we conclude $d_t \mathbf{P}_t = (\text{total force acting on } V_t) \equiv \mathbf{F}_t$. This force is the sum of volumic forces and surface forces, $\mathbf{F}_t = \mathbf{F}_{V_t} + \mathbf{F}_{S_t}$ (Recall that the total surface force vanishes only in the limit of an infinitesimally small volume). By definition, the latter is obtained by integration over

forces acting on surface element, which in turn is given by the stress tensor. In components

$$F_{V_t,i} = \int_{V_t} dV f_i,$$

$$F_{S_t,i} = \int_{S_t} dS \sigma_{ij} n^j = \int_{V_t} dV \partial_j \sigma_{ij}.$$

With the identification $\psi = v_i$ and Eq. (7.3) we thus obtain the balance

$$d_t P_i = \int_{V_t} d^3x \rho \frac{Dv_i}{Dt} = \int_{V_t} d^3x (f_i + \partial_j \sigma_{ij}).$$

This equation holds regardless of the realization of V_t , leading to the condition

$$\boxed{\rho \partial_t v_i + \rho v_j \partial_j v_i = f_i + \partial_j \sigma_{ij}.} \quad (7.4)$$

This is one of several equivalent formulations of the **Navier-Stokes equation**. It describes space and time dependent changes in the velocity field in response to external forces (such as gravity), and internal forces described by the stress tensor. However, before applying the equation as a tool to the prediction of physical properties of liquids, more specific assumptions concerning their internal forces, i.e. the stress tensor must be made. Depending on the model of the stress tensor, the Navier-Stokes equation is capable of describing the physical properties of almost all static and dynamical properties of liquids. While a comprehensive review of different model classes is beyond the scope of the present introduction, we mention the relatively simple, yet practically relevant of incompressible liquids with constant **viscosity**, ν . In this case, minimal further assumptions relying on Galilean invariance and isotropy of the fluid motivate the ansatz

$$\sigma_{ij} = -p \delta_{ij} + \rho \nu (\partial_i v_j + \partial_j v_i),$$

where ν defines the viscosity. (Note that due to the assumed incompressibility, the average over diagonal elements $\sum_i \sigma_{ii} = -3p + 2\rho \nu \nabla \cdot \mathbf{v} = -3p$ continues to be defined by the pressure.) In this case, the stress tensor term is effectively replaced by

$$\partial_j \sigma_{ij} = \rho \nu \Delta v_i - \partial_i p, \quad (7.5)$$

where we again used $\partial_i v_i = 0$. The Navier-Stokes equation then assumes the form

$$\rho \partial_t v_i + \rho v_j \partial_j v_i = f_i - \partial_i p + \rho \nu \Delta v_i, \quad (7.6)$$

of a coupled second order nonlinear differential equations.

7.4 Navier-Stokes equation: two simple examples

Even with the replacement Eq. (7.5), the Navier-Stokes equation is notoriously hard to solve in practice. Analytical solutions are only possible in cases where a combination of symmetries and simplifying assumptions provides an exceptionally high level of tailwind. We here discuss two such simple models solutions as examples.

7.4.1 Flow of ideal fluid around a spherical obstacle

Our first example is defined by a spherical obstacle placed into a fluid which far away from the sphere shows uniform flow described by $\mathbf{v} = v\mathbf{e}_x$. We consider 'moderate' velocities implicitly defined by the following two conditions. They should be fast enough to make viscous effects negligible $\nu = 0$. On the other hand, they should be slow enough to exclude the formation of whirls, i.e. turbulent motion.

The first of these conditions, collapses the Navier-Stokes equation to the much simpler **Euler equation** describing ideal (means non-viscous liquids)

$$\boxed{\rho\partial_t v_i + \rho v_j \partial_j v_i = f_i - \partial_i p.} \quad (7.7)$$

However, the Euler equation remains a complex nonlinear equation whose solution requires simplifying circumstances. Presently, we assume stationarity, $\partial_t v_i = 0$, the absence of external forces $f_i = 0$, and incompressibility, where we aim to resolve the latter condition in terms of a constant density profile $\rho = 0$. We also have boundary conditions in the problem: for $|\mathbf{x}| \gg a$, the radius of our sphere, we require $\mathbf{v}(\mathbf{x}) \rightarrow v\mathbf{x}$. We also know that $v_\perp(\mathbf{x}) = 0$ for $|\mathbf{x}| = a$: there is no velocity flow perpendicular to the spherical surface (why?). We then need to solve the reduced equation

$$\rho v_j \partial_j v_i = -\partial_i p. \quad (7.8)$$

▷ Info. Multiply this equation with v_i and sum over i , and note $v_i \partial_j v_i = \frac{1}{2} \partial_j |\mathbf{v}|^2$, to obtain the

equation $v_i \partial_i \left(\frac{\rho |\mathbf{v}|^2}{2} + p \right) = 0$. This is an interesting result. It tells us that along streamlines of the velocity (the directional derivative $v_i \partial_i$) profile of a stationary fluid, the sum of pressure and the kinetic energy density $\rho |\mathbf{v}|^2 / 2$ is constant. Larger velocity means smaller pressure, which is the principle behind **airfoils** such the wings of aircraft and birds. These wings are formed in such a way that the velocity stream along their upper surface exceeds that along the lower surface, resulting in an effective upward pressure.

We now solve this problem, inspired by its analogies to *electrostatics*. Anticipating the absence of vortical flow, we try to find a solution that is a gradient field $\mathbf{v} = \nabla \cdot \phi$. The condition of incompressibility, $\nabla \cdot \mathbf{v} = \Delta \phi = 0$, then becomes analogous to that describing an electric potential in a medium without sources via the Laplace equation. We will solve this equation, and then use Eq. (7.8) to *define* a scalar pressure field satisfying the static Euler equation.

Expressed in terms of ϕ , our boundary conditions read $\phi(\mathbf{x}) = vx$ for $|\mathbf{x}| \rightarrow \infty$, and $\partial_\perp \phi = \partial_r \phi = 0$ at the surface of the sphere, where r is the radial coordinate. The solution of these equations can be found in various ways, we here proceed by educated guessing. In view of the high level of symmetry of the field, one may hope that its multipole expansion, will terminate after dipole order. We thus consider an ansatz superimposing the potentials of

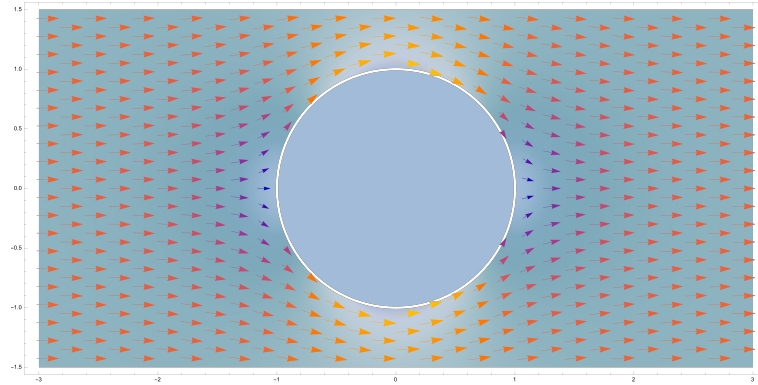


Figure 7.1: Static velocity field of a non-viscous fluid streaming around a sphere. The corresponding pressure profile is indicated by shading where lighter/darker regions correspond to low/high pressure, i.e. high/low velocity.

a constant field, a monopole, and a dipole potential (cf. Eq. (3.15))

$$\phi(\mathbf{x}) = vx + \frac{\xi}{r} + \eta \frac{\mathbf{e}_x \cdot \mathbf{x}}{r^3} = vr \cos \theta + \frac{\xi}{r} + \eta \frac{\cos \theta}{r^2},$$

with constants η , ξ , and θ the polar angle relative to the x -axis. Computing the radial derivative at $r = a$, we obtain the condition

$$v \cos \theta - \frac{\xi}{a^2} - 2\eta \frac{\cos \theta}{a^3}.$$

This needs to be satisfied for all angles θ , requiring $\xi = 0$, and $\eta = \frac{va^3}{2}$. We have thus found the solution

$$\phi(\mathbf{x}) = vx \left(1 + \frac{a^3}{2r^3} \right).$$

From here, it is straightforward to compute the velocity profile by forming gradients, and the pressure from Eq. (7.8), cf. Fig. 7.1.

7.4.2 Steady viscous flow through cylindrical pipe

Consider an infinitely long cylindrical pipe of radius R , and subject to a pressure gradient $d_z p = c = \text{const.}$. Compared to this pressure gradient, other forces are negligible, $f_i = 0$. We wish to determine the stationary fluid velocity distribution. In essence, this is an exercise in eliminating terms in the Navier-Stokes equation, now containing its viscous contribution $\propto \nu$. Stationarity means $\partial_t v_i = 0$. Working in cylindrical coordinates (r, θ, z) , we assume translational and rotational symmetry of the velocity field $v_i = v_i(r)$. Also, it is natural to assume flow parallel to the cylinder axis, $\mathbf{v} = v \mathbf{e}_z$.

We start by writing the equation in a vectorial form as

$$\rho \mathbf{v} \cdot \nabla \mathbf{v} = -\rho \nabla p + \rho \nu \Delta \mathbf{v}.$$

Under the stated assumptions, the term on the left-hand-side drops out, and on the right hand side, the only contribution to the Laplace operator is the one differentiating in r -direction: $\Delta = \frac{1}{r} \partial_r r \partial_r + \dots$. The equation thus collapses to

$$-c + \nu \frac{1}{r} \partial_r r \partial_r v(r) = 0.$$

The best way to go from here is to determine the general solution of the differential equation and to obtain a specific solution by matching boundary conditions. We multiply the equation by r and integrate once to obtain $\frac{c}{2\nu} r^2 = r \partial_r v(r) + C_1$. Division by r followed by another integration yields

$$v(r) = \frac{c}{4\nu} r^2 - C_1 \ln r + C_2.$$

The constants can now be fixed by requiring $\partial_r v(r) = 0$ at the cylinder axis $r = 0$, requiring $C_1 = 0$, and $v(R) = 0$, i.e. absence of flow at the cylinder boundary. This second condition leads to $C_2 = -\frac{c}{4\nu} R^2$, and hence the final result

$$v(r) = \frac{c}{4\nu} (r^2 - R^2).$$

We observe a quadratically increasing velocity profile reaching the maximum value $-\frac{c}{4\nu} = d_z p R^2 / 4\nu R^2$ on the cylinder axis.

For the discussion of more complex use cases of the Navier-Stokes equation, we refer to the literature.

8.0 Appendix

8.1 Vector calculus in curvilinear coordinates

In analytical electrodynamics we often think about problems with an exceptionally high degree of symmetry — rotational symmetry, axial symmetry, etc. These problems are conveniently formulated in non–cartesian coordinates; To solve them we need to express all analytical operations of vector calculus in such coordinates.

8.1.1 Curvilinear orthogonal systems

Focusing on the case of three dimensional space, let $\mathbf{r}(u_1, u_2, u_3)$ be the parameterization of a point in space in terms of three coordinates u_i . (By way of example, you may think of spherical coordinates, $u_1 = r, u_2 = \theta, u_3 = \phi$. Consider the three vectors $\mathbf{v}_i \equiv \partial_{u_i} \mathbf{r}(u)$. In general, these vectors will not be normalized, so let us consider their normalized descendants $\mathbf{e}_i \equiv g_i^{-1} \mathbf{v}_i$, where $g_i \equiv |\mathbf{v}_i|$. For any point \mathbf{r} , the three vectors $\{\mathbf{e}_i\}$ form a system of coordinate vectors pointing in the direction of the coordinate lines of the coordinate system, i.e. those lines along which two of the coordinates are kept constant (see the figure.)

▷ Example: Specifically, for spherical coordinates, $\mathbf{r} = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta)^T$. It is then straightforward to verify that $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$ form the familiar triple of spherical coordinate vectors and that the normalization factors are given by $g_r = 1, g_\theta = r, g_\phi = r \sin \theta$.

Now, most coordinate systems of practical relevance are **orthogonal** in the sense that $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$, i.e. the coordinate vectors form an orthonormal system. An arbitrary vector field \mathbf{v} can be expanded in this coordinate system as $\mathbf{v}(\mathbf{r}) = \sum_i v_i(\mathbf{r}) \mathbf{e}_i(\mathbf{r})$, where $v_i(\mathbf{r}) = \langle \mathbf{v}(\mathbf{r}), \mathbf{e}_i(\mathbf{r}) \rangle$ and the notation indicates that both the coefficient functions and the coordinate vectors may depend on \mathbf{r} .

8.1.2 Vector differential operations in curvilinear coordinates

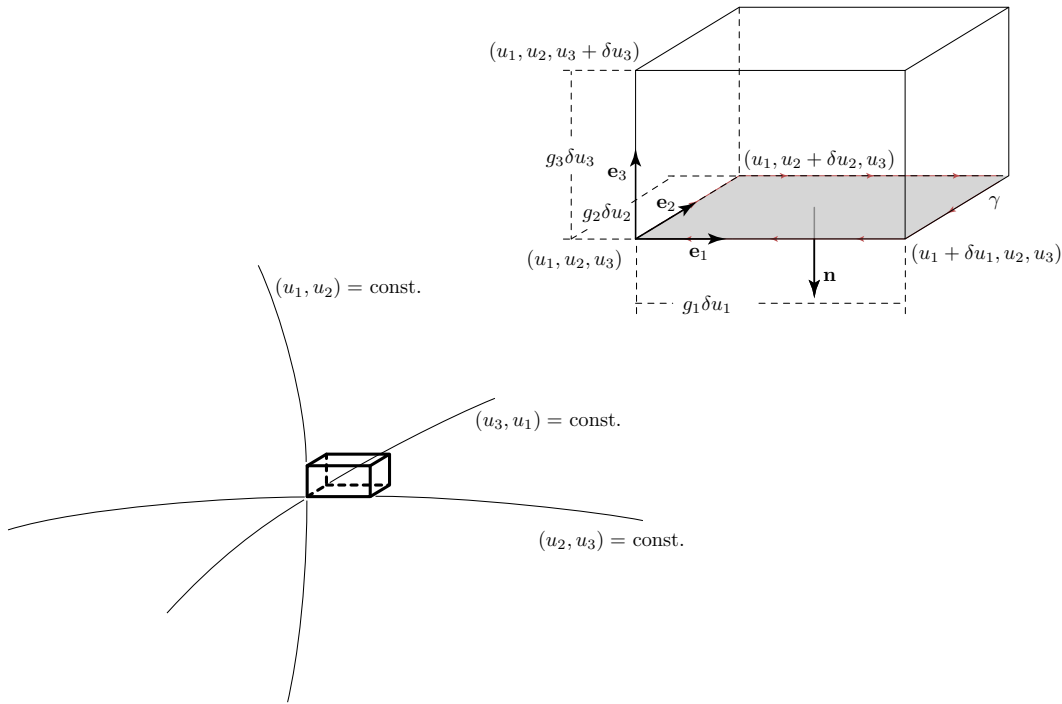


Figure 8.1: On the definition of an infinitesimal reference volume in an orthogonal curvilinear coordinate system.

Gradient

Let $f(\mathbf{r})$ be a differentiable function. The gradient of f , ∇f is a vector that points in the direction of maximal variation of f . Put differently, the of overlap of ∇f with an arbitrary reference vector \mathbf{v} equals the directional derivative of f with respect to \mathbf{v} :

$$\forall \mathbf{v} : \langle \nabla f, \mathbf{v} \rangle \equiv \partial_{\mathbf{v}} f. \tag{8.1}$$

We now use this definition to compute the components of ∇f in an arbitrary orthogonal coordinate system: $\nabla f_i \equiv \langle \nabla f, \mathbf{e}_i \rangle = \partial_{\mathbf{e}_i} f = (g_i)^{-1} \partial_{g_i \mathbf{e}_i} f = (g_i)^{-1} \partial_{\mathbf{v}_i} f$. Now, recall that to compute a directional derivative $\partial_{\mathbf{v}} f$, we need a curve $\gamma(s)$ whose velocity at $s = u$, equals \mathbf{v} : $d_u|_{u=u_0} \gamma(u) = \mathbf{v}(\gamma(u_0))$. Then, $\partial_{\mathbf{v}} f = d_u|_{u=u_0} f(\gamma(u))$. Presently, a curve that does the job is $\mathbf{r}(u_1, u_2, u_3)$: by definition $\partial_{u_i} \mathbf{r} = \mathbf{v}_i$, i.e. $\mathbf{r}(u_1, u_2, u_3)$ is a suitable parameter curve and $\partial_{\mathbf{v}_i} f = \partial_{u_i} f$. Summarizing we have found that $\nabla f_i = g_i \partial_{u_i} f$, or

$$\nabla f = \sum_i \frac{1}{g_i} (\partial_{u_i} f) \mathbf{e}_i. \tag{8.2}$$

▷ Example: For **spherical coordinates**,

$$\nabla f = \partial_r f + \frac{1}{r} \partial_\theta f + \frac{1}{r \sin \theta} \partial_\phi f.$$

Divergence

To obtain generalizations of the cartesian variant $\nabla \cdot \mathbf{v} = \partial_{x_i} v_i$ of the divergence, we first need to give this vector differentiation operation a coordinate-independent meaning. Indeed, the ‘true’ meaning of the divergence — a measure of the source contents of a vector field — is expressed by Gauß’s theorem (2.23). Specifically, for an infinitesimally small reference volume Gauß’s theorem assumes the form

$$\nabla \cdot \mathbf{v} = \frac{1}{V(S)} \int_S d\sigma \mathbf{v} \cdot \mathbf{n},$$

where $V(S)$ denotes the volume of the area enclosed by S and we have assumed continuity of \mathbf{v} , i.e. $\int_{V(S)} d^3x \nabla \cdot \mathbf{v} \simeq V(S) \nabla \cdot \mathbf{v}$. Now, let us assume that $V(S)$ is the cuboid spanned by the three vectors $g_i \mathbf{e}_i \delta u_i$, where the three parameters δu_i are infinitesimal (see Fig. 8.1.) Notice that these vectors describe the change of the position vector \mathbf{r} under changes of the coordinates, e.g. $\mathbf{r}(u_1, u_2, u_3 + \delta u_3) - \mathbf{r}(u_1, u_2, u_3) \simeq \partial_{u_3} \mathbf{r} \delta u_3 = g_3 \mathbf{e}_3 \delta u_3$. Clearly, the volume of $V(S)$ (also denoted by $V(S)$ for brevity) is given by $V(S) = \prod_i g_i \delta u_i$. The integral of \mathbf{v} over the surface S obtains by adding the product of the areas of the faces of $V(S)$ times the components of \mathbf{v} normal to these faces. For example, the area of the bottom face (cf. Fig. 8.1) is given by $\delta u_1 \delta u_2 (g_1 g_2)(u_1, u_2, u_3)$, where the argument (u_1, u_2, u_3) indicates that the metric factors g_i depend on the coordinate data. The unit vector normal to that surface is given by $-\mathbf{e}_3$, i.e. the normal component of \mathbf{v} is $-v_3(u_1, u_2, u_3)$, where v_i are the components in the coordinate system $\{\mathbf{e}_i\}$. The opposite (top) face of the box has an area $\delta u_1 \delta u_2 (g_1 g_2)(u_1, u_2, u_3 + \delta u_3)$, and the normal component of \mathbf{v} is given by $+v_3(u_1, u_2, u_3 + \delta u_3)$. Adding the contribution of the top and the bottom face to the surface integral, we obtain $\delta u_1 \delta u_2 [(g_1 g_2 v_3)(u_1, u_2, u_3 + \delta u_3) - (g_1 g_2 v_3)(u_1, u_2, u_3)] \simeq \delta u_1 \delta u_2 \delta u_3 \partial_{u_3} (g_1 g_2 v_3)$. Adding the contribution of the other faces and dividing by $V(S)$, we thus obtain

$$\nabla \cdot \mathbf{v} = \frac{1}{g_1 g_2 g_3} [\partial_{u_1} (g_2 g_3 v_1) + \partial_{u_2} (g_3 g_1 v_2) + \partial_{u_3} (g_1 g_2 v_3)]. \quad (8.3)$$

For cartesian coordinates this expression reduces to the familiar form given above.

▷ Example: For **spherical coordinates**,

$$(\nabla \cdot \mathbf{v}) = \frac{1}{r^2} \partial_r (r^2 v_r) + \frac{1}{r} \partial_\theta (v_\theta) + \frac{1}{r \sin \theta} \partial_\phi (\sin \theta v_\phi),$$

where $v_r = \mathbf{v} \cdot \mathbf{e}_r$, etc.

Curl

As with Gauß's theorem above, we consider an infinitesimal variant of Stoke's theorem:

$$\mathbf{v} \times \mathbf{n} = \frac{1}{S} \int_{\gamma} d\mathbf{s} \cdot \mathbf{v},$$

where S is the area of an infinitesimal surface element (also denoted by S) bounded by a curve γ and \mathbf{n} is the normal unit vector to S . Specifically, let us consider the example S =(bottom face of the cuboid volume above). Then $S = \delta u_1 \delta u_2 g_1 g_2$ and $\mathbf{n} = -\mathbf{e}_3$. Keeping in mind that the orientation of the contour γ must be chosen so as to conform with the 'right hand rule', the line integral around the rectangular contour surrounding S evaluates to (see the figure)

$$\begin{aligned} \int_{\gamma} d\mathbf{s} \cdot \mathbf{v} &\simeq \delta u_2 (\mathbf{v} \cdot \mathbf{e}_2 g_2)(u_1, u_2, u_3) + \delta u_1 (\mathbf{v} \cdot \mathbf{e}_1 g_1)(u_1, u_2 + \delta u_2, u_3) - \\ &\quad - \delta u_2 (\mathbf{v} \cdot \mathbf{e}_2 g_2)(u_1 + \delta u_1, u_2, u_3) - \delta u_1 (\mathbf{v} \cdot \mathbf{e}_1 g_1)(u_1 + \delta u_1, u_2 + \delta u_2, u_3) \simeq \\ &\simeq -\delta u_1 \delta u_2 (\partial_{u_1}(g_2 v_2) - \partial_{u_2}(g_1 v_1)), \end{aligned}$$

where $v_i = \mathbf{v} \cdot \mathbf{e}_i$ are the components of \mathbf{v} in the coordinate system and the coordinate reference points are evaluated at the corners of the rectangle sitting at the beginning of the outgoing line-stretches. (To first order in the parameters δu_i , the choice of the coordinate convention is immaterial, provided one sticks to a consistent rule.) Substituting this result into the r.h.s. of the infinitesimal variant of Stokes theorem, we obtain $(\nabla \times \mathbf{v})_3 = (g_1 g_2)^{-1} (\partial_{u_1}(g_2 v_2) - \partial_{u_2}(g_1 v_1))$. Analogous results are obtained for the two other components of the curl. Expressed in terms of the Levi-Civita symbol,

$$\boxed{(\nabla \times \mathbf{v})_i = \sum_{jk} \frac{\epsilon_{ijk}}{g_j g_k} \partial_{u_j} (g_k v_k)}. \quad (8.4)$$

For Cartesian coordinates we rediscover the standard formula $(\nabla \times \mathbf{v})_i = \epsilon_{ijk} \partial_j v_k$.

▷ Example: For **spherical coordinates**,

$$\nabla \times \mathbf{v} = \frac{1}{r \sin \theta} [\partial_{\theta}(\sin \theta v_{\phi}) - \partial_{\phi} v_{\theta}] \mathbf{e}_r + \left[\frac{1}{r \sin \theta} \partial_{\phi} v_r - \frac{1}{r} \partial_r (r v_{\phi}) \right] \mathbf{e}_{\theta} + \frac{1}{r} [\partial_r (r v_{\theta}) - \partial_{\theta} v_r] \mathbf{e}_{\phi}.$$

Laplace operator

A generalized expression for the Laplace operator $\Delta = \nabla \cdot \nabla$ is obtained by substituting Eq.(8.2) into Eq.(8.3):

$$\boxed{\Delta = \frac{1}{g_1 g_2 g_3} \left[\partial_{u_1} \frac{g_2 g_3}{g_1} \partial_{u_1} + \partial_{u_2} \frac{g_3 g_1}{g_2} \partial_{u_2} + \partial_{u_3} \frac{g_1 g_2}{g_3} \partial_{u_3} \right]}. \quad (8.5)$$

▷ Example: For **spherical coordinates**,

$$\Delta = \frac{1}{r^2} \partial_r r^2 \partial_r + \frac{1}{r^2 \sin \theta} \partial_\theta \sin \theta \partial_\theta + \frac{1}{r^2 \sin^2 \theta} \partial_\phi^2.$$

For the convenience of the reader, the basic operations of vector calculus in spherical and cylindrical coordinates are summarized in the table below.

Cylindrical (ρ, ϕ, z)	$\nabla f = (\partial_\rho f) \mathbf{e}_\rho + \rho^{-1} (\partial_\phi f) \mathbf{e}_\phi + (\partial_z f) \mathbf{e}_z,$ $\nabla \cdot \mathbf{v} = \rho^{-1} \partial_\rho (\rho v_\rho) + \rho^{-1} \partial_\phi v_\phi + \partial_z v_z,$ $\nabla \times \mathbf{v} = [\rho^{-1} \partial_\phi v_z - \partial_z v_\phi] \mathbf{e}_\rho + [\partial_z v_\rho - \partial_\rho v_z] \mathbf{e}_\phi + \rho^{-1} [\partial_\rho (\rho v_\phi) - \partial_\phi v_\rho] \mathbf{e}_z,$ $\Delta f = \rho^{-1} \partial_\rho \rho \partial_\rho f + \rho^{-2} \partial_\phi^2 f + \partial_z^2 f.$
Spherical (ρ, ϕ, z)	$\nabla f = \partial_r f + r^{-1} \partial_\theta f + (r \sin \theta)^{-1} \partial_\phi f,$ $(\nabla \cdot \mathbf{v}) = r^{-2} \partial_r (r^2 v_r) + (r \sin \theta)^{-1} \partial_\theta (\sin \theta v_\theta) + (r \sin \theta)^{-1} \partial_\phi v_\phi,$ $\nabla \times \mathbf{v} = (r \sin \theta)^{-1} [\partial_\theta (\sin \theta v_\phi) - \partial_\phi v_\theta] \mathbf{e}_r + [(r \sin \theta)^{-1} \partial_\phi v_r - r^{-1} \partial_r (r v_\phi)] \mathbf{e}_\theta +$ $+ r^{-1} [\partial_r (r v_\theta) - \partial_\theta v_r] \mathbf{e}_\phi.$ $\Delta f = r^{-2} \partial_r r^2 \partial_r f + (r^2 \sin \theta)^{-1} \partial_\theta \sin \theta \partial_\theta f + (r^2 \sin^2 \theta)^{-1} \partial_\phi^2 f.$

Table 8.1: The basic operations of vector calculus in spherical and in cylindrical coordinates

Bibliography

- [1] J. D. Jackson. *Classical Electrodynamics*. Wiley, 1975.
- [2] R. Sexl and H. Urbantke. *Relativity, Groups, Particles*. Springer, 2001.