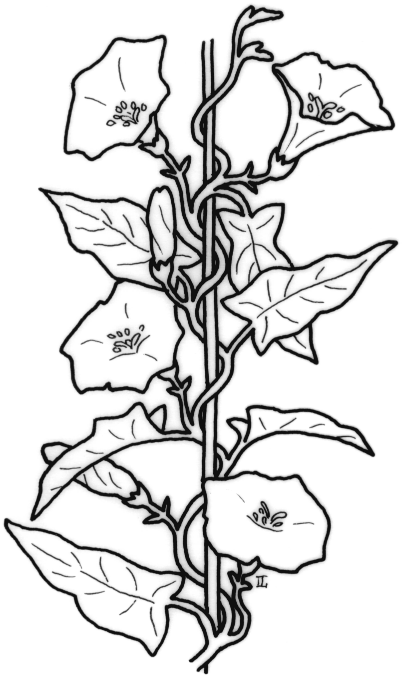
Contact and symplectic geometry in electromagnetism

Matias Dahl

5th September 2002



Convolvulus arvensis

Lonicera sempervirens

Matias Dahl (matias.dahl-at-hut.fi)
Contact and symplectic geometry in electromagnetism

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Espoo, 5th September 2002.

Supervisor: Professor Erkki Somersalo
Instructor: Doctor Kirsi Peltonen

Acknowledgements

First of all, I would like to thank my instructor Doctor Kirsi Peltonen and my supervisor Professor Erkki Somersalo for their expert guidance. I would also like to thank Doctor Perttu Puska, Professor Ismo Lindell, and Doctor Jarmo Malinen for many valuable discussions.

The present work has been carried out at the department of mathematics at the Helsinki University of Technology. For this opportunity, I would like to thank Professor Olavi Nevanlinna.

The cover art is drawn by Inge Löök after [1].

Last, but not least, I would like to express my deepest gratitude to my mother and father for their endless support.

Matias Dahl
5th September 2002

TEKNINSKA HÖGSKOLAN I HELSINGFORS
SAMMANDRAG AV DIPLOMARBETET

Utfört av	Matias Dahl
Avdelning	Elektro- och telekommunikationsteknik
Huvudämne	Elmagnetism (S-96)
Biämne	Matematik (Mat-1)
Arbetets namn	Kontakt och symplektisk geometri i elmagnetism
Title in English	Contact and symplectic geometry in electromagnetism
Professur	Matematik

Sammandrag

Detta arbete är uppdelat i två delar. I den första delen visar vi att genom att betrakta Helmholtz dekompositionen och Bohren dekompositionen i Fourier rymden så kan dessa två kombineras till en dekomposition. Detta ger en matematisk dekomposition som delar upp ett godtyckligt vektorfält i $\mathbb{R}^3$ i tre delar. De viktigaste egenskaperna hos denna dekomposition är att den kommuterar med både curl operatorn och tidsderivator. Därav följer att vi kan andvända dekompositionen för att dela upp Maxwells ekvationer utan att göra antaganden om det underliggande elektromagnetiska ämnet. När vi utför denna dekomposition på de traditionella Maxwells ekvationerna i $\mathbb{R}^3$ (för fälten $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}$) ser vi att ekvationerna delas upp i tre okopplade ekvationssystem. När ett ämne introduceras i dessa ekvationer blir de antingen kopplade eller förblir okopplade beroende på ämnets egenskaper. De uppdelade Maxwells ekvationerna innehåller som ett specialfall Bohren dekompositionen.

I andra delen av detta arbete börjar vi med en introduktion till kontaktgeometri och symplektisk geometri. Genom att studera kända lösningar visar vi att de uppdelade fälten i de uppdelade Maxwells ekvationerna alltid ser ut att ge upphov till kontaktstrukturer. Som ett exempel är de uppdelade fälten för en planvåg dess höger- och vänsterhänt cirkulärt polariserade komponenter. Dessa ger båda upphov till en egen kontaktstruktur. Vi visar också att varje kontaktstruktur ger upphov till sin egen Carnot-Carathéodory metrik. Geodeserna för dessa metriker ser ut att beskriva färdvägen för cirkulärt polariserade vågor.

Antal sidor	103
Nyckelord	kontaktgeometri, symplektisk geometri, elmagnetism, Bohren dekomposition, Moses dekomposition, Fourier transformation

HELSINKI UNIVERSITY OF TECHNOLOGY
ABSTRACT OF MASTER'S THESIS

Author	Matias Dahl
Department	Electrical and communications engineering
Major subject	Electromagnetism (S-96)
Minor subject	Mathematics (Mat-1)

Title	Contact and symplectic geometry in electromagnetism
Title in Swedish	Kontakt och symplektisk geometri i elmagnetism
Chair	Mathematics

Abstract

The present work is divided into two parts. In the first part, we show that, by working in Fourier space, the Bohren decomposition and the Helmholtz's decomposition can be combined into one decomposition. This yields a completely mathematical decomposition, which decomposes an arbitrary vector field on $\mathbb{R}^3$ into three components. The most important properties of this decomposition is that it commutes with both the curl operator and with the time derivative. We can therefore apply this decomposition to Maxwell's equations without assuming anything about the media. When we apply this decomposition to the traditional Maxwell's equations (for $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}$) on $\mathbb{R}^3$, we will see that Maxwell's equations split into three completely uncoupled sets of equations. Also, when a medium is introduced, the decomposed Maxwell's equations either remain uncoupled, or become coupled depending on the complexity of the medium. As a special case, the decomposed Maxwell's equations contain the Bohren decomposition.

In the second part of this work, we begin with an introduction to contact and symplectic geometry and then study their relation to electromagnetism. By studying examples, we show that the decomposed fields in the decomposed Maxwell's equations always seem to induce contact structures. For instance, for a planewave, the decomposed fields are the right and left hand circulary polarized components and each of these induce their own contact structure. Moreover, we show that each contact structure induces its own *Carnot-Carathéodory metric*, and the path traversed by the circularly polarized waves seem to coincide with the geodesics of these metrics.

Number of pages	103
Keywords	contact geometry, symplectic geometry, electromagnetism, Bohren decomposition, Moses decomposition, Fourier transformation

Contents

1 Introduction

One of the main subjects of the present work will be handedness in electromagnetism. In Sections 1.1 and 1.2 we will give a more detailed description of this subject and the present work. However, since handedness is a very general phenomena, we begin this work with an introduction to handedness as a general phenomena in nature.

An object, which is not identical to its mirror image is said to be *handed*. Such an object can always exist in two mirror symmetric forms. For instance, the human hand is handed (with the above meaning), and can therefore exist as a “right” hand or as a “left” hand. Other examples of handed objects are golf clubs and scissors. Also, in nature, plants, snails, and sea-shells typically twist so that they are not identical to their own mirror images and are thus handed. A surprising property in nature is that when an object is handed, then it usually has a *preferred handedness*. For a species, such as snails, this means the following. If, within a species, one form occurs more frequently than its mirror image, then the species has a *preferred handedness*. In the human body, the heart is usually on the left side, and the liver is on the right side. The placements of these organs thus give a preferred handedness to the human body. Most persons are also right-handed (with the traditional meaning), so there is a functional difference between right and left. The front page of this work shows two plants with preferred handedness. The *convolvulus arvensis* (on the left) ordinarily winds to the right, and the *lonicera sempervirens* (on the right) ordinarily winds to the left. Due to mutations, however, there are also specimens that twist in the “wrong” direction. Depending on the species, these can be extremely rare or they can exist in equal numbers as the dominant form [1]. Handedness also exist on the smaller scale in nature. For instance, human smell and taste are handed. In other words, these senses can distinguish molecules from their mirror images. As an example, the molecules giving the smell to oranges and lemons are mirror images of each other, and the molecules giving the taste to dill and peppermint are mirror images of each other [1]. Also, the DNA helix has a preferred handedness [1]. Handedness also occur on the large scale in nature. For instance, hurricanes are handed. Their handedness are determined by the Coriolis force; hurricanes north of the equator twist with opposite handedness than hurricanes south of the equator.

From the above we can make the observation that in nature, right and left are not in symmetrical positions. We can thus say that “nature is handed”, or, in the words of Luis Pasteur, *L’univers est dissymmétrique* [1]. This asymmetry seems to be very general. It exists both on the small molecular scale and on the large scale. One possible explanation for this phenomena is that the weak nuclear force (the weakest of the four natural forces) has a preferred handedness [1]. We can also make another observations, which will be relevant to the present work. Namely, in nature, things (both living an dead) seem to organize themselves in the form of a helix. This helix form can be found in the DNA molecule, in plants, in animals (see e.g. page 8 in [3]), in bathtubs [2], and in hurricanes. One explanation for this organization is that the *helicoid* (which looks like a spiral staircase) is a minimal surfaces in $\mathbb{R}^3$ [4]. One can therefore suspect that the helix configuration is very stable. Since the helix is not mirror symmetric, this organization can be seen as one source of handedness in nature. Another form, which also frequently appears in nature is the spiral. It can be seen as a two dimensional analogue to the helix, and it can be found in galaxies, plants, and Julia sets [3].

1.1 Handed behavior in electromagnetics

In electromagnetism, handedness exists on two levels: handedness in the *electromagnetic fields* and handedness in the *electromagnetic media*. The most simple handed electromagnetic fields are the right and left hand circulary polarized planewaves. These have the forms of helices and are therefore not mirror symmetric. Similarly, *handed media*, or *chiral media*, is media, which is not mirror symmetric to itself. This mirror asymmetry in the media can be either on the microscopic level or on the macroscopic level. On a microscopic level, for instance, quartz is not mirror symmetric to itself. On a macroscopic level, handed media can be made from small helices twisted out of copper wire.

The aim of the present work is to study the handed behavior of electromagnetism. By this we mean the interaction between mirror asymmetry of electromagnetic media and mirror asymmetry of the electromagnetic fields. A typical question, for instance, is to study how a mirror asymmetry in the media affects the fields. Since the electromagnetic fields depend on the media, one could suspect that a mirror asymmetry in the media would give rise to a mirror asymmetry also in the electromagnetic fields and also vice-versa. This argumentation is known as *Curie's principle* (see page 37 of this work), and it is indeed the case. One manifestation of the handed behavior of electromagnetism is that an isotropic homogeneous media is chiral, if and only if right and left hand circulary polarized wave solutions propagate with different phase velocities. This means that in chiral media, right and left hand circulary polarized wave solutions are not mirror images of each other.

Since waves in chiral media can propagate with different phase velocities depending on their polarization, it is clear that chiral media have more degrees of freedom than, say, homogeneous isotropic media. As can be expected, this fact complicates the mathematical analysis of such a media. On the other hand, by properly controlling these additional degrees of freedom, one can construct more intelligent electromagnetic systems. In modern electromagnetism, there are many (more or less equivalent) macroscopic models for chiral media. One such mathematical model is given by the following constitutive equations:

$$\mathbf{D} = \epsilon\mathbf{E} + \xi\mathbf{H}, \tag{1}$$

$$\mathbf{B} = \mu\mathbf{H} + \zeta\mathbf{E}. \tag{2}$$

(In these equations we have used standard notation for the time harmonic electromagnetic fields, and $\epsilon, \mu, \xi, \zeta$ are complex scalars that describe the media [5, 6].) With the above constitutive equations one can, for instance, mathematically show that right and left hand circulary polarized waves can propagate with a different phase velocities. The main disadvantage, however, of the above model for chiral media is that it is algebraic. That is, although equations 1-2 do model chiral media, the equations in themselves do no have a direct geometrical or physical interpretation. For instance, from equations 1-2 we can not directly conclude that they model media where right and left hand circulary polarized waves can propagate with different phase velocities. This means that when we translate our physical model for chiral media into a mathematical one, we loose the geometric insight that we might have about handed behavior, circulary polarized waves, and mirror asymmetry. We already noticed that handed behavior seems to play an important role in all of nature. In particular, we know that handed behavior plays an important role in electromagnetism. With this in view, it is motivated to search for a formalism for electromagnetism, which directly would describe this handed behavior. One of the main results of the present work is to describe one such formalism.

The present work is divided into two parts. In the first part (Sections 2-4), we develop a mathematical decomposition for vector fields on $\mathbb{R}^3$, which we apply to Maxwell's equations. With this decomposition we will see that that the traditional Maxwell's equations split into three completely uncoupled sets of equations. This yield a set of *decomposed Maxwell's equations*, which are completely equivalent to the traditional non-relativistic Maxwell's equations. We can therefore say that these decomposed equations form an alternative formalism for electromagnetism. The main advantage of these decomposed Maxwell's equations is that their fundamental quantities have direct interpretation in terms of handedness. For instance, the decomposed Maxwell's equations contain as a special case the *Bohren decomposition*. The Bohren decomposition, in turn, contains as a special case the decomposition of a plane wave into right and left hand circulary polarized components. However, in sharp contrast to the Bohren decomposition, it is not necessary to introduce any constitutive equations to derive the decomposed Maxwell's equations. For this reason the decomposed Maxwell's equations are valid in arbitrary media.

Sections 2-4 are organized as follows. We begin in Section 2 by defining *helicity* for a vectorfield $\mathbf{F}$ as $\int_{\mathbb{R}^3} \mathbf{F} \cdot \nabla \times \mathbf{F} d\mathbf{x}$. We also show how the quantity $\mathbf{F} \cdot \nabla \times \mathbf{F}$ is related to polarization if $\mathbf{F}$ is a planewave. Then, in Section 3, we define a mathematical decomposition, which decomposes an arbitrary vectorfield into three orthogonal components: one with positive helicity, one with negative helicity, and one with zero helicity. We will call this decomposition the *helicity decomposition*. It is derived as a mathematical generalization of the Bohren decomposition and a refinement of the Helmholtz's decomposition. In Section 4 we introduce Maxwell's equations and show that under this decomposition, they decompose with no assumptions on the media. We also study how the constitutive equations decompose in different media. In particular, for the decomposed Maxwell's equations, we give an alternative mathematical model for chiral media. From this model it is seen directly that right and left hand circulary polarized waves propagate with different phase velocities.

1.2 Geometry of electromagnetism

In modern electromagnetism there are numerous mathematical formalisms for writing down the traditional Maxwell's equations and the constitutive equations. One such formalism is with the use of differential forms. The main advantage of this formalism is that Maxwell's equations become completely topological, i.e., they can be formulated on an arbitrary smooth 3-manifold. In particular, this shows that Maxwell's equations do not rely on an inner product, so by using differential forms, it is possible to remove one unnecessary mathematical structure from the formalism. Another advantage of this formalism is that all the quantities in the formalism divide into forms of different degrees. For instance, the electric field intensity is a 1-form; it only can be integrated over a curve, magnetic flux is a 2-form; it can only be integrated over a surface, and charge density is a 3-form; it can only be integrated over a volume. Since Maxwell's equations for differential forms are purely topological, they are form invariant under any diffeomorphism. This topological character of Maxwell's equations can be understood by writing Maxwell's equations in integral form using Stokes theorem. For instance, from Gauss' law, $dD = \rho$, we obtain that the amount of charge inside a closed surface can be calculated by measuring the electric flux entering and leaving the surface. The other Maxwell's equations have similar interpretations [7].

The main problem, however, with differential forms is that it is not clear how the constitutive equations should be formulated. One approach is to directly translate the constitutive equations from the vector formalism into an equivalent set of equations for differential forms. This can be done by introducing a Riemannian structure, i.e., a Riemannian metric, onto the 3-manifold. If this is done, then the constitutive equations become completely "metrical"; they only depend on the Riemannian metric. There are, however, some problems with this approach. First, if the media possesses electric and magnetic anisotropy, then one needs two Riemannian metrics to model the media [8]; one to describe the electric anisotropy, and one to describe the magnetic anisotropy. Unfortunately, these metrics have no simple geometrical or physical interpretations. One sought feature, for instance, would be that the the geodesics for the metrics would have a physical significance such as describing the path traversed by a ray of light. However, since such a path depends on the polarization of the wave, and since the above metrics do not take polarization into account, the geodesics can not have such properties. A similar approach has also been studied in relativistic electromagnetics [9]. Then one obtains one pseudo-Riemannian metric, but neither this metric seems to have a clear geometrical interpretation. We thus have the same problem as for the chiral constitutive equations in $\mathbb{R}^3$: introducing a Riemannian structure gives a model for media, but the model has no clear interpretation. A second reason why a Riemannian geometry can be questioned is that it can be seen as a generalization of a Cartesian geometry. Cartesian geometry, again, is based on our perception of reality, which is based on electromagnetism. Indeed, if we measure the length of an object with a rigid measuring rod, then the measured length depends on the electromagnetic forces which bind the rod together ([10], page 132). This means that there is no *absolute* way of measuring length without electromagnetism. Therefore, if we base our description of electromagnetic media on a Riemannian structure, then there is a risk for a circular explanation.

The aim of the second part of this work is to try to describe the *internal geometry* of electromagnetism. In other words, the aim is to find a geometric structure, which would describe the geometry of space as an electromagnetic wave would see space. What we here exactly mean by geometry is not clear since there does not seem to exist any such geometric structure for electromagnetism. In this work we will neither present any such canonical geometric structure for electromagnetism. However, we will show ample evidence, which suggests two things. First, in order to study this geometry, one must take into account the handed behavior of electromagnetism. For instance, in the scattering of planewaves one must take into account polarization. For this purpose, the decomposed Maxwell's equations form an ideal framework. Second, we will show that the geometry of electromagnetism seems to be related to contact and symplectic geometry. The aim of the second part of this work (Sections 5-6) will be to introduce these two geometries and study their relation to electromagnetism.

1.3 Contact and symplectic geometry

One can very roughly say that if the fundamental quantity in Riemannian geometry is length, then the fundamental quantity in symplectic geometry is directed area, and the fundamental quantity in contact geometry is a certain twisting behavior. It also holds that contact geometry is always odd dimensional, and symplectic geometry is always even dimensional. What is more, contact geometry and symplectic geometry are dual in the sense that they have many common results, and there are many connections between the two. A characteristic feature for both contact and symplectic geometry is that they have both been found in numerous areas of physics and mathematics.

Examples where contact geometry have been found include optics [10], hydrodynamics [11], knot theory [12], the Monge-Ampere equation [13], Huygens's principle [13], control theory [13], thermodynamics [14], and heat flow [15]. Contact geometric methods have also been applied to real problems. For instance, in molecular biology, the twisting of the DNA molecule has been analyzed using contact geometric methods. Another example is from plasma physics, where the optimum shape for fusion reactors has been analyzed using contact geometric methods [16]. Historically, contact geometry can be traced back to the work of the Norwegian mathematician Sophus Lie. In 1872 he introduced *contact transformations* as a geometric tool to study differential equation systems. The modern usage of contact structures on odd dimensional manifolds began in the 1950s. In Section 5 we will give an introduction to contact geometry, and in Section 6 we show that known solutions to Helmholtz's equation induce contact structures.

The even dimensional analogue theory to contact geometry is symplectic geometry. This geometry has also been found in numerous areas of mathematics and physics. In Section 5.6 we will see how symplectic geometry arises as the natural geometry of phase-space in Hamilton's equations, i.e., the equations of classical mechanics. The coordinates in phase-space are the coordinates of all location and momentum vectors for all particles in the system. One point in phase space thus determines the location and momentum vectors for all particles in the mechanical system. In consequence, a solution to a mechanical system is a path in phase-space. Phase space and its geometry was first studies by Poincaré. The English adverb symplectic comes from the Greek word $\sigma\nu\mu\pi\lambda\epsilon\kappa\tau\iota\kappa us$, which means plaited together, or woven. The corresponding Latin word is complex. It is instructive to think of symplectic geometry as the geometry of the space of location vectors and momentum vectors, which have been "plaited together" into phase space. More generally one can think of symplectic geometry as a something obtained by "plating together" mathematics and physics [17].

By treating mechanical problems in phase-space, it is typical that the problems usually becomes much more simpler. For example, knowledge of this internal geometry, or structure, of phase-space has led to completely new methods for numerically solving Hamilton's equations. As an example, these methods have been used to numerically simulate the outer solar system for one billion years [18]. Unfortunately, the resulting numerical solution is not quantitatively correct. For instance, the precise angles of the planets can not be predicted. However, from the solution, it is possible to make quantitative predictions. This is quite remarkable when one takes into account that the system is chaotic. Compared to other methods for simulating the solar system, the symplectic method also has other advantages. It is fast, elegant, simple, and has a small error [18]. A similar "symplectic FDTD method" has been studied in electromagnetics [19]. Symplectic geometry also arises in geometric optics [10]. In microlocal analysis, symplectic geometry is also used to generalize Fourier integral operators to manifolds [20, 21].

In Section 4.4.1 we derive a generalization of the *dual transformation* [5] for the decomposed Maxwell's equations. This transformations has 16 degrees of freedom, and in Section 4.4.1 we show that this transformation is related to symplectic geometry.

In the last section of this work (Section 7) we draw the conclusions and give some suggestions for further work.

2 Helicity and Beltrami fields

In Sections 2-4 of this work we will work with possibly complex valued *vector fields*. These are vector fields defined on an open simply connected set $\Omega \subset \mathbb{R}^3$ with possibly complex component functions. If Ω has a boundary, we also assume that the boundary is smooth. In Sections 2-4 of this work we shall only work with Cartesian coordinates. We further assume that the component functions of all vector fields are Lebesgue measurable functions $\Omega \to \mathbb{C}$. The *Lebesgue integral* of a measurable function $f : \Omega \to \mathbb{C}$ is denoted by $\int_\Omega f(\mathbf{x})d\mathbf{x}$. Similarly, the integral of a vector field $\mathbf{F}$ is defined componentwise, and is denoted by $\int_\Omega \mathbf{F}(\mathbf{x})d\mathbf{x}$. In this work, $i = \sqrt{-1}$ is the complex unit, and $\Re\{x\}$ and $\Im\{x\}$ are the real and imaginary parts of a complex number (or vector) x. All vectors will be written in boldface.

2.1 Helicity

In this section we will see that helicity is a scalar associated with a vector field that measures how much the field rotates, or twists, about itself. If the field has only right-hand rotation, its helicity is positive, and if it has only left-hand rotation, its helicity is negative. Further, if the field has both right-hand and left-hand rotation, its helicity can be positive, negative, or zero.

Definition 2.1 (Helicity) *Let $\boldsymbol{F}$ and $\boldsymbol{G}$ be vector fields on a simply connected open set $\Omega \subset \mathbb{R}^3$. The* helicity *of $\boldsymbol{F}$ is the real number defined as*

$$\mathscr{H}(\boldsymbol{F}) = \int_\Omega \boldsymbol{F} \cdot \nabla \times \boldsymbol{F}\, d\boldsymbol{x}. \tag{3}$$

We will also say that $\boldsymbol{F} \cdot \nabla \times \boldsymbol{F}$ is the helicity density *of $\boldsymbol{F}$. The* cross-helicity *of $\boldsymbol{F}$ and $\boldsymbol{G}$ is defined as*

$$\mathscr{H}(\boldsymbol{F}, \boldsymbol{G}) = \int_\Omega \boldsymbol{F} \cdot \nabla \times \boldsymbol{G}\, d\boldsymbol{x}. \tag{4}$$

In the above definition we have not defined the precise function space in which $\mathbf{F}$ and $\mathbf{G}$ should be members. However, the aim of the present section is only to give a short heuristic introduction to helicity. For instance, in Example 2.2 we show how helicity is related to polarization. We therefore postphone the definition of this function space to Section 3. In the present section, we therefore tacitly assume that all objects are sufficiently smooth and well behaved so that all derivatives and integrals are well behaved.

It should also be pointed out that the above definition of helicity is non-standard [23, 24]. In fluid mechanics and plasma physics, the helicity of a divergence-free vector field $\mathbf{F}$ is defined as

$$\mathscr{H}_{\text{fluid}}(\mathbf{F}) = \int_\Omega \mathbf{F} \cdot \nabla \times^{-1} \mathbf{F}\, d\mathbf{x}.$$

If $\mathbf{F}$ vanishes on the boundary of Ω, then $\mathscr{H}_{\text{fluid}}$ does not depend on the curl-free term of $\nabla\times^{-1}$ (see e.g. [23] page 121). Therefore $\mathscr{H}_{\text{fluid}}$ is well defined. Since $\mathscr{H}_{\text{fluid}}(\nabla \times \mathbf{F}) = \mathscr{H}(\mathbf{F})$, the two definitions are closely related. The non-standard definition of helicity is motivated for two reasons. First, we will calculate helicity for vector fields with non-zero divergence. For such fields, the inverse of the $\nabla\times$ operator (and hence $\mathscr{H}_{\text{fluid}}$) is not well defined. Another motivation for Definition 2.1 will be given in Section 5.3, where we show how helicity (as defined in Definition 2.1) is related to contact structures.

We next show how helicity is related to the polarization of planewaves in electromagnetics. A *time harmonic plane-wave* is a real valued vector field in $\mathbb{R}^3$ that can be written as

$$\mathbf{F}(z,t) \quad = \quad \Re\left\{\mathbf{A}e^{i(kz-\omega t)}\right\}$$

for some Cartesian coordinates x, y, z, some positive real numbers k, ω, and a complex constant vector $\mathbf{A}$, which has no z-component. For this wave, the z-axis is the *direction of propagation*, and t is the time parameter. The vector $\mathbf{A}$ determines the *polarization* of the wave. The *axes of rotation* of $\mathbf{F}$ are $\Re\{\mathbf{A}\}$ and $\Im\{\mathbf{A}\}$. If $\Re\{\mathbf{A}\}\times\Im\{\mathbf{A}\} = 0$, then $\mathbf{F}$ is *linearly polarized*. If $|\Re\{\mathbf{A}\}|^2 = |\Im\{\mathbf{A}\}|^2$ and $\Re\{\mathbf{A}\}\cdot\Im\{\mathbf{A}\} = 0$, then $\mathbf{F}$ is *circulary polarized*. If $\mathbf{F}$ is neither linearly polarized nor circulary polarized, then $\mathbf{F}$ is *elliptically polarized*. Circulary polarized waves are further handed. We say that a circulary polarized plane-wave is *right-hand polarized*, if its helicity density is negative, and *left-hand polarized*, if its helicity density is positive. This definition is motivated by the next example.

Example 2.2 (Helicity and polarization) Let us define

$$\begin{aligned}\mathbf{E}_\pm(z,t) &= \Re\left\{(\mathbf{u}_x \pm i\mathbf{u}_y)e^{i(kz-\omega t)}\right\} \\ &= \cos(kz-\omega t)\mathbf{u}_x \mp \sin(kz-\omega t)\mathbf{u}_y.\end{aligned}$$

For these fields, $\nabla \times \mathbf{E}_\pm = \pm k\mathbf{E}_\pm$. In other words, the fields are parallel and anti-parallel to their own curl. Hence the helicity densities of $\mathbf{E}_+$ and $\mathbf{E}_-$ are positive, respectively negative, so $\mathbf{E}_+$ is left-hand circulary polarized and $\mathbf{E}_-$ is right-hand circulary polarized. Figure 1 shows these fields for $t = 0$: $\mathbf{E}_-$ rotates around the positive z-axis using the "right-hand rule", and $\mathbf{E}_+$ rotates around the positive z-axis using the "left-hand rule".

Adding the above fields yields a linearly polarized planewave $\mathbf{E} = \mathbf{E}_+ + \mathbf{E}_- = 2\cos(kz - \omega t)\mathbf{u}_x$. Since $\nabla \times \mathbf{E} = k(\mathbf{E}_+ - \mathbf{E}_-) = -2k\sin(kz - \omega t)\mathbf{u}_y$, it follows that $\mathscr{H}(\mathbf{E}) = 0$. This shows that a linearly polarized planewave carries no helicity, but it can be decomposed into two plane-waves with positive/negative helicity. $\square$

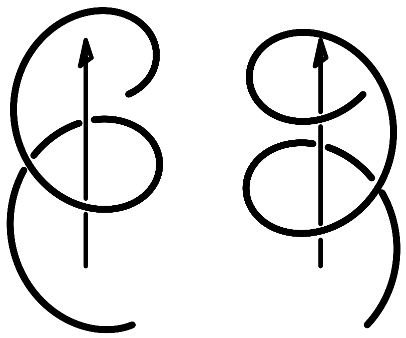


Figure 1: $\mathbf{E}_+$ and $\mathbf{E}_-$ in Example 2.2.

The next lemma shows the basic properties of helicity. Part *b)* shows that $\mathscr{H}$ is a non-linear functional.

Lemma 2.3 *Let $\boldsymbol{F}$ and $\boldsymbol{G}$ be C^∞ vector-fields on a simply connected open set $\Omega \subset \mathbb{R}^3$ with a smooth boundary such that both vector fields vanish on the boundary of Ω. Then*

a) $\mathscr{H}(\boldsymbol{F}, \boldsymbol{G}) = \mathscr{H}(\boldsymbol{G}, \boldsymbol{F})$.

b) For all α, β in $\mathbb{R}$,

$$\mathscr{H}(\alpha\boldsymbol{F} + \beta\boldsymbol{G}) = \alpha^2\mathscr{H}(\boldsymbol{F}) + 2\alpha\beta\mathscr{H}(\boldsymbol{F}, \boldsymbol{G}) + \beta^2\mathscr{H}(\boldsymbol{G}).$$

Proof. Since the fields vanish on the boundary of Ω, part *a)* follows from the identity $\nabla \cdot \big(\mathbf{F} \times \mathbf{G}\big) = \nabla \times \mathbf{F} \cdot \mathbf{G} - \mathbf{F} \cdot \nabla \times \mathbf{G}$ using Stokes theorem. Part *b)* follows from part *a)*. $\square$

2.2 Beltrami fields

Helicity is closely related to *Beltrami fields*. These are vector fields $\mathbf{F} : \Omega \to \mathbb{R}^3$ in a simply connected open set $\Omega \subset \mathbb{R}^3$ that satisfy

$$\nabla \times \mathbf{F} = f\mathbf{F}$$

for some function $f : \Omega \to \mathbb{R}$. (Here, again, we assume that all objects are sufficiently smooth.) Geometrically, the above equation states that the rotation of $\mathbf{F}$ is everywhere parallel to the field. A characteristic feature for such fields is a constant twisting of the field. If $f > 0$, the field has positive helicity, and if $f < 0$, the field has negative helicity. If f is constant, then the field is said to be a *Trkalian field*. The fields $\mathbf{E}_+$ and $\mathbf{E}_-$ in Example 2.2 are Trkalian fields. Trkalian fields on $\mathbb{R}^3$ are classified in [25].

Beltrami fields appear in surprisingly many areas of physics. In plasma physics Beltrami fields are also called *force free fields*. For instance, the magnetic field inside ball lightnings and fusion reactors have been modeled by Beltrami fields [26, 24]. Also, in fluid mechanics, the motion of particles in tornadoes and waterspouts have be modeled by Beltrami fields [27]. Beltrami fields also appear in gravitation research, quark physics and thermoacoustics [27]. In electromagnetics, Beltrami fields are also called *wave fields* [5, 6, 28]. In Section 3.1.2 we will show that solutions to Helmholtz's equation $\nabla \times (\nabla \times \mathbf{E}) = k^2\mathbf{E}$ can be written as the sum of two Beltrami fields. In electromagnetics this decomposition is known as the Bohren decomposition.

In [11] it is shown that there is a one-to-one correspondence (up to a scaling) between non-vanishing Beltrami fields and contact structures. In Section 6 we will use this result to derive contact structures from electromagnetic field solutions.

3 Helicity decomposition of vector fields in $\mathbb{R}^3$

The *Bohren decomposition* is a very useful tool in electromagnetics and it has been studied in numerous references (e.g [5, 6, 27, 29, 30]). Essentially, it takes a solution to Helmholtz's equation $\nabla \times (\nabla \times \mathbf{E}) = k^2\mathbf{E}$ and decomposes it as $\mathbf{E} = \mathbf{E}_+ + \mathbf{E}_-$ such that $\nabla \times \mathbf{E}_\pm = \pm k\mathbf{E}_\pm$. In other words, the Bohren decomposition takes a solution to the (second order) Helmholtz's equation and decomposes it into a sum of two solutions to the (first order) Beltrami equation. This decomposition has two drawbacks. First, it is a physical decomposition, i.e., it can only be applied to solutions to Helmholtz's equations. Second, in order to form Helmholtz's equation from the most general Maxwell's equations, it is necessary to make quite strong assumptions on the media.

In this section we derive a generalization of the Bohren decomposition, which we will call the *helicity decomposition*. This decomposition decomposes an *arbitrary* vector field on $\mathbb{R}^3$ into three components: one with positive helicity, one with negative helicity, and one with zero helicity. To motivate its definition, we shall use the Helmholtz's decomposition and the Bohren decomposition as models. By considering these decompositions in Fourier space, we can combine them into one mathematical decomposition, which will be the helicity decomposition. This derivation, or motivation, is done in Section 3.1. Since the aim of that section is only to *motivate* the helicity decomposition, the approach is completely formal. In Section 3.2, the helicity decomposition is rigorously defined and its basic properties are derived. Further, in Section 3.3 we show how the decomposition behaves for time dependent fields, and in Section 3.4 we show how the helicity decomposition is related to the *Moses decomposition* presented in [31]. This decomposition is well-known in fluid mechanics [31, 32, 33, 34, 35]. Mathematically, the helicity decomposition can also be seen as a *Krein decomposition* with respect of helicity [36].

From a physical standpoint, the main results of this section are that the helicity decomposition commutes with the curl operator and with the time derivative. Thus the helicity decomposition has all the properties necessary to decompose Maxwell's equations in arbitrary media (see Section 4). It should, however, be emphasized that the helicity decomposition is a mathematical decomposition. As such, it can be applied to any vector field.

3.1 Motivation

We begin by motivating the definition of the helicity decomposition. For this purpose it is not necessary to precisely define the function spaces on which we work. We only assume that the underlying space is $\mathbb{R}^3$ with Cartesian coordinates. On this space, we study possibly complex valued vector fields. We will also need the Fourier transform, which we, at this point, only define formally. For a vector field $\mathbf{F}$, we define its *Fourier transform* by

$$\hat{\mathbf{F}}(\mathbf{k}) \quad = \quad \int_{\mathbb{R}^3} \mathbf{F}(\mathbf{x}) e^{-2\pi i \mathbf{k}\cdot\mathbf{x}} d\mathbf{x}. \tag{5}$$

Given the transformed field $\hat{\mathbf{F}}$, the original field is formally given by

$$\mathbf{F}(\mathbf{x}) \quad = \quad \int_{\mathbb{R}^3} \hat{\mathbf{F}}(\mathbf{k}) e^{2\pi i \mathbf{k}\cdot\mathbf{x}} d\mathbf{k}. \tag{6}$$

We will refer to $\mathbf{k}$-space as the *Fourier space* and to $\mathbf{x}$-space as the *physical space*.

3.1.1 Helmholtz's decomposition

Helmholtz's decomposition states that a vector field can be decomposed into two parts; one part whose curl vanishes and another part whose divergence vanishes. In this section we derive this result in Fourier space. Moreover, in Fourier space we give a simple geometric interpretation of Helmholtz's decomposition.

We begin by considering a vector field $\mathbf{F}$ as written in equation 6. We further assume that $\int_{\mathbb{R}^3} \mathbf{F}(\mathbf{x}) d\mathbf{x} = 0$, so that $\hat{\mathbf{F}}$ does not have a singularity (in the sense of a δ-peak) at $\mathbf{k} = 0$. We can therefore remove this point from Fourier space without modifying $\mathbf{F}$ and write

$$\mathbf{F}(\mathbf{x}) \quad = \quad \int_{\mathbb{R}^3\setminus\{0\}} \hat{\mathbf{F}}(\mathbf{k}) e^{2\pi i \mathbf{k}\cdot\mathbf{x}} d\mathbf{k}.$$

Since $\mathbf{k} = 0$ is excluded, the sphere $|\mathbf{k}| =$ constant is well-defined. We can therefore decompose $\hat{\mathbf{F}}$ as $\hat{\mathbf{F}} = \hat{\mathbf{F}}_t + \hat{\mathbf{F}}_n$ such that $\hat{\mathbf{F}}_t$ is tangential to the $\mathbf{k}$-sphere, and $\mathbf{F}_n$ is normal to the $\mathbf{k}$-sphere. With dyadic algebra (see Appendix A),

$$\begin{aligned} \hat{\mathbf{F}}_n(\mathbf{k}) &= \mathbf{u}_r \mathbf{u}_r \cdot \hat{\mathbf{F}}(\mathbf{k}), \\ \hat{\mathbf{F}}_t(\mathbf{k}) &= \left(\bar{\bar{\mathbf{I}}} - \mathbf{u}_r \mathbf{u}_r\right) \cdot \hat{\mathbf{F}}(\mathbf{k}), \end{aligned}$$

where $\mathbf{u}_r = \frac{\mathbf{k}}{|\mathbf{k}|}$. Since the Fourier transform is one-to-one (when defined between suitable function spaces), the above decomposition induces a unique decomposition also in physical space. Namely,

$$\begin{aligned} \mathbf{F}(\mathbf{x}) &= \int_{\mathbb{R}^3\setminus\{0\}} \hat{\mathbf{F}}_n(\mathbf{k}) e^{2\pi i \mathbf{k}\cdot\mathbf{x}} d\mathbf{k} + \int_{\mathbb{R}^3\setminus\{0\}} \hat{\mathbf{F}}_t(\mathbf{k}) e^{2\pi i \mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \\ &= \mathbf{F}_n(\mathbf{x}) + \mathbf{F}_t(\mathbf{x}), \end{aligned} \tag{7}$$

where $\mathbf{F}_n(\mathbf{x})$ and $\mathbf{F}_t(\mathbf{x})$ are defined as the corresponding terms on the first line.

We next show that the decomposition $\mathbf{F} = \mathbf{F}_n + \mathbf{F}_t$ is the Helmholtz's decomposition into divergence- and curl-free components. This follows from the identities

$$\begin{aligned} \nabla \times (\mathbf{v}e^{i\mathbf{k}\cdot\mathbf{x}}) &= i\mathbf{k} \times \mathbf{v}e^{i\mathbf{k}\cdot\mathbf{x}}, \\ \nabla \cdot (\mathbf{v}e^{i\mathbf{k}\cdot\mathbf{x}}) &= i\mathbf{k} \cdot \mathbf{v}e^{i\mathbf{k}\cdot\mathbf{x}}. \end{aligned}$$

In these, $\nabla\times$ and $\nabla\cdot$ are taken with respect of $\mathbf{x}$, and $\mathbf{v}$ is a vector independent of $\mathbf{x}$. Since our approach is completely formal, we can assume that $\nabla\times$ and $\nabla\cdot$ can be taken in under the integral signs in $\mathbf{F}_n$ and $\mathbf{F}_t$. Then, since $\hat{\mathbf{F}}_n$ is normal to the $|\mathbf{k}|$-sphere, $\nabla \times \mathbf{F}_n = 0$, i.e., $\mathbf{F}_n$ is *curl-free*. Similarly, since $\hat{\mathbf{F}}_t$ is tangential to the $|\mathbf{k}|$-sphere, $\nabla \cdot \mathbf{F}_t = 0$, i.e., $\mathbf{F}_t$ is *divergence-free*.

The geometric interpretation in Fourier space is now as follows. In Fourier space the divergence-free component is tangential to the **k**-sphere and the curl-free component is normal to the **k**-sphere.

3.1.2 The Bohren decomposition

In this section we first derive the Bohren decomposition using methods from functional analysis. After that we consider the decomposition in Fourier space and see that it can be combined with the Helmholtz's decomposition to yield a mathematical decomposition.

To derive the traditional Bohren decomposition, we begin with Helmholtz's equations for the electric field $\mathbf{E}$,

$$\nabla \times (\nabla \times \mathbf{E}) = k^2\mathbf{E}. \tag{8}$$

Here k is a real scalar depending of the electrical properties of the medium and the frequency of the wave. In order to form Helmholtz's equations, we have also assumed that the fields in Maxwell's equations are time harmonic. Hence $\mathbf{E}$ is a complex valued vector field.

Helmholtz's equations can be read as an equation for the operator $\frac{1}{k}\nabla\times$. In fact, Helmholtz's equations states that the square of this operator is the identity operator, but only when applied to $\mathbf{E}$. In general, a linear operator $L : X \to X$ on a vector space X for which L^2 is the identity operator is an *involution*. For involutions, we have the following theorem from functional analysis.

Proposition 3.1 *Let X be a vector space where I is the identity operator and $L : X \to X$ is a linear involution. Then L induces two operators $P_\pm = \frac{1}{2}\big(I \pm L\big)$*

for which

$$\begin{aligned} P_+ + P_- &= I, \\ P_\lambda P_\kappa &= \delta_{\lambda\kappa} P_\lambda, \\ L P_\lambda &= \lambda P_\lambda, \end{aligned}$$

where λ and κ are in ± 1.

Here $\delta_{\lambda\kappa}$ is the *Kronecker delta symbol*: $\delta_{\lambda\kappa} = 0$ if $\lambda \neq \kappa$ and $\delta_{\lambda\kappa} = 1$ if $\lambda = \kappa$.

Proof. The first claim is seen from the definition of $P_\pm$. Using $L^2 = I$, we have

$$\begin{aligned} P_\lambda P_\kappa &= \frac{1}{4}(I + \lambda L)(I + \kappa L) \\ &= \frac{1}{4}\big((1 + \lambda\kappa)I + (\lambda + \kappa)L\big), \end{aligned}$$

from which the second claim follows. The third claim follows from $LP_\lambda = \frac{1}{2}\big(L + \lambda I\big) = \lambda P_\lambda$. □

Inspecting the above proof shows that it only relies on the algebraic property $L^2 = I$. We can therefore append E to the right side of all formulas in the above proposition and obtain the following variant:

Proposition 3.2 *Let X be a vector space where I is the identity operator. Furthermore, let $L : X \to X$ be a linear operator such that $L^2 E = E$ for some E in X. Then E decomposes into $E_\pm = \frac{1}{2}\big(E \pm LE\big)$ and*

$$\begin{aligned} E_+ + E_- &= E, \\ P_\lambda E_\kappa &= \delta_{\lambda\kappa} E_\lambda, \\ L E_\lambda &= \lambda E_\lambda, \end{aligned}$$

where λ and κ are in ± 1 and $P_\lambda = \frac{1}{2}(I + \lambda L)$.

Applying Proposition 3.2 to Helmholtz's equation gives the decomposition $\mathbf{E} = \mathbf{E}_+ + \mathbf{E}_-$ with

$$\mathbf{E}_\pm = \frac{1}{2}\big(\mathbf{E} \pm \frac{1}{k}\nabla \times \mathbf{E}\big) \tag{9}$$

and $\nabla \times \mathbf{E}_\pm = \pm k \mathbf{E}_\pm$. In electromagnetic theory, this decomposition is the *Bohren decomposition*. Next we show that the Bohren decomposition can be generalized to an arbitrary vector field; the assumption that $\mathbf{E}$ is a solution to Helmholtz's equation is unnecessary.

In the previous section, we saw that the Helmholtz's decomposition had a simple interpretation in Fourier space. It is therefore motivated to study also the Bohren decomposition in Fourier space. Since $\mathbf{E}$ is divergence-free (as a solution to Helmholtz's equation), $\hat{\mathbf{E}} = \hat{\mathbf{E}}_t$, so for the Fourier transformed fields $\hat{\mathbf{E}}_\pm$, equation 9 reads

$$\hat{\mathbf{E}}_\pm \;=\; \frac{1}{2}\big(\bar{\bar{\mathbf{I}}} \pm 2\pi i \frac{1}{k}\mathbf{k} \times \bar{\bar{\mathbf{I}}}\big) \cdot \hat{\mathbf{E}}_t.$$

Here $\mathbf{k}$ is the Fourier variable, and k defines the physical properties of the medium. We next show that $\hat{\mathbf{E}}_t$ is non-zero only when $k = 2\pi|\mathbf{k}|$. We can then simplify the above equation. Fourier transforming Helmholtz's equation yields $(2\pi)^2 i\mathbf{k} \times (i\mathbf{k} \times \hat{\mathbf{E}}_t) = k^2\hat{\mathbf{E}}_t$. By the identity $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$, it follows that $(2\pi)^2|\mathbf{k}|^2\hat{\mathbf{E}}_t = k^2\hat{\mathbf{E}}_t$, so either $\hat{\mathbf{E}}_t(\mathbf{k})$ is zero, or $(2\pi)^2|\mathbf{k}|^2 = k^2$. Hence $\hat{\mathbf{E}}_t$ can only be non-zero on the sphere $k = 2\pi|\mathbf{k}|$ and

$$\hat{\mathbf{E}}_\pm \;=\; \frac{1}{2}\big(\bar{\bar{\mathbf{I}}} \pm i\mathbf{u}_r \times \bar{\bar{\mathbf{I}}}\big) \cdot \hat{\mathbf{E}}_t. \tag{10}$$

Since the origin is excluded from Fourier space, $\hat{\mathbf{E}}_\pm$ are well-defined. Here we note that if $\hat{\mathbf{E}}_t$ is indeed non-zero only on the sphere $|\mathbf{k}| = k$, then $\mathbf{E}$ would be identically zero since the Lebesgue measure of this sphere (in $\mathbb{R}^3$) is zero. Therefore, in the above reasoning, $\hat{\mathbf{E}}$ should formally be interpreted as a distribution.

From equation 10, we see that the decomposition $\hat{\mathbf{E}}_t = \hat{\mathbf{E}}_+ + \hat{\mathbf{E}}_-$ corresponds to multiplying $\hat{\mathbf{E}}_t$ in Fourier space by $\frac{1}{2}\big(\bar{\bar{\mathbf{I}}} \pm i\mathbf{u}_r \times \bar{\bar{\mathbf{I}}}\big)$. The key observation is that this operation is independent of k, i.e., the free parameter in Helmholtz's equation from which we started. Since the decomposition is independent of this parameters, we can drop the assumption that $\mathbf{E}$ is a solution to Helmholtz's equation. To decompose $\mathbf{E}$, we only need to assume that $\mathbf{E}$ is a divergence-free vector field, and that $\mathbf{k} = 0$ can be removed from Fourier space without changing $\mathbf{E}$. Under these assumptions, we can combine the above decomposition with Helmholtz's decomposition and obtain the decomposition $\mathbf{E} = \mathbf{E}_n + \mathbf{E}_+ + \mathbf{E}_-$ in physical space. This decomposition is the *helicity decomposition*.

In the next section, we properly define the helicity decomposition for *square integrable vector fields*. For such, any set with zero measure in Fourier space can be removed without modifying the physical field. It is therefore clear that one can remove the point $\mathbf{k} = 0$ without modifying the physical field. Formally, this point is both tangential and normal to the $|\mathbf{k}|$-sphere. In *Hodge's decomposition* (the generalization of Helmholtz's decomposition to manifolds [37]), this component is the *harmonic component*, which is both curl-free and divergence-free. In $\mathbb{R}^3$, this is the constant component of a vector field. In [38], the helicity decomposition is developed on the 3-torus. On this space, the harmonic component must first be removed in order to decompose a field.

3.2 The helicity decomposition

In this section we properly define the helicity decomposition and derive its basic properties. For this reason, we first define the Fourier transform more precisely. Since we will apply the decomposition to Maxwell's equations, the natural space to operate on is L^2. This space consists of possibly complex valued vector fields $\mathbf{F}$ for which $\int_{\mathbb{R}^3} \mathbf{F} \cdot \mathbf{F}^* d\mathbf{x} < \infty$. We then also write $\mathbf{F} \in L^2$. Here, $\mathbf{F}^*$ is the complex conjugate of $\mathbf{F}$. With the inner product $(\mathbf{F}, \mathbf{G}) = \int_{\mathbb{R}^3} \mathbf{F} \cdot \mathbf{G}^* d\mathbf{x}$, L^2 is a Banach space [41] with the norm $||\mathbf{F}|| = \sqrt{(\mathbf{F}, \mathbf{F})}$. Similarly, a vector field in L^1 is a vector field, whose all component functions are L^1 functions, i.e., such functions $f : \mathbb{R}^3 \to \mathbb{C}$ for which $\int_{\mathbb{R}^3} |f(\mathbf{x})| d\mathbf{x} < \infty$.

For L^1 vector fields, the Fourier transform and its inverse are well-defined and given by equations 5 and 6 [41]. However, for vector fields in L^2, these equations do not define a transform. For instance, if $\mathbf{F}$ is an L^2 vector field, then Equation 5 does not necessarily converge. Next we define (following [42], pages 16-18) the L^2 Fourier transform as a limit of the L^1 Fourier transform. For this we will need some results, which are collected in the next lemma.

Lemma 3.3 *[42]*

a) The space $L^1 \cap L^2$ is dense in L^2.

b) If $\boldsymbol{F} \in L^1 \cap L^2$, then the L^1 Fourier transform $\hat{\boldsymbol{F}}$ is an element in L^2.

c) If $\boldsymbol{F}$ and $\boldsymbol{G}$ are vector fields in $L^1 \cap L^2$, then $(\hat{\boldsymbol{F}}, \hat{\boldsymbol{G}}) = (\boldsymbol{F}, \boldsymbol{G})$, where $\hat{\boldsymbol{F}}$ and $\hat{\boldsymbol{G}}$ are the L^1 Fourier transforms of $\boldsymbol{F}$ and $\boldsymbol{G}$.

If $\mathbf{F} \in L^2$, then its Fourier transform is defined as follows. Since $L^1 \cap L^2$ is dense in L^2, there exists a sequence $\mathbf{F}_n \in L^1 \cap L^2$, such that $\mathbf{F}_n \to \mathbf{F}$ (in the L^2 norm) as $n \to \infty$. For each $\mathbf{F}_n$, the L^1 Fourier transform $\hat{\mathbf{F}}_n$ is well-defined. Further, since a sequence converges if and only if it is a Cauchy sequence, it follows from property *b)* and *c)* that $\hat{\mathbf{F}}_n$ is a converging sequence in L^2. Since L^2 is closed [41], it follows that $\hat{\mathbf{F}}_n$ converges to an element in L^2, say $\hat{\mathbf{F}}$, which is uniquely determined by $\mathbf{F}$: If $\mathbf{F}'_n$ is another sequence converging (in the L^2 norm) to $\mathbf{F}$, then $||\hat{\mathbf{F}}_n - \hat{\mathbf{F}}'_n|| = ||\mathbf{F}_n - \mathbf{F}'_n|| \to 0$ by part *c)*. We then define $\hat{\mathbf{F}}$ to be the L^2 Fourier transform of $\mathbf{F}$. We shall use the same notation for both the L^1 and the L^2 Fourier transform. If $\mathbf{F}$ is a vector field in L^1 or L^2, then both its L^1 and L^2 Fourier transform are written as $\hat{\mathbf{F}} = \mathscr{F}\mathbf{F}$. From the context it will always be clear which transform is meant. We will also see that on $L^1 \cap L^2$, the two transforms coincide.

If $\mathbf{F} \in L^2$, then one can choose the sequence in $L^1 \cap L^2$ as $\mathbf{F}_n(\mathbf{x}) = \chi_n(\mathbf{x})\mathbf{F}(\mathbf{x})$, where $\chi_n(\mathbf{x}) = 0$, when $|\mathbf{x}| \geq n$, and $\chi_n(\mathbf{x}) = 1$, when $|\mathbf{x}| < n$ [42]. Then

$$\mathscr{F}\{\mathbf{F}(\mathbf{x})\}(\mathbf{k}) = \lim_{n\to\infty} \int_{\mathbb{R}^3} \chi_n(\mathbf{x})\mathbf{F}(\mathbf{x})e^{-2\pi i\mathbf{k}\cdot\mathbf{x}}d\mathbf{x}, \tag{11}$$

where $\lim$ denotes the limit in the L^2 norm. By *Lebesgue's dominated convergence theorem* [41], it follows that the restriction of the L^2 Fourier transform to L^1 vector fields is the L^1 Fourier transform.

Since the L^2 inner product is continuous [39], part *b)* in Lemma 3.3 also holds in L^2: If $\mathbf{F}, \mathbf{G} \in L^2$, then $(\mathbf{F}, \mathbf{G}) = (\mathscr{F}\mathbf{F}, \mathscr{F}\mathbf{G})$. Hence $\mathscr{F}$ is a continuous linear mapping $\mathscr{F} : L^2 \to L^2$. It is also one-to-one [42], and hence has an inverse $\mathscr{F}^{-1} : L^2 \to L^2$. If $\hat{\mathbf{F}} \in L^2$, then the inverse is given by [42]

$$\mathscr{F}^{-1}\{\hat{\mathbf{F}}(\mathbf{k})\}(\mathbf{x}) = \lim_{n\to\infty} \int_{\mathbb{R}^3} \chi_n(\mathbf{k})\hat{\mathbf{F}}(\mathbf{k})e^{2\pi i\mathbf{k}\cdot\mathbf{x}}d\mathbf{k}. \tag{12}$$

Curl and divergence in L^2

For a vector field $\mathbf{F} \in L^2$, we define curl and divergence as

$$\begin{aligned} \nabla \times \mathbf{F} &= \mathscr{F}^{-1}\{2\pi i\mathbf{k} \times \hat{\mathbf{F}}\}, &(13)\\ \nabla \cdot \mathbf{F} &= \mathscr{F}^{-1}\{2\pi i\mathbf{k} \cdot \hat{\mathbf{F}}\}, &(14)\end{aligned}$$

where in equation 14, $\mathscr{F}^{-1}$ is the inverse of the L^2 scalar Fourier transform. If $|\mathbf{k}|\hat{\mathbf{F}}$ is a L^2 vector field, then $\nabla \times \mathbf{F}$ is also an L^2 vector field, and $\nabla \cdot \mathbf{F}$ is an L^2 function ($\int_{\mathbb{R}^3} |\nabla \cdot \mathbf{F}|^2 d\mathbf{x} < \infty$). These results follow from the inequalities $|\mathbf{u}_r \times \hat{\mathbf{F}}| \leq 2|\hat{\mathbf{F}}|$ and $|\mathbf{u}_r \cdot \hat{\mathbf{F}}| \leq 2|\hat{\mathbf{F}}|$ valid for all complex vectors $\hat{\mathbf{F}}$.

In order for curl and divergence to be well-defined, it is natural to restrict the physical vector fields to the function space

$$L^2_{\text{curl}} = \{\mathbf{F} \,|\, \mathbf{F} \in L^2,\ |\mathbf{k}|\hat{\mathbf{F}} \in L^2\}.$$

Elements in L^2_{curl} have finite energy and finite helicity.

The helicity decomposition

We are now in position to define the helicity decomposition. Motivated by Sections 3.1.1-3.1.2, we define for $\lambda = 0, \pm 1$ in the Fourier space $\mathbb{R}^3$ the dyadics

$$\overline{\overline{\mathbf{P}}}_\lambda(\mathbf{k}) = \begin{cases} \frac{1}{2}\big(\overline{\overline{\mathbf{I}}} + i\lambda\mathbf{u}_r \times \overline{\overline{\mathbf{I}}}\big) \cdot \overline{\overline{\mathbf{P}}}_t & \text{when } \lambda = \pm 1, \mathbf{k} \neq 0, \\ \mathbf{u}_r\mathbf{u}_r & \text{when } \lambda = 0, \mathbf{k} \neq 0, \\ 0 & \text{when } \mathbf{k} = 0, \end{cases}$$

where $\overline{\overline{\mathbf{P}}}_t = (\overline{\overline{\mathbf{I}}} - \mathbf{u}_r\mathbf{u}_r)$. Since we have set $\overline{\overline{\mathbf{P}}}_\lambda$ to zero when $\mathbf{k} = 0$, $\overline{\overline{\mathbf{P}}}_\lambda$ are only dyadics almost everywhere. Despite this we call $\overline{\overline{\mathbf{P}}}_\lambda$ dyadics.

The helicity decomposition in physical space is obtained by lifting the operators $\overline{\overline{\mathbf{P}}}_+, \overline{\overline{\mathbf{P}}}_-, \overline{\overline{\mathbf{P}}}_0$ from Fourier space onto the physical space. The induced operators in physical space, which we denote by π_+, π_-, π_0, are defined as follows.

Definition 3.4 (Helicity decomposition) *Let $\boldsymbol{F}$ be a real valued vector field in L^2_{curl}. For $\lambda = 0, \pm 1$,*

$$\pi_\lambda \boldsymbol{F} = \mathscr{F}^{-1}\big\{\overline{\overline{\boldsymbol{P}}}_\lambda \cdot \mathscr{F}\{\boldsymbol{F}\}\big\}.$$

(In L^2_{curl}, we say that a vector field $\mathbf{F}$ is real valued, if $\mathbf{F} = \mathbf{F}^*$.) If $\mathbf{F} \in L^2_{\text{curl}}$, then $\hat{\mathbf{F}} \in L^2$ and $|\overline{\overline{\mathbf{P}}}_\lambda \cdot \hat{\mathbf{F}}| \leq 5|\hat{\mathbf{F}}|$, so the above definition is well-defined. From the definition, we also see that the decomposed fields are independent of the value of $\overline{\overline{\mathbf{P}}}_\lambda$ at $\mathbf{k} = 0$. For the decomposed fields, we will also use the notation $\mathbf{F}_\lambda = \pi_\lambda \mathbf{F}$ and $\hat{\mathbf{F}}_\lambda = \overline{\overline{\mathbf{P}}}_\lambda \cdot \hat{\mathbf{F}}$. To index the decomposed fields, we exclusively use the Greek letters λ and κ. These only take values in $\{0, \pm 1\}$.

The next theorem describes the basic properties of the helicity decomposition. The proof of this theorem is based on Lemma 3.6, which describes the algebraic properties of the dyadics $\overline{\overline{\mathbf{P}}}_\lambda$ in Fourier space.

Theorem 3.5 (Helicity decomposition) *On $\mathbb{R}^3$ with Cartesian coordinates, let $\boldsymbol{F}$ be a real valued vector field in L^2_{curl}. Then*

$$\boldsymbol{F} = \boldsymbol{F}_0 + \boldsymbol{F}_+ + \boldsymbol{F}_-,$$

and each $\boldsymbol{F}_\lambda = \pi_\lambda \boldsymbol{F}$ is a real valued vector field in L^2_{curl}.

The decomposed fields satisfy $\nabla \cdot \boldsymbol{F}_\pm = 0$, $\nabla \times \boldsymbol{F}_0 = 0$, $\mathscr{H}(\boldsymbol{F}_0) = 0$, $\mathscr{H}(\boldsymbol{F}_+) \geq 0$, and $\mathscr{H}(\boldsymbol{F}_-) \leq 0$. Also, if we have that $\mathscr{H}(\boldsymbol{F}_+) = 0$, then $\boldsymbol{F}_+ = 0$, and if we have that $\mathscr{H}(\boldsymbol{F}_-) = 0$, then $\boldsymbol{F}_- = 0$.

Lemma 3.6 *For all λ, κ in $0, \pm 1$, and $\boldsymbol{k} \neq 0$,*

a) $\overline{\overline{\boldsymbol{P}}}_0(\boldsymbol{k}) + \overline{\overline{\boldsymbol{P}}}_+(\boldsymbol{k}) + \overline{\overline{\boldsymbol{P}}}_-(\boldsymbol{k}) = \overline{\overline{\boldsymbol{I}}}$.

b) The $\overline{\overline{\boldsymbol{P}}}_\lambda(\boldsymbol{k})$ operators are orthogonal projections, i.e.,

$$\overline{\overline{\boldsymbol{P}}}_\lambda(\boldsymbol{k}) \cdot \overline{\overline{\boldsymbol{P}}}_\kappa(\boldsymbol{k}) = \delta_{\lambda\kappa}\overline{\overline{\boldsymbol{P}}}_\lambda(\boldsymbol{k}).$$

c) $i\boldsymbol{k} \times \overline{\overline{\boldsymbol{P}}}_\lambda(\boldsymbol{k}) = \lambda|\boldsymbol{k}|\overline{\overline{\boldsymbol{P}}}_\lambda(\boldsymbol{k})$.

Proof of lemma 3.6. Part *a)* follows from $\overline{\overline{\mathbf{P}}}_0+(\overline{\overline{\mathbf{P}}}_+ +\overline{\overline{\mathbf{P}}}_-) = \mathbf{u}_r\mathbf{u}_r+\overline{\overline{\mathbf{P}}}_t = \overline{\overline{\mathbf{I}}}$. For part *b)* suppose first that both λ and κ are in ± 1. Since $\overline{\overline{\mathbf{L}}} = i\mathbf{u}_r \times \overline{\overline{\mathbf{I}}}$ is an involution for tangential fields, it follows, by Proposition 3.1, that $\overline{\overline{\mathbf{P}}}'_\lambda \cdot \overline{\overline{\mathbf{P}}}'_\kappa = \delta_{\lambda\kappa}\overline{\overline{\mathbf{P}}}'_\lambda$ for $\overline{\overline{\mathbf{P}}}'_\lambda = \frac{1}{2}(\overline{\overline{\mathbf{I}}} + i\lambda\mathbf{u}_r \times \overline{\overline{\mathbf{I}}})$. Writing out $\overline{\overline{\mathbf{P}}}'_\lambda \cdot \overline{\overline{\mathbf{P}}}_t \cdot \mathbf{x} = \overline{\overline{\mathbf{P}}}_t \cdot \overline{\overline{\mathbf{P}}}'_\lambda \cdot \mathbf{x}$ for an arbitrary vector $\mathbf{x}$ shows that $\overline{\overline{\mathbf{P}}}'_\lambda \cdot \overline{\overline{\mathbf{P}}}_t = \overline{\overline{\mathbf{P}}}_t \cdot \overline{\overline{\mathbf{P}}}'_\lambda$. Then, since $\overline{\overline{\mathbf{P}}}_t \cdot \overline{\overline{\mathbf{P}}}_t = \overline{\overline{\mathbf{P}}}_t$, and since $\overline{\overline{\mathbf{P}}}'_\lambda \cdot \overline{\overline{\mathbf{P}}}_t = \overline{\overline{\mathbf{P}}}_\lambda$, part *b)* follows for λ, κ in ± 1. Next, suppose $\lambda = 0$ and κ is in ± 1. Then $\overline{\overline{\mathbf{P}}}_0 \cdot \overline{\overline{\mathbf{P}}}_\pm = \overline{\overline{\mathbf{P}}}_0 \cdot \overline{\overline{\mathbf{P}}}'_\pm \cdot \overline{\overline{\mathbf{P}}}_t = \overline{\overline{\mathbf{P}}}_0 \cdot \overline{\overline{\mathbf{P}}}_t \cdot \overline{\overline{\mathbf{P}}}'_\pm = 0$ since $\overline{\overline{\mathbf{P}}}_0 \cdot \overline{\overline{\mathbf{P}}}_t = 0$. One similarly proves the claim when $\kappa = 0$ and $\lambda = \pm 1$ and when $\lambda = \kappa = 0$. Part *c)* is clear when $\lambda = 0$. When $\lambda = \pm 1$, we know, by Proposition 3.1, that $i\mathbf{u}_r \times \overline{\overline{\mathbf{P}}}'_\lambda = \lambda\overline{\overline{\mathbf{P}}}'_\lambda$. The claim follows by multiplying scalarly by $|\mathbf{k}|$ and dot-multiplying from the right with $\overline{\overline{\mathbf{P}}}_t$. □

Using part *c)* we obtain the following useful relation

$$\nabla \times \mathbf{F}_\lambda = \mathscr{F}^{-1}\big\{2\pi\lambda|\mathbf{k}|\hat{\mathbf{F}}_\lambda\big\}. \tag{15}$$

This equation shows that curl is a scalar operator for the decomposed fields in Fourier space.

Proof of Theorem 3.5. By the comment after Definition 3.4, it follows that all $\mathbf{F}_\lambda$ are in L^2_{curl}. To prove that $\mathbf{F}_\lambda = \mathbf{F}^*_\lambda$ we show that $\mathscr{F}^{-1}\hat{\mathbf{F}}_\lambda = \big(\mathscr{F}^{-1}\hat{\mathbf{F}}_\lambda\big)^*$. Writing out this condition using equation 12, inverting the integration coordinate in one integral, and using the fact that the Fourier transform is one-to-one yields the equivalent condition $\hat{\mathbf{F}}^*_\lambda(\mathbf{k}) = \hat{\mathbf{F}}_\lambda(-\mathbf{k})$. Since $\mathbf{F}$ is real, $\hat{\mathbf{F}}$ must satisfy this condition, and since it holds that $\overline{\overline{\mathbf{P}}}^*_\lambda(\mathbf{k}) = \overline{\overline{\mathbf{P}}}_\lambda(-\mathbf{k})$, we have that $\mathbf{F}_\lambda = \mathbf{F}^*_\lambda$. Using the relation $(\mathbf{F}, \mathbf{G}) = (\hat{\mathbf{F}}, \hat{\mathbf{G}})$ and equation 15, we obtain the relation $\mathscr{H}(\mathbf{F}_\lambda) = 2\pi\lambda(\hat{\mathbf{F}}_\lambda, |\mathbf{k}|\hat{\mathbf{F}}_\lambda)$. It follows that $\mathscr{H}(\mathbf{F}_+) \geq 0$ and that $\mathscr{H}(\mathbf{F}_-) \leq 0$. Also, if we have that $\mathscr{H}(\mathbf{F}_\pm) = 0$, then $\mathbf{F}_\pm(\mathbf{x}) = 0$. □

The next property is essential for decomposing Maxwell's equations.

Theorem 3.7 *If $\boldsymbol{F}$ is a real valued vector field in L^2_{curl}, then for all λ,*

$$\pi_\lambda\big(\nabla \times \boldsymbol{F}\big) = \nabla \times (\pi_\lambda \boldsymbol{F}).$$

Proof. Using properties *a)* and *b)* in Lemma 3.6 and equation 15, we have

$$\begin{aligned}
\pi_\lambda\big(\nabla \times \mathbf{F}\big) &= \sum_\kappa \pi_\lambda \nabla \times \mathbf{F}_\kappa \\
&= \sum_\kappa \mathscr{F}^{-1}\{\overline{\overline{\mathbf{P}}}_\lambda \cdot 2\pi\kappa|\mathbf{k}|\hat{\mathbf{F}}_\kappa\} \\
&= \nabla \times \big(\pi_\lambda \mathbf{F}\big).
\end{aligned}$$

□

Proposition 3.8 *Let $\boldsymbol{F}$ and $\mathbf{G}$ be real valued vector fields in L^2_{curl}. Then*

a) For all λ and κ, $\pi_\lambda \pi_\kappa \boldsymbol{F} = \delta_{\lambda\kappa}\pi_\lambda \boldsymbol{F}$.

b) For all λ, the operators π_λ are self-adjoint *[39] in the L^2 inner product, $(\pi_\lambda \boldsymbol{F}, \boldsymbol{G}) = (\boldsymbol{F}, \pi_\lambda \boldsymbol{G})$.*

c) If λ and κ are distinct, then $\boldsymbol{F}_\lambda$ and $\boldsymbol{G}_\kappa$ are orthogonal, $(\boldsymbol{F}_\lambda, \boldsymbol{G}_\kappa) = 0$, and their cross-helicity is zero, $\mathscr{H}(\boldsymbol{F}_\lambda, \boldsymbol{G}_\kappa) = 0$.

d) $\mathscr{H}(\boldsymbol{F}, \boldsymbol{G}) = \mathscr{H}(\boldsymbol{G}, \boldsymbol{F})$.

Proof. Part *a)* follows from part *b)* in Lemma 3.6. For part *b)*, we need to show that $\overline{\overline{\mathbf{P}}}^*_\lambda = \overline{\overline{\mathbf{P}}}^{\mathrm{T}}_\lambda$ for all λ. The relation is clear for $\lambda = 0$. For $\lambda = \pm 1$, we have (using notation from the proof of Lemma 3.6) that $\overline{\overline{\mathbf{P}}}^{\mathrm{T}}_\lambda = \overline{\overline{\mathbf{P}}}^{\mathrm{T}}_t \cdot \overline{\overline{\mathbf{P}}}'^{\mathrm{T}}_\lambda = \overline{\overline{\mathbf{P}}}_t \cdot \overline{\overline{\mathbf{P}}}'_{-\lambda} = \overline{\overline{\mathbf{P}}}^*_\lambda$. Part *c)* follows from the relation $\mathscr{H}(\mathbf{F}) = \sum_\lambda (\mathbf{F}_\lambda, \nabla \times \mathbf{F}_\lambda)$, which is a consequence of parts *a)* and *b)* and Theorem 3.7. Part *d)* follows from part *c)*, from the relation $(\mathbf{F}, \mathbf{G}) = (\hat{\mathbf{F}}, \hat{\mathbf{G}})$, and from equation 15. □

The above proposition shows that the images of the π_λ operators decompose L^2_{curl} into three components, $L^2_{\mathrm{curl}} = \pi_+ L^2_{\mathrm{curl}} + \pi_- L^2_{\mathrm{curl}} + \pi_0 L^2_{\mathrm{curl}}$. Any two vector fields from distinct components are both orthogonal and have zero cross-helicity. For arbitrary real vector fields $\mathbf{F}, \mathbf{G} \in L^2_{\mathrm{curl}}$, we have the relations

$$
\begin{aligned}
||\mathbf{F}||^2 &= ||\mathbf{F}_+||^2 + ||\mathbf{F}_-||^2 + ||\mathbf{F}_0||^2 \\
&= ||\hat{\mathbf{F}}_+||^2 + ||\hat{\mathbf{F}}_-||^2 + ||\hat{\mathbf{F}}_0||^2, \\
(\mathbf{F}, \mathbf{G}) &= (\mathbf{F}_+, \mathbf{G}_+) + (\mathbf{F}_-, \mathbf{G}_-) + (\mathbf{F}_0, \mathbf{G}_0) \\
&= (\hat{\mathbf{F}}_+, \hat{\mathbf{G}}_+) + (\hat{\mathbf{F}}_-, \hat{\mathbf{G}}_-) + (\hat{\mathbf{F}}_0, \hat{\mathbf{G}}_0), \\
\mathscr{H}(\mathbf{F}) &= \mathscr{H}(\mathbf{F}_+) + \mathscr{H}(\mathbf{F}_-), \\
\mathscr{H}(\mathbf{F}, \mathbf{G}) &= \mathscr{H}(\mathbf{F}_+, \mathbf{G}_+) + \mathscr{H}(\mathbf{F}_-, \mathbf{G}_-).
\end{aligned}
$$

In particular, helicity does not depend on the 0-component. Since the operators π_λ are linear operators, they decompose the space L^2 into three linear vector spaces: $\pi_+ L^2_{\mathrm{curl}}$, $\pi_- L^2_{\mathrm{curl}}$, and $\pi_0 L^2_{\mathrm{curl}}$.

Proposition 3.9 *If $\lambda = \pm 1$, then $(\cdot, \cdot)_\lambda = \lambda \mathscr{H}(\cdot, \cdot)$ is an inner product for real valued vector fields in $\pi_\lambda L^2_{curl}$.*

Proof. It is clear that $(\cdot, \cdot)_\lambda$ is linear in both arguments. By part *d)* in Proposition 3.8, $(\cdot, \cdot)_\lambda$ is also symmetric, and by Theorem 3.5, $(\cdot, \cdot)_\lambda$ is positive definite. □

Corollary 3.10 *If $\lambda = \pm 1$, then $|| \cdot ||_\lambda = \sqrt{(\cdot, \cdot)_\lambda}$ is a norm for real valued vector fields in $\pi_\lambda L^2_{curl}$.*

We now have the following expression for helicity:

$$\mathscr{H}(\mathbf{F}) = ||\mathbf{F}_+||_+^2 - ||\mathbf{F}_-||_-^2.$$

This formula explains the degenerate character of helicity observed in Example 2.2.

Proposition 3.11 *Let $\boldsymbol{F}$ be a real valued vector field in L^2_{curl}. If $\lambda = \pm 1$, then*

$$||\boldsymbol{F}_\lambda||_\lambda^2 \geq C||\boldsymbol{F}_\lambda||^2$$

for some $C \in (0, 2\pi]$.

Proof. Let B is the Cartesian unit ball in $\mathbb{R}^3$ with origin as center. We then have that

$$||\mathbf{F}_\lambda||_\lambda^2 = 2\pi \int_B |\mathbf{k}||\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k} + 2\pi \int_{\mathbb{R}^3 \setminus B} |\mathbf{k}||\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k}.$$

If $\int_B |\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k}$ vanishes, the claim follows since $|\mathbf{k}||\hat{\mathbf{F}}_\lambda|^2 \geq |\hat{\mathbf{F}}_\lambda|^2$ outside B. Otherwise, there exists a constant $C \in (0, 1)$, such that $\int_B |\mathbf{k}||\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k} = C \int_B |\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k}$ and we have that

$$\begin{aligned} ||\mathbf{F}_\lambda||_\lambda^2 &\geq 2\pi C \int_B |\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k} + 2\pi \int_{\mathbb{R}^3 \setminus B} |\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k} \\ &\geq 2\pi C \int_{\mathbb{R}^3} |\hat{\mathbf{F}}_\lambda|^2 d\mathbf{k} \\ &= 2\pi C ||\mathbf{F}_\lambda||^2. \end{aligned}$$

□

We end this section with a remark about the helicity decomposition and $\mathscr{H}_{\text{fluid}}$. From equation 15, it follows that $\nabla \times^{-1} \mathbf{F}_\lambda = \mathscr{F}^{-1}\left\{\frac{\lambda}{2\pi|\mathbf{k}|}\hat{\mathbf{F}}_\lambda\right\}$ assuming that $\frac{1}{|\mathbf{k}|}\hat{\mathbf{F}} \in L^2$. Thus all the results we have proved for the decomposed fields concerning $\mathscr{H}$ can also be proved for $\mathscr{H}_{\text{fluid}}$ by exchanging $2\pi|\mathbf{k}| \leftrightarrow \frac{1}{2\pi|\mathbf{k}|}$. In Proposition 3.11, however, this exchange turns the inequality sign and gives another bound for C. In [23] (page 122), an alternative proof of Proposition 3.11 is given (for $\mathscr{H}_{\text{fluid}}$) using Poincaré's inequality.

3.2.1 Invariance of Cartesian coordinates

We next show that the helicity decomposition is invariant under coordinate rotations, changes of origin, and positive scalings. Hence the decomposition does not depend on the choice of Cartesian coordinates. A physical interpretation is that different observers in $\mathbb{R}^3$ obtain the same decomposition of a given field. We also show that under coordinate inversions, the $+$-component and the $-$-component exchange roles. This is a typical handed result.

Proposition 3.12 *Let $\boldsymbol{F}$ be real valued vector fields in L^2_{curl}. Then*

a) Under the coordinate scaling/inversion $\boldsymbol{x} \mapsto \alpha\boldsymbol{x}$ ($\alpha \neq 0$), the transformed field $\boldsymbol{G}(\boldsymbol{x}) = \alpha\boldsymbol{F}(\frac{1}{\alpha}\boldsymbol{x})$ decomposes as

$$\begin{aligned} \boldsymbol{G}_0(\boldsymbol{x}) &= \alpha\boldsymbol{F}_0(\frac{1}{\alpha}\boldsymbol{x}), \\ \boldsymbol{G}_\lambda(\boldsymbol{x}) &= \alpha\boldsymbol{F}_{\lambda\,\mathrm{sgn}\alpha}(\frac{1}{\alpha}\boldsymbol{x}), \quad (\lambda = \pm 1), \end{aligned}$$

where sgn α is the signum function*; sgn $\alpha = +1$ for $\alpha > 0$, and sgn $\alpha = -1$ for $\alpha < 0$.*

b) Under the change of origin $\boldsymbol{x} \mapsto \boldsymbol{x} - \boldsymbol{x}_0$ ($\boldsymbol{x}_0 \in \mathbb{R}^3$), the transformed field $\boldsymbol{G}(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{x} - \boldsymbol{x}_0)$ decomposes as

$$\boldsymbol{G}_\lambda(\boldsymbol{x}) = \boldsymbol{F}_\lambda(\boldsymbol{x} - \boldsymbol{x}_0)$$

for all λ.

c) If $\overline{\overline{\boldsymbol{R}}}$ is a rotational dyad, then, under the rotation $\boldsymbol{x} \mapsto \overline{\overline{\boldsymbol{R}}}\cdot\boldsymbol{x}$, the rotated field $\boldsymbol{G}(\boldsymbol{x}) = \overline{\overline{\boldsymbol{R}}} \cdot \boldsymbol{F}(\overline{\overline{\boldsymbol{R}}}^{-1} \cdot \boldsymbol{x})$ decomposes as

$$\boldsymbol{G}_\lambda(\boldsymbol{x}) = \overline{\overline{\boldsymbol{R}}} \cdot \boldsymbol{F}_\lambda(\overline{\overline{\boldsymbol{R}}}^{-1} \cdot \boldsymbol{x}).$$

In part *a)* the vector field $\alpha\mathbf{F}(\frac{1}{\alpha}\mathbf{x})$ is constructed in the following way. If $\mathbf{x}$ is the scaled coordinate, then $\mathbf{F}(\frac{1}{\alpha}\mathbf{x})$ is the unscaled value of $\mathbf{F}$ at $\mathbf{x}$. Thus $\alpha\mathbf{F}(\frac{1}{\alpha}\mathbf{x})$ is the scaled value of $\mathbf{F}$ at $\mathbf{x}$. In *c)* the transformed field is constructed in the same way [31].

Proof. For part *a)* we have in L^1 that $\mathscr{F}\{\mathbf{F}(\frac{1}{\alpha}\mathbf{x})\}(\mathbf{k}) = |\alpha|^3\mathscr{F}\{\mathbf{F}(\mathbf{x})\}(\alpha\mathbf{k})$. Since $L^1 \cap L^2$ is dense in L^2, the result also holds when $\mathbf{F} \in L^2$. Using $\overline{\overline{\mathbf{P}}}_\lambda(\mathbf{k}) = \overline{\overline{\mathbf{P}}}_{\lambda\mathrm{sgn}\,\alpha}(\alpha\mathbf{k})$, it follows that

$$\begin{aligned} \pi_\lambda\mathbf{G}(\mathbf{x}) &= \mathscr{F}^{-1}\{\overline{\overline{\mathbf{P}}}_\lambda(\mathbf{k}) \cdot \hat{\mathbf{G}}(\mathbf{k})\} \\ &= \mathscr{F}^{-1}\{\alpha|\alpha|^3\overline{\overline{\mathbf{P}}}_{\lambda\,\mathrm{sgn}\alpha}(\alpha\mathbf{k}) \cdot \hat{\mathbf{F}}(\alpha\mathbf{k})\} \\ &= \alpha\mathbf{F}_{\lambda\mathrm{sgn}\alpha}(\frac{1}{\alpha}\mathbf{x}). \end{aligned}$$

For part *b)* we have in L^1 that $\hat{\mathbf{G}}(\mathbf{k}) = e^{-2\pi i\mathbf{k}\cdot\mathbf{x}_0}\hat{\mathbf{F}}(\mathbf{k})$. Again, since $L^1 \cap L^2$ is dense in L^2, the result also holds in L^2. Then

$$\begin{aligned}\pi_\lambda \mathbf{G}(\mathbf{x}) &= \mathscr{F}^{-1}\{\overline{\overline{\mathbf{P}}}_\lambda(\mathbf{k}) \cdot \hat{\mathbf{G}}(\mathbf{k})\} \\ &= \mathscr{F}^{-1}\{e^{-2\pi i\mathbf{k}\cdot\mathbf{x}_0}\hat{\mathbf{F}}_\lambda(\mathbf{k})\} \\ &= \mathbf{F}_\lambda(\mathbf{x} - \mathbf{x}_0).\end{aligned}$$

To prove part *c)*, we first calculate the L^2 Fourier transform of the rotated field $\mathbf{G}(\mathbf{x}) = \overline{\overline{\mathbf{R}}} \cdot \mathbf{F}(\overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{x})$,

$$\begin{aligned}\hat{\mathbf{G}}(\mathbf{k}) &= \lim_{n\to\infty} \overline{\overline{\mathbf{R}}} \cdot \int_{\mathbb{R}^3} \chi_n(\mathbf{x})\mathbf{F}(\overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{x}) e^{-2\pi i(\overline{\overline{\mathbf{R}}}^{-1}\cdot\mathbf{k})\cdot(\overline{\overline{\mathbf{R}}}^{-1}\cdot\mathbf{x})} d\mathbf{x} \\ &= \lim_{n\to\infty} \overline{\overline{\mathbf{R}}} \cdot \int_{\mathbb{R}^3} \chi_n(\mathbf{z})\mathbf{F}(\mathbf{z}) e^{-2\pi i(\overline{\overline{\mathbf{R}}}^{-1}\cdot\mathbf{k})\cdot\mathbf{z}} d\mathbf{z} \\ &= \overline{\overline{\mathbf{R}}} \cdot \hat{\mathbf{F}}(\overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{k}).\end{aligned}$$

On the second line in the above calculation, we made the coordinate change $\mathbf{z} = \overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{x}$. Writing out the projection operators as in the proof of *a)* and *b)*, we see part *c)* is equivalent to the relation $\overline{\overline{\mathbf{P}}}_\lambda(\mathbf{k}) \cdot \overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{R}}} \cdot \overline{\overline{\mathbf{P}}}_\lambda(\overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{k})$. For $\lambda = 0$, this follows from $\mathbf{k} \cdot \overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{k}$ and $|\mathbf{k}| = |\overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{k}|$. When $\lambda = \pm 1$, we have that

$$\overline{\overline{\mathbf{P}}}_\lambda(\mathbf{k}) = \frac{1}{2}\big(\overline{\overline{\mathbf{I}}} + \lambda i \mathbf{u}_r(\mathbf{k}) \times \overline{\overline{\mathbf{I}}} - \mathbf{u}_r(\mathbf{k})\mathbf{u}_r(\mathbf{k})\big). \tag{16}$$

We therefore need to show that $\big((\mathbf{u}_r \times \overline{\overline{\mathbf{I}}}) \cdot \overline{\overline{\mathbf{R}}}\big) \cdot \mathbf{x} = \big(\overline{\overline{\mathbf{R}}} \cdot (\mathbf{u}_r(\overline{\overline{\mathbf{R}}}^{-1} \cdot \mathbf{k}) \times \overline{\overline{\mathbf{I}}})\big) \cdot \mathbf{x}$ for all $\mathbf{x}$. This follows by setting $\mathbf{u}$ to $\mathbf{u}_r(\mathbf{k})$, $\mathbf{v}$ to $\overline{\overline{\mathbf{R}}} \cdot \mathbf{x}$ and $\overline{\overline{\mathbf{R}}}$ to $\overline{\overline{\mathbf{R}}}^{-1}$ in Lemma A.3 in Appendix A. (The inverse of a rotational dyad is also a rotational dyad.) $\square$

In Section 6.2.1 we show that the Bohren decomposition contains, as a special case, the decomposition of a plane wave into one right-hand circulary polarized component and one left-hand circulary polarized component. Since the helicity decomposition is a generalization of the Bohren decomposition (see Section 4.3.2), it follows that a general affine transformation $\mathbf{x} \mapsto \overline{\overline{\mathbf{A}}} \cdot \mathbf{x}$ may couple the helicity decomposition. For instance, consider a right-hand circular polarized plane wave. If the transformation is not a rotation in the plane perpendicular to the direction of propagation, the polarization changes into an elliptical polarization, which also has a left-hand polarized component.

3.2.2 Invariance under spatial convolutions

We next show that the helicity decomposition is invariant under spatial convolutions.

Definition 3.13 *Let $g : \mathbb{R}^3 \to \mathbb{R}$ is a scalar function on $\mathbb{R}^3$, and let $\boldsymbol{F}$ be a vector field. Then the* spatial convolution *of g and $\boldsymbol{F}$ is*

$$\big(g * \boldsymbol{F}\big)(\boldsymbol{x}) = \int_{\mathbb{R}^3} g(\boldsymbol{x}-\boldsymbol{y})\boldsymbol{F}(\boldsymbol{y})d\boldsymbol{y}.$$

Proposition 3.14 *Let $\boldsymbol{F} \in L^2_{curl}$ and let $g : \mathbb{R}^3 \to \mathbb{R}$ be an L^1 function. Then*

$$\pi_\lambda(g * \boldsymbol{F})(\boldsymbol{x}) \quad = \quad (g * \pi_\lambda \boldsymbol{F})(\boldsymbol{x}).$$

Proof. Since $\mathbf{F} \in L^2$, there is a sequence $\mathbf{F}_n$ in $L^1 \cap L^2$ converging to $\mathbf{F}$ in the L^2 norm. By Young's inequality [41], the sequence $g * \mathbf{F}_n$ is also in $L^1 \cap L^2$, it converges to $g * \mathbf{F}$, and $g * \mathbf{F}$ is an element in L^2. By a standard result for the L^1 Fourier transform, we have for each n that $\mathscr{F}\{g * \mathbf{F}_n\} = \hat{g}\mathscr{F}\mathbf{F}_n$, where $\hat{g} = \int_{\mathbb{R}^3} g(\mathbf{x})e^{-2\pi i \mathbf{x}\cdot\mathbf{k}}d\mathbf{x}$ is the L^1 Fourier transform for the scalar function g [41]. Taking $n \to \infty$ shows that $\mathscr{F}\{g * \mathbf{F}\} = \hat{g}\mathscr{F}\mathbf{F}$ whence

$$\begin{aligned} \pi_\lambda(g * \mathbf{F}) &= \mathscr{F}^{-1}\{\overline{\overline{\mathbf{P}}}_\lambda \cdot \hat{g}\hat{\mathbf{F}}\} \\ &= \mathscr{F}^{-1}\{\hat{g}\mathscr{F}\{\mathbf{F}_\lambda\}\} \\ &= g * \pi_\lambda \mathbf{F}. \end{aligned}$$

□

3.2.3 Local properties of the helicity decomposition

One of the disadvantages of the helicity decomposition is that it does not preserve the support of a vector field. In other words, if we wish to decompose a vector field, which is non-zero only in a small region of $\mathbb{R}^3$, the decomposed fields do not necessarily vanish outside this region. This is also the case for the Helmholtz's decomposition [43].

3.3 Helicity decomposition for time dependent vector fields

The helicity decomposition of a time dependent vector field $\mathbf{F}(\mathbf{x}, t)$ is defined pointwise for each t in some interval $J \subset \mathbb{R}$. For this operation to be well-defined, $\mathbf{F}(\mathbf{x}, t)$ must be an element in L^2_{curl} for each $t \in J$. In this section we show that, at least formally, the helicity decomposition for time dependent fields commutes with the time derivative and temporal convolutions.

Time derivatives
To show that the helicity decomposition commutes with time derivatives, it is

necessary to show that

$$\frac{\partial}{\partial t}\pi_\lambda \mathbf{F} = \pi_\lambda \frac{\partial}{\partial t}\mathbf{F}.$$

If we assume that $\frac{\partial}{\partial t}$ can be taken in under both integral signs in π_λ, and if we assume that $\frac{\partial}{\partial t}\mathbf{F}$ is also in L^2_{curl} for all $t \in J$, then the claim follows. The precise conditions when integration and derivation by a parameter commutes are given by Theorem 8.12 in [40]. Essentially, it must be possible to majorize each component-function of $\mathbf{F}$ and $\hat{\mathbf{F}}$ by an L^1 function that does not depend on t.

Time convolutions

If $g : \mathbb{R} \to \mathbb{R}$ is a scalar function, and $\mathbf{F}(\mathbf{x}, t)$ is a time dependent vector field, then the *temporal convolution* of g and $\mathbf{F}$ is the time-dependent vector field

$$\big(g * \mathbf{F}\big)(\mathbf{x}, t) = \int_{-\infty}^{\infty} g(t-\tau)\mathbf{F}(\mathbf{x}, \tau) d\tau.$$

If we can assume that $g * \mathbf{F} \in L^2$ for each t, and that the order of integration can be changed in $\pi_\lambda(g * \mathbf{F})(\mathbf{x}, t)$, we have that

$$\pi_\lambda(g * \mathbf{F})(\mathbf{x}, t) = (g * \pi_\lambda \mathbf{F})(\mathbf{x}, t).$$

3.4 The Moses decompositions

In this section we show that the helicity decomposition derived in the previous sections is, in fact, the so called *Moses decomposition*. This result ties together three decompositions: the Helmholtz decomposition, the Bohren decomposition, and the Moses decomposition. The helicity decomposition is a refinement of the Helmholtz's decomposition, a generalization of the Bohren decomposition, and the Moses decomposition is one way of writing down the helicity decomposition in basis.

The Moses decomposition was introduced in 1971, and it has been studied in many papers, mostly related to fluid dynamics [31, 32, 38]. It was first applied to electromagnetics in homogeneous isotropic media in the original paper [31]. In the same paper, the decomposition is also given a mathematical derivation. In [44], the Moses decomposition is studied in a relativistic setting.

The Moses decomposition is defined as follows. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2$ be an orthonormal basis for $\mathbb{R}^3$, and let $\mathbf{k} = \sum k_i \mathbf{u}_i$ and $k = |\mathbf{k}|$. Then the Moses decomposition introduces the *complex basis* $\mathbf{Q}_\lambda(\mathbf{k})$ in Fourier space $\mathbb{R}^3 \backslash \{0\}$

$$\mathbf{Q}_0(\mathbf{k}) \;\; = \;\; -(k_1, k_2, k_3)/k,$$

and for $\lambda = \pm 1$,

$$\mathbf{Q}_\lambda(\mathbf{k}) \;=\; -\frac{\lambda}{\sqrt{2}}\Big(\frac{k_1(k_1+i\lambda k_2)}{k(k+k_3)} - 1, \frac{k_2(k_1+i\lambda k_2)}{k(k+k_3)} - i\lambda, \frac{k_1+i\lambda k_2}{k}\Big).$$

The properties of the $\mathbf{Q}_\lambda$ vectors are investigated in [31]. In the same reference the definition of the $\mathbf{Q}_\lambda$ vectors is also motivated. Here we only mention that they are both orthonormal, i.e., $\mathbf{Q}_\lambda(\mathbf{k}) \cdot \mathbf{Q}^*_\kappa(\mathbf{k}) = \delta_{\lambda\kappa}$ and complete. Moreover, they satisfy $\mathbf{k} \times \mathbf{Q}(\mathbf{k}) = -i\lambda|\mathbf{k}|\mathbf{Q}_\lambda(\mathbf{k})$ for $\lambda = \pm 1$, which is the key property, which makes the Moses decomposition well behaved under curl.

The *Moses decomposition* $\mathbf{F} = \mathbf{F}_+ + \mathbf{F}_- + \mathbf{F}_0$ in physical space is defined as the projection onto the $\mathbf{Q}_\lambda$ basis in Fourier space. If we denote the projection operators in physical space by $\tilde{\pi}_\lambda$ then

$$\tilde{\pi}_\lambda \mathbf{F} = \mathscr{F}^{-1}\{f_\lambda(\mathbf{k})\mathbf{Q}_\lambda(\mathbf{k})\},$$

where $f_\lambda(\mathbf{k}) = \mathscr{F}\{\mathbf{F}\}(\mathbf{k}) \cdot \mathbf{Q}^*_\lambda(\mathbf{k})$ and $\lambda = 0, \pm 1$.

The main advantage of the Moses decomposition is that each component is written using only one scalar function. The disadvantage, on the other hand, is that the basis vectors $\mathbf{Q}_\lambda$ are not defined when $k + k_3 = 0$ [31].

We next show that $\overline{\overline{\mathbf{P}}}_\lambda = \mathbf{Q}_\lambda \mathbf{Q}^*_\lambda$ for all λ up to the aforementioned singularity. To prove this, we show that the 3×3 matrices corresponding to these dyadics are equal. For $\lambda = 0$ this is clear, so let us assume that $\lambda = \pm 1$. To construct the 3×3 matrices corresponding to the $\overline{\overline{\mathbf{P}}}_\lambda$ dyadics, we first note that the matrix corresponding to $\mathbf{u}_r(\mathbf{k})\mathbf{u}_r(\mathbf{k})$ is

$$\frac{1}{k^2}\begin{pmatrix} k_1k_1 & k_1k_2 & k_1k_3 \\ k_2k_1 & k_2k_2 & k_2k_3 \\ k_3k_1 & k_3k_2 & k_3k_3 \end{pmatrix}.$$

Also, the matrix corresponding to $\mathbf{u}_r \times \overline{\overline{\mathbf{I}}}$ is given by equation 67 in Appendix A. Using these results and equation 16 gives the matrix representations for $\overline{\overline{\mathbf{P}}}_\lambda$,

$$P_\lambda = \frac{1}{2k^2}\begin{pmatrix} k_2^2 + k_3^2 & -k_1k_2 - i\lambda k k_3 & -k_1k_3 + i\lambda k k_2 \\ -k_2k_1 + i\lambda k k_3 & k_1^2 + k_3^2 & -k_2k_3 - i\lambda k k_1 \\ -k_3k_1 - i\lambda k k_2 & -k_3k_2 + i\lambda k k_1 & k_1^2 + k_2^2 \end{pmatrix}. \tag{17}$$

It should now be obvious that it is much easier to manipulate the $\overline{\overline{\mathbf{P}}}_\lambda$ operators using dyadic algebra than using standard matrix algebra. For instance, proving how the above matrix (or the $\mathbf{Q}_\lambda$-vectors) behaves under the most general rotational matrix in Fourier space is probably quite difficult.

To show that the Moses decomposition is equal to the helicity decomposition we need to show that $\mathbf{Q}_\lambda \mathbf{Q}_\lambda^{*\,\mathrm{T}} = P_\lambda$ when the $\mathbf{Q}_\lambda$ vectors are taken as column vectors. Writing out and manipulating the coordinate expressions for $\mathbf{Q}_\lambda \mathbf{Q}_\lambda^{*\,\mathrm{T}}$ seems to be very tedious. However, applying the FullSimplify command in Mathematica (version 4.2) shows that the equality holds. However, it must be pointed out that in this calculation it is necessary to cancel the factor $k + k_3$ in the $\mathbf{Q}_\lambda$ vectors. Hence the result is only true up to the singularity on the line $k + k_3 = 0$. Since this line has zero measure, it does not modify the decomposition in physical space.

4 Helicity decomposition in electromagnetics

In this section we apply the helicity decomposition to Maxwell's equations. Without any assumptions on the media we prove that Maxwell's equations for the fields $\mathbf{E}, \mathbf{D}, \mathbf{B}\,\mathbf{H}$ decompose into three uncoupled sets of equations; one set involving only the $+$-components, one set involving only the $-$-components, and one set involving only the 0-components. This result is an immediate consequence from the fact that the helicity decomposition commutes with the curl operator and time derivatives.

In Section 4.1, we begin by formulating Maxwell's equations and the constitutive equations in different media. Then, in Section 4.2 we decompose Maxwell's equations, and in Section 4.3 we decompose the constitutive equations. In Section 4.4, we show that Maxwell's equations take a very simple form when formulated using the Moses representations for the decomposed fields in Fourier space. In Section 4.4.1 we use this formulation and derive a connection between electromagnetism and symplectic geometry

4.1 Electromagnetics in complex media

Maxwell's equation form a set of linear first order differential equations that describe how the electromagnetic fields are related to their sources. These equations can be written down in a variety of different mathematical formalisms [5]. However, to apply the helicity decomposition to Maxwell's equations, we formulate them using vectors in $\mathbb{R}^3$ with Cartesian coordinates. We therefore assume that Maxwell's equations are formulated in a non-relativistic setting. Maxwell's equations then read

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} - \mathbf{M}, \tag{18}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}, \tag{19}$$

$$\nabla \cdot \mathbf{D} = \rho, \tag{20}$$

$$\nabla \cdot \mathbf{B} = \rho_m. \tag{21}$$

In the above, $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic field intensities, $\mathbf{D}$ and $\mathbf{B}$ are the electric and magnetic flux densities, and ρ and $\mathbf{J}$ are the charge density and current. We have also included ρ_m and $\mathbf{M}$, which are the magnetic charge density and current. These are included for symmetry reasons. Physically, they might or might not exist. However, many practical problems can be solved by introducing *equivalent sources* which, in general, also contain magnetic sources [5, 6]. The standard Maxwell's equations are recovered at any time by setting $\rho_m = 0$ and $\mathbf{M} = 0$.

If the sources are known, equations 18-21 do not alone determine the fields since there are more unknowns than there are equations. Therefore, in order to make Maxwell's equations into a predictive theory, the number of unknowns must be reduced. This is done by introducing additional equations that model the electromagnetic properties of the medium. These equations are the *constitutive equations*, which, in general, can be very complicated non-linear functionals. For instance, magnetic media usually possess hysteresis. We will assume the constitutive equations are of the form

$$\begin{aligned} \mathbf{D} &= \mathbf{D}(\mathbf{E}, \mathbf{H}), & (22)\\ \mathbf{B} &= \mathbf{B}(\mathbf{E}, \mathbf{H}). & (23) \end{aligned}$$

In other words, we assume that **D** and **B** can be solved as functionals of **E** and **H**.

a) In the most simple medium the constitutive equations read

$$\begin{aligned} \mathbf{D} &= \epsilon \mathbf{E}, & (24)\\ \mathbf{B} &= \mu \mathbf{H}, & (25) \end{aligned}$$

where $\epsilon > 0$ is the *permittivity* of the material, and $\mu > 0$ is the *permeability* of the material. This media is called *linear non-dispersive isotropic media*. If ϵ and μ are constants (as, for instance, in vacuum), then the media is *homogeneous*. Otherwise, the media is *inhomogeneous*.

b) In *linear time-dispersive isotropic media*, the constitutive equations take the form

$$\begin{aligned} \mathbf{D} &= \epsilon * \mathbf{E}, & (26)\\ \mathbf{B} &= \mu * \mathbf{H}, & (27) \end{aligned}$$

where $*$ is temporal convolution, and ϵ and μ are functions $\mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$. If τ is the temporal variable, then by causality, ϵ and μ must vanish for $\tau < 0$.

c) In *non-dispersive linear bi-isotropic media*, the electric and magnetic fields are coupled as

$$\begin{aligned} \mathbf{D} &= \epsilon \mathbf{E} + \xi \mathbf{H}, & (28)\\ \mathbf{B} &= \mu \mathbf{H} + \zeta \mathbf{E}. & (29) \end{aligned}$$

Here, $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}$ are the time-harmonic fields and $\epsilon, \xi, \mu, \zeta$ are complex scalars. The geometric or physical interpretation of the above (algebraic) equations is not clear (see Section 1.1). However, it can be shown that with the above constitutive equations, the properties of the medium depends on the handedness of the fields. This means that a right-hand polarized wave can propagate with different velocity than a left-hand polarized wave.

In Section 4.3, we will see that the above equations take a particularly simple form when formulated using the decomposed fields. With this formulation it is directly seen that the response of the medium depends on the helicity of the fields.

d) In the *general linear bi-anisotropic media*, the constitutive equations read

$$\mathbf{D} = \overline{\overline{\epsilon}} \cdot \mathbf{E} + \overline{\overline{\xi}} \cdot \mathbf{H}, \tag{30}$$

$$\mathbf{B} = \overline{\overline{\mu}} \cdot \mathbf{H} + \overline{\overline{\zeta}} \cdot \mathbf{E}, \tag{31}$$

where $\overline{\overline{\epsilon}}, \overline{\overline{\xi}}, \overline{\overline{\mu}}$, and $\overline{\overline{\zeta}}$ are real dyads (or complex dyads for time harmonic fields). In this medium, the properties of the medium can depend on both the handedness of the fields and the direction of propagation of the fields. Such behavior can be found in certain crystals [30].

The above special cases can be seen as linear approximations of equations 22-23.

4.2 Decomposition of Maxwell's equations

In this section, we apply the helicity decomposition to Maxwell's equations. For this purpose, we shall assume that all the vector fields in Maxwell's equations are time dependent vector fields in a function space where the helicity decomposition is defined and time derivatives commute with the decomposition. We also use the same notation as in Section 3. For instance, we write the decomposed components of the electric field $\mathbf{E}$ as $\mathbf{E}_\lambda = \pi_\lambda \mathbf{E}$.

The $+$-component of the first two Maxwell's equations 18-19 are

$$\nabla \times \mathbf{E}_+ = -\frac{\partial \mathbf{B}_+}{\partial t} - \mathbf{M}_+, \tag{32}$$

$$\nabla \times \mathbf{H}_+ = \frac{\partial \mathbf{D}_+}{\partial t} + \mathbf{J}_+, \tag{33}$$

the $-$-components are

$$\nabla \times \mathbf{E}_- = -\frac{\partial \mathbf{B}_-}{\partial t} - \mathbf{M}_-, \tag{34}$$

$$\nabla \times \mathbf{H}_- = \frac{\partial \mathbf{D}_-}{\partial t} + \mathbf{J}_-, \tag{35}$$

and the 0-components are

$$\begin{aligned}\frac{\partial \mathbf{B}_0}{\partial t} &= -\mathbf{M}_0, &(36)\\ \frac{\partial \mathbf{D}_0}{\partial t} &= -\mathbf{J}_0. &(37)\end{aligned}$$

Further, inserting $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_+ + \mathbf{D}_-$ and $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_+ + \mathbf{B}_-$ into equations 20-21 gives

$$\begin{aligned}\nabla \cdot \mathbf{D}_0 &= \rho, &(38)\\ \nabla \cdot \mathbf{B}_0 &= \rho_m. &(39)\end{aligned}$$

Equations 32-39 constitute the decomposed Maxwell's equations. These equations give an alternative, but completely equivalent formulation for non-relativistic electromagnetism in $\mathbb{R}^3$. Here, of course, when we say electromagnetism, we mean it in the broad sense, and not as the theory of the *electric* and the *magnetic* field. These fields are not present in the above equations. In fact, none of the original fields $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}, \mathbf{J}$ or $\mathbf{M}$ are present in equations 32-39. Instead, each of these have split into three components, and each component is governed by its own set of equations: the $+$-components are governed by equations 32-33, the $-$-components are governed by equations 34-35, and the 0-components are governed by equations 36-39. One interpretation of the above is that the fundamental quantities in electromagnetic field theory are not the 6 vector fields $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{H}$, $\mathbf{J}$, and $\mathbf{M}$, but the 18 decomposed fields $\mathbf{E}_\lambda, \mathbf{D}_\lambda, \mathbf{B}_\lambda, \mathbf{H}_\lambda, \mathbf{J}_\lambda$, and $\mathbf{M}_\lambda$. By Section 3.4, we know that all the decomposed fields depend only on 18 complex functions in Fourier space. Since the decomposed fields are real valued, it follows that the decomposed fields have the same degrees of freedom as the original fields, which depend on 18 real scalar coordinate functions in $\mathbb{R}^3$. However, for the decomposed fields each of the 18 components have a clear physical interpretation. This is not true for the 18 Cartesian component functions of the original fields since the choice of coordinate axes is arbitrary, i.e., does not depend on physics.

We now see that the decomposed Maxwell's equations for the decomposed fields give a much more detailed view of electromagnetism than the traditional Maxwell's equations. For instance, we immediately see the handed nature of electromagnetism. The fields with positive helicity are governed by a different sets of equations than the fields with negative helicity. Structurally, these sets of equations are both identical to the traditional Maxwell's equations 18-19. Hence electromagnetism is symmetric with with respect of the $+$- and the $-$-components of the fields, or, alternatively, electromagnetism does not prefer one handedness over the other. Another important observation is that there is no coupling between the different sets of equations. For instance, the fields $\mathbf{E}_+, \mathbf{D}_+, \mathbf{B}_+, \mathbf{H}_+$ do not depend on $\mathbf{E}_-, \mathbf{D}_-, \mathbf{B}_-, \mathbf{H}_-$, which physically means that these fields propagate independently of each other; the fields with positive helicity do not "see" the fields with negative helicity. This is in sharp contrast to the traditional Maxwell's equations in Cartesian coordinates, where the curl operator couples the x, y and z components of the fields [31]. No such coupling exist between the $+$, $-$ and 0 fields in the decomposed Maxwell's equations.

From the decomposed Maxwell's equations, it can also be seen that the decomposed components of the fields are completely determined by the corresponding components of the sources. This result can be interpreted trough *Curie's principle*. It is a general principle in science, which states that a symmetry in the effect can be traced back to a symmetry in the cause [3].

Since the 0-components of the electromagnetic fields are curl-free, we can identify them with the static fields, i.e., the fields that do not radiate. Correspondingly, the $\pm$-components are the radiating fields.

The main disadvantages of the decomposition is that it does not preserve the support of fields. For instance, even if $\mathbf{J}$ is non-zero only in some small region of $\mathbb{R}^3$ (for instance inside an antenna), the decomposed fields $\mathbf{J}_\lambda$ can be non-zero in all of $\mathbb{R}^3$ (see Section 3.2.3). The same is true for, say, the electric field in a waveguide. For sources this is problematic since the Green's dyad is singular in the origin [5]. We will not consider this problem.

The conservation of charge is obtained in the usual way. Taking the divergence of equations 36-37, and using equations 38-39 yields

$$
\begin{aligned}
\nabla \cdot \mathbf{J}_0 &= -\frac{\partial \rho}{\partial t}, \\
\nabla \cdot \mathbf{M}_0 &= -\frac{\partial \rho_m}{\partial t}.
\end{aligned}
$$

In particular we see that the conservation of charge only involves the 0-components of the sources [31].

It is of some interest to note that the decomposed Maxwell's equations do not involve the fields $\mathbf{E}_0$ and $\mathbf{H}_0$. This is because they are present in the original Maxwell's equations only behind the curl operator, which maps them to zero. These fields are determined by the constitutive equations, which are studied in the next section.

Since we have not introduced any constitutive equations, the helicity decomposition of Maxwell's equations is valid in *any* media. It is therefore motivated to say that the helicity decomposition in Maxwell's equations is a topological decomposition (see Section 1.2).

The helicity decomposition can also be performed on the time harmonic Maxwell's equations. To do this, it is, however, first necessary to generalize the helicity decomposition to complex valued vector fields. For such vector fields, helicity should be defined as $\mathscr{H}(\mathbf{F}) = \int_{\mathbb{R}^3} \mathbf{F} \cdot \nabla \times \mathbf{F}^* d\mathbf{x}$. We shall not study this decomposition.

4.3 Decomposition of constitutive equations

In the previous section we saw that using the helicity decomposition, Maxwell's equations decompose into three uncoupled parts. This result was independent of any choice of media. We also saw that there were numerous advantages of treating these decomposed fields as fundamental quantities in electromagnetism. It is therefore also motivated to seek a formulation for the constitutive equations in terms of these fields. Ideally, such a formulation could give qualitative information about the coupling of say $\mathbf{D}_-$ and $\mathbf{E}_+$ in different scattering problems. However, even for simple geometries such as a dielectric sphere, it seems to be very difficult to find such a formulation for the constitutive equations. For instance, if $\mathbf{D} = \epsilon(\mathbf{x})\mathbf{E}$, where $\epsilon(\mathbf{x})$ is a real function, then

$$\mathbf{D}_\lambda = \pi_\lambda(\epsilon(\mathbf{x})\mathbf{E}_+) + \pi_\lambda(\epsilon(\mathbf{x})\mathbf{E}_-) + \pi_\lambda(\epsilon(\mathbf{x})\mathbf{E}_0).$$

From this equation we can only conclude that depending on the properties of $\epsilon(\mathbf{x})$ there might be coupling between $\mathbf{E}_+$, $\mathbf{E}_-$, $\mathbf{E}_0$, and $\mathbf{D}_\lambda$. Unfortunately, this equation gives no deeper insight or qualitative information about the scattering process.

In this section we apply the helicity decomposition to the the constitutive equations in Section 4.1. We shall say that the constitutive equations *decompose*, if $\mathbf{D}_\lambda$ and $\mathbf{B}_\lambda$ only depend on $\mathbf{E}_\lambda$ and $\mathbf{H}_\lambda$ for all λ. Otherwise, we say that the constitutive equations *couple* the decomposition. The result of this section is that in "simple" media e.g. vacuum or bi-isotropic homogeneous media, the constitutive equations decompose, but in more complex media such as bi-anisotropic or non-homogeneous media, the constitutive equations couple the decomposition.

4.3.1 Isotropic medium

In the isotropic non-dispersive homogeneous medium we have by linearity that

$$\mathbf{D}_\lambda = \epsilon\mathbf{E}_\lambda, \tag{40}$$
$$\mathbf{B}_\lambda = \mu\mathbf{H}_\lambda. \tag{41}$$

It follows that in this medium, the medium does not couple the decomposition. In other words, the response of this medium does not depend on the helicity of the fields. Since the helicity decomposition commutes with spatial convolutions, the above also holds in homogeneous time-dispersive media. However, if the media is non-homogeneous, then, in general, the media couples the decomposition.

4.3.2 Bi-isotropic medium

Instead of applying the helicity decomposition directly to equations 28-29, we work backwards. We first formulate the constitutive equations for the decomposed fields and then show that they are of the same form as equations 28-29.

If we assume that the constants ϵ and μ in equations 40-41 are functions of λ, we obtain the constitutive equations

$$\mathbf{D}_\lambda = \epsilon_\lambda \mathbf{E}_\lambda, \tag{42}$$

$$\mathbf{B}_\lambda = \mu_\lambda \mathbf{H}_\lambda \tag{43}$$

for all λ. These equations contain six real (constant) medium parameters; ϵ_0, μ_0 describe the response of fields with zero helicity, ϵ_+, μ_+ describe the response of the fields with positive helicity, and ϵ_-, μ_- describe the response of the fields with negative helicity. Next we show that if we disregard the 0-components of the fields (which are zero for sourceless problems), then the above equations are, at least formally, of the same form as the bi-isotropic constitutive equations 28-29. This result is only formally true since we must use the same technique of restriction the support of the fields in Fourier space as in Section 3.1.2.

We begin with the decomposed sourceless Maxwell's equations for the $+$- and the $-$-components of the fields. Assuming that the fields are time-harmonic (with the time convention $e^{-i\omega t}$) these equations read

$$\nabla \times \mathbf{E}_\pm = i\omega\mu_\pm \mathbf{H}_\pm, \tag{44}$$

$$\nabla \times \mathbf{H}_\pm = -i\omega\epsilon_\pm \mathbf{E}_\pm. \tag{45}$$

From these we can form Helmholtz's equation for $\mathbf{E}_\pm$ and $\mathbf{H}_\pm$. Repeating the argument from Section 3.1.2, it follows that the Fourier transforms of $\mathbf{E}_\pm$ and $\mathbf{H}_\pm$ are non-zero only when $|\mathbf{k}| = \frac{1}{2\pi}\omega\sqrt{\epsilon_\pm\mu_\pm}$. We therefore define $k_\pm = \omega\sqrt{\epsilon_\pm\mu_\pm}$. For $\lambda = \pm 1$, we then have

$$\begin{aligned} 0 &= \pi_{-\lambda}\mathscr{F}^{-1}\hat{\mathbf{E}}_\lambda \\ &= \frac{1}{2}\mathscr{F}^{-1}\big\{\hat{\mathbf{E}}_\lambda - i\lambda\mathbf{u}_r(\mathbf{k}) \times \hat{\mathbf{E}}_\lambda\big\} \\ &= \frac{1}{2}\mathscr{F}^{-1}\big\{\hat{\mathbf{E}}_\lambda - \frac{\lambda}{k_\lambda}2\pi i\mathbf{k} \times \hat{\mathbf{E}}_\lambda\big\} \\ &= \frac{1}{2}\big(\mathbf{E}_\lambda - \frac{\lambda}{k_\pm}\nabla \times \mathbf{E}_\lambda\big). \end{aligned}$$

The same calculation also holds for $\mathbf{H}_\lambda$. This shows that the decomposed fields are Beltrami fields,

$$\nabla \times \mathbf{E}_\pm = \pm k_\pm \mathbf{E}_\pm, \tag{46}$$
$$\nabla \times \mathbf{H}_\pm = \pm k_\pm \mathbf{H}_\pm. \tag{47}$$

If the media is not homogeneous, the above reasoning in not valid since then the support of the Fourier transform can not be restricted to $|\mathbf{k}| = \frac{1}{2\pi} k_\pm$.

We now proceed and show that equations 42-43 are of the same form as the chiral constitutive equations 28-29. First, combining equations 44-45 and 46-47 we obtain

$$\mathbf{E}_\lambda = \lambda i \eta_\lambda \mathbf{H}_\lambda, \tag{48}$$

where $\eta_\lambda = \sqrt{\mu_\lambda/\epsilon_\lambda}$. Then, from $\mathbf{E} = \mathbf{E}_+ + \mathbf{E}_-$ and $\mathbf{H} = \mathbf{H}_+ + \mathbf{H}_-$ we have

$$\begin{aligned} \mathbf{E}_+ &= \frac{\eta_+}{\eta_+ + \eta_-}(\mathbf{E} + i\eta_- \mathbf{H}), \\ \mathbf{E}_- &= \frac{\eta_-}{\eta_+ + \eta_-}(\mathbf{E} - i\eta_+ \mathbf{H}). \end{aligned}$$

Also, from equation 48 it follows that

$$\begin{aligned} \mathbf{H}_+ &= \frac{-i}{\eta_+ + \eta_-}(\mathbf{E} + i\eta_- \mathbf{H}), \\ \mathbf{H}_- &= \frac{i}{\eta_+ + \eta_-}(\mathbf{E} - i\eta_+ \mathbf{H}), \end{aligned}$$

The above four equations give the decomposed fields in terms of $\mathbf{E}$ and $\mathbf{H}$. From the undecomposed Maxwell's equations we can further express $\mathbf{H}$ using $\nabla \times \mathbf{E}$ and $\mathbf{E}$ using $\nabla \times \mathbf{H}$. This shows that the decomposition, in this case, is local. For instance, the decomposition of $\mathbf{E}$ is completely determined by $\mathbf{E}$ and $\nabla \times \mathbf{E}$.

Inserting the above expressions for the decomposed fields into $\mathbf{D} = \epsilon_+ \mathbf{E}_+ + \epsilon_- \mathbf{E}_-$ and $\mathbf{B} = \mu_+ \mathbf{H}_+ + \mu_- \mathbf{H}_-$ yields

$$\mathbf{D} = \frac{\epsilon_+\sqrt{\epsilon_-\mu_+} + \epsilon_-\sqrt{\epsilon_+\mu_-}}{\sqrt{\epsilon_-\mu_+} + \sqrt{\epsilon_+\mu_-}}\mathbf{E} + i(\epsilon_+ - \epsilon_-)\frac{\sqrt{\mu_-\mu_+}}{\sqrt{\epsilon_-\mu_+} + \sqrt{\epsilon_+\mu_-}}\mathbf{H}, \tag{49}$$
$$\mathbf{B} = \frac{\mu_+\sqrt{\epsilon_+\mu_-} + \mu_-\sqrt{\epsilon_-\mu_+}}{\sqrt{\epsilon_-\mu_+} + \sqrt{\epsilon_+\mu_-}}\mathbf{H} + i(\mu_- - \mu_+)\frac{\sqrt{\epsilon_+\epsilon_-}}{\sqrt{\epsilon_-\mu_+} + \sqrt{\epsilon_+\mu_-}}\mathbf{E}. \tag{50}$$

We now recognize the chiral constitutive equations. We have thus shown that equations 42-43 are of the same form as equations 28-29. The two equations are, however, not equivalent. Equations 49-50 are formulated for the time-harmonic fields and equations 28-29 are formulated for the fields in the time-domain. Since $\epsilon_\pm$ and $\mu_\pm$ are real constants, it follows that in equations 49-50, the time harmonic counterparts of ϵ and μ are real, and the time harmonic counterparts of ξ and ζ are purely imaginary. If we would formulate equations 28-29 and equations 42-43 for the time harmonic fields (whence $\epsilon, \mu, \xi, \zeta$, and $\epsilon_\lambda, \mu_\lambda$ would all be complex constants), then we could probably show that the two formalisms are equivalent. However, introducing complex constants for modeling the medium would take us one step away from the geometry of electromagnetism. For this reason the above approach has been chosen. It does not yield the most general constitutive equations possible, but the constitutive equations (equations 42-43) have a clear geometrical and physical interpretation.

4.3.3 General linear bi-anisotropic medium

The general linear bi-anisotropic medium decomposes as

$$\begin{aligned}\mathbf{D}_\lambda &= \pi_\lambda(\overline{\overline{\epsilon}} \cdot \mathbf{E}) + \pi_\lambda(\overline{\overline{\xi}} \cdot \mathbf{H}),\\ \mathbf{B}_\lambda &= \pi_\lambda(\overline{\overline{\mu}} \cdot \mathbf{H}) + \pi_\lambda(\overline{\overline{\zeta}} \cdot \mathbf{E}).\end{aligned}$$

Here we have only used linearity. Even if we assume that the medium is homogeneous, we can not assume that e.g. $\overline{\overline{\epsilon}}$ commutes with $\overline{\overline{\mathbf{P}}}_\lambda$. Therefore we can not simplify these equations further.

4.3.4 General constitutive equations

The most general constitutive equations (given by equations 22-23) decomposes as

$$\mathbf{D}_\lambda = \mathbf{D}_\lambda(\mathbf{E}_0, \mathbf{E}_+, \mathbf{E}_-, \mathbf{H}_0, \mathbf{H}_+, \mathbf{H}_-), \tag{51}$$

$$\mathbf{B}_\lambda = \mathbf{B}_\lambda(\mathbf{H}_0, \mathbf{H}_+, \mathbf{H}_-, \mathbf{E}_0, \mathbf{E}_+, \mathbf{E}_-). \tag{52}$$

4.4 Scalar formulation for electromagnetics

In this section we use the Moses representation for the decomposed fields to formulate the decomposed Maxwell's equations. Under one assumption (involving linearity and coupling of the fields), we shall see that Maxwell's equations can be written as a set of linear ordinary differential equations in Fourier space. These can be solved using traditional methods. In Section 4.4.1, we use this to derive a generalization of the dual transformation with 16 degrees of freedom. With this dual transformation we draw a connection to symplectic geometry.

We shall use the same notation as in Section 3.4. If $\mathbf{E}$ is a vector field, then $\mathbf{E}_\lambda = \mathscr{F}^{-1}\{e_\lambda \mathbf{Q}_\lambda\}$, so, for λ in $\{0, \pm 1\}$, e_λ are functions $\mathbb{R}^3 \to \mathbb{C}$ that determines $\mathbf{E}_\lambda$. Similarly, we denote by $d_\lambda, b_\lambda, h_\lambda, j_\lambda, m_\lambda$ the functions that determine $\mathbf{D}_\lambda, \mathbf{B}_\lambda, \mathbf{H}_\lambda, \mathbf{J}_\lambda, \mathbf{M}_\lambda$. Then, defining

$$\mathsf{e} = \begin{pmatrix} e_- \\ h_+ \\ e_+ \\ h_- \end{pmatrix}, \quad \mathsf{J} = \left(\begin{array}{cc|cc} & & 1 & \\ & & & 1 \\ \hline -1 & & & \\ & -1 & & \end{array}\right), \qquad \mathsf{f} = \begin{pmatrix} b_+ \\ d_- \\ b_- \\ d_+ \end{pmatrix}, \quad \mathsf{s} = \begin{pmatrix} m_+ \\ j_- \\ m_- \\ j_+ \end{pmatrix},$$

Maxwell's equations take the form

$$-2\pi|\mathbf{k}|\mathsf{J}\mathsf{e} = \partial_t \mathsf{f} + \mathsf{s}, \tag{53}$$

and

$$\begin{aligned} \partial_t b_0 &= -m_0, & (54)\\ \partial_t d_0 &= -j_0, & (55)\\ -2\pi i|\mathbf{k}|b_0 &= \hat{\rho}_m, & (56)\\ -2\pi i|\mathbf{k}|d_0 &= \hat{\rho}, & (57) \end{aligned}$$

where $\hat{\rho}$ and $\hat{\rho}_m$ are the L^2 scalar Fourier transforms of ρ and ρ_m, and ∂_t is the time derivative. This is seen by inserting the Moses representations of the fields into the decomposed Maxwell's equations and using equations 13, 14 and the relations $\mathbf{k} \times \mathbf{Q}_\lambda = -i\lambda|\mathbf{k}|\mathbf{Q}_\lambda$, $\mathbf{Q}_0 = -\mathbf{u}_r(\mathbf{k})$. The variables e, f and s have been chosen so that Maxwell's equations can be written using the 4×4 matrix J. In Section 5.6, we will see that this way of writing Maxwell's equations show a similarity to Hamilton's equations.

4.4.1 A generalization of the dual transformation

In electromagnetics, the traditional *dual transformation* is a linear variation of the field variables determined by 3 complex constants [5, 22]. This transformation can be used to transform solutions from one problem into solutions to other problems. Using the dual transformation, one can, for instance, transform a solution from one medium to another. Next we show that with the above scalar formulation for electromagnetism, we can define a dual transformation with 16 free parameters, which can be functions of the Fourier variable. We also show that this dual transformation takes a very simple form if the transformation matrix is a *symplectic* matrix. In Section 5.6 we will see how these matrices are related to Hamilton's equations.

To perform the dual transformation in equation 53, we make the assumption that we can write $\mathsf{f} = \mathsf{M}\mathsf{e}$ for some invertible 4×4 matrix M possibly a function of $\mathbf{k}$, but not a function of t. A necessary condition for this to hold is that there is no coupling between the 0-components and the $\pm$-components of the fields. In scattering problems, this assumption should hold with good accuracy if the scatterer is much smaller than the wavelength of the wave. Under this assumption, equation 53 reads

$$-2\pi|\mathbf{k}|\mathsf{J}\mathsf{e} = \partial_t \mathsf{M}\mathsf{e} + \mathsf{s}. \tag{58}$$

This is a set of ordinary differential equation which can be solved using traditional methods [31]. If equation 58 holds, then we say that e is the solution corresponding to the source s in the medium M. Next we introduce a transformation matrix T, which we assume is an arbitrary, possibly complex, invertible 4×4 matrix, which can depend on $\mathbf{k}$, but not on t. By writing $\mathsf{e} = \mathsf{T}\mathsf{T}^{-1}\mathsf{e}$, equation 58 can be manipulated into the form

$$-2\pi|\mathbf{k}|\mathsf{J}(\mathsf{T}^{-1}\mathsf{e}) = \partial_t \mathsf{J}\mathsf{T}^{-1}\mathsf{J}^{-1}\mathsf{M}\mathsf{T}(\mathsf{T}^{-1}\mathsf{e}) + \mathsf{J}\mathsf{T}^{-1}\mathsf{J}^{-1}\mathsf{s}.$$

This gives the following duality transformation. If e is the solution corresponding to the source s in the medium M, then the transformed field $\mathsf{e}' = \mathsf{T}^{-1}\mathsf{e}$ is a solution corresponding to the source $\mathsf{s}' = \mathsf{J}\mathsf{T}^{-1}\mathsf{J}^{-1}\mathsf{s}$ in the medium $\mathsf{M}' = \mathsf{J}\mathsf{T}^{-1}\mathsf{J}^{-1}\mathsf{M}\mathsf{T}$.

This dual transformation mathematically describes the duality between polarization and media. It has the following interpretations.

a) If we change the polarization of a solution to Maxwell's equations, then this change must be compensated by a change in the medium.

b) A solution to Maxwell's equations in one media can be transformed to another media, by compensating the polarization of the solution.

Here, by a change in polarization, we mean the general mapping $\mathsf{e} \to \mathsf{T}^{-1}\mathsf{e}$, and by a change in the medium, we means the corresponding mapping for M.

In Section 5.6 we will see that *symplectic matrices* are related to transformations that leave Hamilton's equations form invariant: A real 4×4 matrix T is *symplectic* if $\mathsf{T}\mathsf{J}\mathsf{T}^{\mathrm{T}} = \mathsf{J}$. It is natural to extend this definition to complex matrices with the condition $\mathsf{T}\mathsf{J}\mathsf{T}^{\mathrm{H}} = \mathsf{J}$, where $\mathsf{T}^{\mathrm{H}} = \mathsf{T}^{\mathrm{T}\,*}$. Then, if we assume that T is symplectic, then s' and M' take the simple forms $\mathsf{s}' = \mathsf{T}^{\mathrm{H}}\mathsf{s}$ and $\mathsf{M}' = \mathsf{T}^{\mathrm{H}}\mathsf{M}\mathsf{T}$. These forms are obtained using $\mathsf{J}^{-1} = \mathsf{J}^{\mathrm{T}} = -\mathsf{J}$. In particular, then the transformation rule for M is essentially that of a 2-tensor. That would suggest that the decomposed Maxwell's equations have some relation with symplectic geometry. In the next section we give an introduction to symplectic geometry and the closely related contact geometry.

5 Contact and Symplectic geometry

Contact geometry is the study of *contact structures*. These are certain topological structures that can exist on odd dimensional manifolds. Similarly, *symplectic geometry* is the study of *symplectic structures*. These are also certain topological structures, but these can only exist on even dimensional manifolds. These theories are dual in the sense that they are closely related and have many results in common.

Since both contact and symplectic structures are purely topological structures, they do not depend on any metric structure of the underlying space. Therefore it is not motivated to study these structures using standard vector analysis, where geometry and topology is intertwined. For instance, the curl operator depends on both geometry and topology; the right-hand rule requires a metric, or an orientation, and differentiation requires topology. For these reasons, we will use the language of differential forms on manifolds for studying contact and symplectic structures.

We will use the same definition of a manifold as in [45]. An *n-dimensional manifold*, which we denote by M^n (or by M), is a topological Hausdorff space with countable base that is locally homeomorphic to $\mathbb{R}^n$ [45]. In addition, we shall always assume that all transition functions are C^∞-smooth. That is, we shall only consider C^∞-smooth manifolds. The space of differential p-forms on M^n is denoted by $\Omega^p(M^n)$, and the tangent space of M^n is denoted by TM^n. When we consider an object at some point x in M^n, we use x as a sub-index on the object. For example, $\Omega^1_x(M^n)$ is the set of 1-forms originating from x. The Einstein summing convention is used throughout.

In this work, we shall hereafter assume that all mathematical objects (e.g. functions, p-forms and vector fields) are C^∞-smooth. This is a standard assumption in differential geometry. However, since the natural function space for electromagnetism is L^2_{curl}, this assumption gives some mathematical problems when studying "contact and symplectic geometry in electromagnetism". We shall not study this problem.

5.1 Contact structures

To define a contact structure, we will need the *Lie bracket*.

Definition 5.1 (Lie bracket) *[37, 46] Let X and Y be two vector fields on a manifold M^n. Then the* Lie bracket *$[X, Y]$ is defined as $[X, Y] = XY - YX$.*

If x^i is are local coordinates for M^n, $X = X^i \frac{\partial}{\partial x^i}$, and $Y = Y^i \frac{\partial}{\partial x^i}$, then

$$\begin{aligned}
XY - YX &= X^i \frac{\partial}{\partial x^i} Y^j \frac{\partial}{\partial x^j} - Y^k \frac{\partial}{\partial x^k} X^l \frac{\partial}{\partial x^l} \\
&= X^i \frac{\partial Y^j}{\partial x^i} \frac{\partial}{\partial x^j} + X^i Y^j \frac{\partial^2}{\partial x^i x^j} - Y^k \frac{\partial X^l}{\partial x^k} \frac{\partial}{\partial x^l} - Y^k X^l \frac{\partial^2}{\partial x^k x^l} \\
&= X^i \frac{\partial Y^j}{\partial x^i} \frac{\partial}{\partial x^j} - Y^k \frac{\partial X^l}{\partial x^k} \frac{\partial}{\partial x^l},
\end{aligned}$$

since partial derivatives commute. In particular, $[X, Y]$ is a vector field.

If α is a 1-form and X and Y are vector fields on an n-manifold, then

$$d\alpha(X, Y) = X(\alpha(Y)) - Y(\alpha(X)) - \alpha([X, Y]). \tag{59}$$

If $X = X^i \frac{\partial}{\partial x^i}$ in some local coordinates $x^1, \cdots, x^n$, and if f is a real function, then, by definition, $X(f) = X^i \frac{\partial f}{\partial x^i}$. In [37], equation 59 is proved using Cartan's formula.

Definition 5.2 (Contact structure) *Let M be a three-manifold, and let ξ be a planefield on TM. Then ξ is a* contact structure*, if for each point $p \in M$ there exist some vector fields X, Y (defined in some neighborhood U of p) such that in U,* $\operatorname{span}\{X, Y\} = \xi$*, but $[X, Y] \notin \xi$.*

Definition 5.3 (Planefield) *A* planefield *ξ on a 3-manifold M^3 is a smooth mapping $p \mapsto \xi_p$ defined for all $p \in M^3$ such that ξ_p is a 2-dimensional vector subspace of T_pM^3.*

Example 5.4 On $\mathbb{R}^3$ with coordinates x, y and z, let

$$\begin{aligned}
X &= \frac{\partial}{\partial x}, \\
Y &= \frac{\partial}{\partial y} - x\frac{\partial}{\partial z}.
\end{aligned}$$

Since X and Y are linearly independent, $\dim \operatorname{span}\{X, Y\} = 2$. Hence $\operatorname{span}\{X, Y\}$ is a planefield. Since $[X, Y] = -\frac{\partial}{\partial z}$, $[X, Y]$ is not in $\operatorname{span}\{X, Y\}$, and the planefield $\operatorname{span}\{X, Y\}$ is a contact structure on $\mathbb{R}^3$. $\square$

On a 3-manifold, a two dimensional planefield is locally determined as the kernel of a 1-form. A contact structure ξ that globally can be written as the kernel of a 1-form is said to be *transversally oriented*. Then $\xi = \ker \alpha$ for some $\alpha \in \Omega^1(M^3)$, and α is said to be a *contact form* for ξ. We will only consider such contact structures. Due to the next theorem, this is a standard assumption in contact geometry.

Theorem 5.5 (Frobenius theorem) *[37, 47] Let α be a 1-form on a 3-manifold. The planefield $\xi = \ker\alpha$ is a contact structure if and only if $\alpha \wedge d\alpha$ is nowhere zero.*

Proof. Suppose $\alpha \wedge d\alpha$ is not zero at some point p in M^3. In some neighborhood U of p we can find vector fields X, Y, Z such that, $\ker\alpha = \operatorname{span}\{X, Y\}$ and $Z \notin \ker\alpha$. Then equation 59 implies that $\alpha \wedge d\alpha(X, Y, Z) = -\alpha([X, Y])\alpha(Z)$. Since X, Y, Z are linearly independent, $\alpha \wedge d\alpha(X, Y, Z)$ does not vanish, and $[X, Y] \notin \ker\alpha$ whence $\ker\alpha$ is a contact structure in U.

Conversely, suppose $\ker\alpha$ is a contact structure. Then, in some neighborhood U, there are vector fields X_1, X_2 such that $\operatorname{span}\{X_1, X_2\} = \ker\alpha$, but $[X_1, X_2] \notin \ker\alpha$. We next show that for arbitrary linearly independent vector fields Y_1, Y_2, Y_3 in U, $\alpha \wedge d\alpha(Y_1, Y_2, Y_3)$ does not vanish. Since $TU = \operatorname{span}\{X_1, X_2, [X_1, X_2]\}$, we can set $X_3 = [X_1, X_2]$, and write $Y_i = \sum_{j=1}^{3} c_{ij} X_j$ for some coefficients c_{ij}. Using the *Levi-Civita permutation symbol* ε,

$$\varepsilon_{k_1\cdots k_m} = \varepsilon^{k_1\cdots k_m} = \begin{cases} +1 & \text{when } k_1\cdots k_m \text{ is an even permutation,} \\ -1 & \text{when } k_1\cdots k_m \text{ is an odd permutation,} \\ 0 & \text{when } k_i = k_j, \text{ for some } \ i \neq j, \end{cases}$$

we then have

$$\begin{aligned} \alpha \wedge d\alpha(Y_1, Y_2, Y_3) &= \sum_{i,j,k=1}^{3} c_{1i}c_{2j}c_{3k} \ \alpha \wedge d\alpha(X_i, X_j, X_k) \\ &= \sum_{i,j,k=1}^{3} c_{1i}c_{2j}c_{3k}\varepsilon_{ijk} \ \alpha \wedge d\alpha(X_1, X_2, [X_1, X_2]) \\ &= -\det c_{ij} \ (\alpha([X_1, X_2]))^2. \end{aligned}$$

Since $\dim\operatorname{span}\{X_1, X_2, X_3\} = \dim\operatorname{span}\{Y_1, Y_2, Y_3\} = 3$, $\det c_{ij} \neq 0$ and the claim follows. □

The above theorem shows that if M^3 has a contact structure, then the contact structure induces an orientation on M^3 given by the volume-form $\alpha \wedge d\alpha$. It is then possible to compare *orientations* of contact structures as follows. The contact structures $\ker\alpha$ and $\ker\alpha'$ have the *same orientation*, if $\alpha \wedge d\alpha = f\alpha' \wedge d\alpha'$ for a positive function f. Similarly, $\ker\alpha$ and $\ker\alpha'$ have *opposite orientations*, if $\alpha \wedge d\alpha = f\alpha' \wedge d\alpha'$ for a negative function f. A *positive (negative) function* only takes values greater (smaller) than zero.

If we scale α by some positive or negative function f, then the planefield $\ker\alpha$ is clearly invariant. Since $(f\alpha) \wedge d(f\alpha) = f^2\alpha \wedge d\alpha$, the contact condition is also invariant under scalings of α. In particular, $\ker \alpha$ and $\ker f\alpha$ have the same orientation. However, the induced volume-form $\alpha \wedge d\alpha$ depends on the choice of α.

At this point we have not assumed that M^3 has any geometric structure (say a Riemannian metric). We have only assumed that M^3 has a manifold structure. However, we have seen that if M^3 has a transversally oriented contact structure, then the structure induces an orientation on M^3. Also, if a distinguished contact form is specified, it induces a volume-form on M^3. These are the first geometrical properties of contact structures.

Example 5.6 (Standard structures on $\mathbb{R}^3$) In Example 5.4, we saw that the plane-field spanned by X and Y was a contact structure. This contact structure is transversally oriented. To find a contact form for span$\{X, Y\}$, we let $\alpha = \alpha_x dx + \alpha_y dy + \alpha_z dz$, and put $\alpha(X) = 0$ and $\alpha(Y) = 0$. Then $\alpha_x = 0$ and $\alpha_y = \alpha_z x$, so $\mathrm{span}\{X, Y\} = \ker \alpha$, where

$$\alpha = xdy + dz.$$

Then $\alpha \wedge d\alpha = dx \wedge dy \wedge dz$. If we put

$$\alpha' = xdy - dz,$$

then $\alpha' \wedge d\alpha' = -dx \wedge dy \wedge dz$. By Frobenius theorem $\ker \alpha$ and $\ker \alpha'$ are both contact structures. By comparing their induced volume-forms, we see that they have opposite orientations. Usually, either $\ker \alpha$ or $\ker \alpha'$ are called the *standard contact structure* on $\mathbb{R}^3$. Since we have no reason to prefer one orientation over the other, we call $\ker \alpha$ and $\ker \alpha'$ the *standard contact structures* on $\mathbb{R}^3$.

In Figure 2 the planefields $\ker \alpha$ and $\ker \alpha'$ are plotted in the xy-plane. Since α and α' do not depend on z, the planefields are only plotted for $z = 0$. The plots show the vector spaces that α maps to zero as small tiles. For instance, when $x = 0$, α and α' equal dz and $-dz$. That means that (at $x = 0$) all vectors in the xy-plane are mapped to zero. In the figures, the tiles at $x = 0$ are thus oriented perpendicular to the z-direction. The tiles in the figures do not have an orientation in the sense of a bi-vector [48]. If we scale a contact form by -1, we obtain the same contact structure, it has the same induced orientation, and the same volume-form. $\square$

In the above we have given two conditions for a planefield to be a contact structure; either using the Lie bracket (Definition 5.2) or using a contact form (Proposition 5.5). We next give a another characterization of contact structures; a contact structure is a planefield that is nowhere integrable. This result is also known as Frobenius theorem.

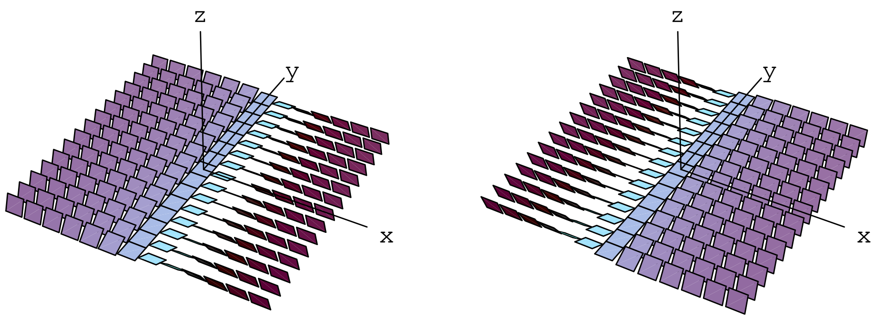


Figure 2: Standard structures on $\mathbb{R}^3$.

Definition 5.7 *Let ξ be a planefield on a 3-manifold M. Then ξ is* integrable *at $p \in M$, if there exists a smooth surface S passing trough p such that ξ is tangential to S in some neighborhood of p. Moreover, ξ is* integrable*, if ξ is integrable at each point of M.*

Theorem 5.8 (Frobenius theorem) *[37] A planefield ξ on a 3-manifold M^3 is a contact structure, if and only if ξ is everywhere non-integrable.*

Proof. We only prove that integrability implies that $\alpha \wedge d\alpha = 0$. The other direction is more technical (see e.g. [37, 47]). Suppose $\ker \alpha$ is integrable at some point $p \in M$. That is, there is a smooth surface S passing trough p such that $TS = \ker \alpha$ in some neighborhood of p. In some (possibly smaller) neighborhood $U \subset M^3$ of p, we can find a function $f : U \to \mathbb{R}$, such that $S \cap U = f^{-1}(0)$. Since f is constant on $S \cap U$, it follows that df is proportional to α, so $\alpha = \lambda df$ for some non-vanishing function $\lambda : U \to \mathbb{R}$. Then $d\alpha = d\lambda \wedge df = \frac{1}{\lambda} d\lambda \wedge \alpha$. Wedge-multiplying by α gives $\alpha \wedge d\alpha = 0$ on $S \cap U$. $\square$

5.2 Contact structures from Beltrami fields

Theorem 5.8 shows that if a planefield is a contact structure, it can not be a tangent to any smooth surface. In consequence, contact structures must be constantly twisting so that the planes, i.e., vector sub-spaces, can not be "stitched" together into a smooth surface. This characteristic twisting can, for instance, be seen in

Figure 2.

In Section 2, we noted that Beltrami fields also posses a characteristic twisting. It is therefore not surprising that there is a connection between Beltrami fields and contact structures. This correspondence is established in [11]. It is shown that every Beltrami field induces a contact structure and a converse: If ξ is a contact structure, then there exist a Riemannian metric and a vector field X (determined up to a scaling) such that $\nabla \times X = X$.

In this section we prove Proposition 5.13, which shows how contact structures can be generated from Beltrami fields on 3-manifolds. For this reason, we must first generalize the definition of the curl operator and Beltrami fields to 3-manifold. On a manifold, the curl operator splits into two operators: the *Hodge star operator* $*$, and the *exterior derivative* d. The Hodge star operator depends only on the Riemannian metric. The exterior derivative, on the other hand, is purely topological and depends only on the differentiable structure of the manifold.

To transform vectors into 1-forms and vice-versa, we use the standard isomorphisms induced by the Riemannian metric $g = g_{ij} dx^i \otimes dx^j$ [37]. By contracting the metric with the vector field $X = X^i \frac{\partial}{\partial x^i}$, we obtain the 1-form $X^\flat = \iota_X g = g(X, \cdot) = g_{ij} X^i dx^j$. This $\flat$-mapping transforms vector fields into 1-forms. More generally, the *contraction* of a vector field X and a tensor (or a differential form) α is defined as $\iota_X \alpha(\cdots) = \alpha(X, \cdots)$. Since g_{ij} is positive definite, the $\flat$-mapping also has an inverse, a $\sharp$-mapping. If $\alpha = \alpha_i dx^i$ is a 1-form, then $\alpha^\sharp = g^{ij} \alpha_i \frac{\partial}{\partial x^j}$, where g^{ij} are the elements of the matrix $(g_{ij})^{-1}$ [37].

Definition 5.9 *[49] Let M^n be a n-dimensional orientable manifold with a Riemannian metric $g = g_{ij} dx^i \otimes dx^j$. The* Hodge star operator *is the linear operator*

$$* : \Omega^p(M^n) \to \Omega^{n-p}(M^n)$$

that maps the basis elements of $\Omega^p(M^n)$ as

$$*(dx^{i_1} \wedge \cdots \wedge dx^{i_p}) = \frac{\sqrt{|g|}}{(n-p)!} g^{i_1 l_1} \cdots g^{i_p l_p} \varepsilon_{l_1 \cdots l_p\, l_{p+1} \cdots l_n} dx^{l_{p+1}} \wedge \cdots \wedge dx^{l_n},$$

where $|g| = \det g_{ij}$.

Generally $** = (-1)^{p(n-p)} id$, where id is the identity operator in $\Omega^p(M^n)$ [49]. In three dimensions, $** = id$ for all $p = 0, \cdots, 3$. In Cartesian coordinates, the metric tensor is $g = dx \otimes dx + dy \otimes dy + dz \otimes dz$, and the Hodge star operator is

$$*dx = dy \wedge dz, \qquad *dy = dz \wedge dx, \qquad *dz = dx \wedge dy.$$

Definition 5.10 *[11, 37] Let M^3 be a Riemannian manifold. The* curl *of the vector field X is the vector field $\nabla \times X$ for which $(\nabla \times X)^\flat = *dX^\flat$.*

Definition 5.11 *[11] Let X be a vector field on a Riemannian 3-manifold M^3. If $\nabla \times X = fX$ for some function $f : M^3 \to \mathbb{R}$, then X is said to be a* Beltrami vector field*. If f is a nowhere zero, then X is said to be a* rotational Beltrami vector field*.*

We see that if X is a Beltrami vector field, then $*dX^\flat = fX^\flat$. Reading this as an equation for the 1-form $X^\flat$, it is motivated to call $X^\flat$ a Beltrami 1-form.

Definition 5.12 *Let M^3 be a Riemannian manifold. A 1-form α on M^3 is a* Beltrami 1-form*, if $*d\alpha = f\alpha$ for some function $f : M^3 \to \mathbb{R}$. If f is nowhere zero, then α is a* rotational Beltrami 1-form*.*

Since $** = id$, the equation $*d\alpha = f\alpha$ implies that $d\alpha = f * \alpha$. This shows how a Beltrami 1-form depend on both the geometry and the topology of the underlying space. The next proposition shows that certain Beltrami fields induce contact structures. Since contact structures depend only on the topology of space, we can say that the induced contact structures extract a topological component from the Beltrami fields.

Theorem 5.13 (Etnyre, Ghrist) *[11] Let M^3 be a Riemannian manifold, and let α be a rotational Beltrami 1-form on M^3. If α is nowhere zero, then α is a contact form on M^3.*

Proof. Since α is a Beltrami 1-form, $*d\alpha = f\alpha$ for some non-zero function $f : M^3 \to \mathbb{R}$. Then $\alpha \wedge d\alpha = f\alpha \wedge *\alpha$. Below we show that $\alpha \wedge *\alpha$ vanishes only where α is zero. Then, since α and f are non-vanishing, $\alpha \wedge d\alpha$ is nowhere zero and α is a contact form.

To complete the proof, we show that $\alpha \wedge *\alpha$ is zero only where α is zero. We first show that $\varepsilon_{lmn}\varepsilon^{imn} = 2\delta_l^i$, when the variables range over $1, 2, 3$. Here δ_l^i is the *Kronecker delta symbol*: $\delta_i^j = 1$, when $i = j$, and $\delta_i^j = 0$, when $i \neq j$. For $\varepsilon_{lmn}\varepsilon^{imn}$ to be non-zero, we must have $l = i$. Then, with $l = i$ fixed, m and n can only take two values.

Both $\wedge$ and $*$ are well behaved under coordinate changes [49]. Therefore, it suffices to check the claim for some local coordinates x^1, x^2, x^3. If $\alpha = \alpha_i dx^i$, then

$$\begin{aligned}
\alpha \wedge *\alpha &= \alpha_i \alpha_j \frac{\sqrt{|g|}}{2} g^{jl} \varepsilon_{lmn}\varepsilon^{imn} dx^1 \wedge dx^2 \wedge dx^3 \\
&= \alpha_i \alpha_j g^{ji} \sqrt{|g|} dx^1 \wedge dx^2 \wedge dx^3.
\end{aligned}$$

We next show that $|g| > 0$ and that g^{ij} is positive definite whence the claim follows. We know that g_{ij} is positive definite, or equivalently, all its eigenvalues are positive. Therefore $|g|$ (being the product of the eigenvalues of g_{ij}) is positive. Under inversion, eigenvalues are transformed as $\lambda \mapsto 1/\lambda$. Therefore all eigenvalues of g^{ij} are positive and g^{ij} is positive definite. □

If either f or α vanishes at some points on M^3, these points can be removed. One then obtains a contact structure on the punctured manifold [11]. In Section 2.2, we noted that Beltrami fields are found in numerous areas of physics. If we can assume that these fields do not vanish, then the above theorem shows how they induce contact structures. In Section 6 we use the above theorem to derive contact structures from electromagnetism using the Bohren decomposition.

Example 5.14 (Standard overtwisted contact structures) Let us define

$$\alpha_\pm = \cos(kx)\, dz \pm \sin(kx)\, dy.$$

For these, we have that $*d\alpha_\pm = \pm k\alpha_\pm$, so $\alpha_\pm$ are rotational Beltrami 1-forms. Hence, by Theorem 5.13, it follows that $\ker \alpha_\pm$ are contact structures (unless $k = 0$). We also have that $\alpha_\pm \wedge d\alpha_\pm = \pm k dx \wedge dy \wedge dz$ from which it follows that $\ker \alpha_\pm$ are contact structures with opposite orientations.

The structures $\ker \alpha_\pm$ are the *standard overtwisted contact structures on* $\mathbb{R}^3$. In Figure 3 these are plotted when $k = 1$ and x range from $-\pi$ to π. It should be pointed out that although $\alpha_\pm$ are $2\pi/k$-periodic, $\ker \alpha_\pm$ are only π/k-periodic.

This example suggests that the orientation of a contact structure is related to the handedness of the twisting in the contact structure. This is the main result of the next section. □

5.3 Helicity of contact structures

In this section we show that $\alpha\wedge d\alpha$ is the equivalent to the helicity density $\mathbf{F}\cdot\nabla\times\mathbf{F}$ of a vector field. In Section 2 we noted that $\mathbf{F}\cdot\nabla\times\mathbf{F}$ is a measure of the twisting of a vector field. Therefore, since $\alpha \wedge d\alpha$ can never vanish, i.e., change sign, in a contact structure, we can interpret contact structures as everywhere twisting structures with a constant handedness.

Lemma 5.15 *Let α and β be two 1-forms on a Riemannian 3-manifold. Then $g(\alpha^\sharp, \beta^\sharp)\, dV = \alpha \wedge *\beta$, where dV is the volume form induced by the Riemannian metric.*

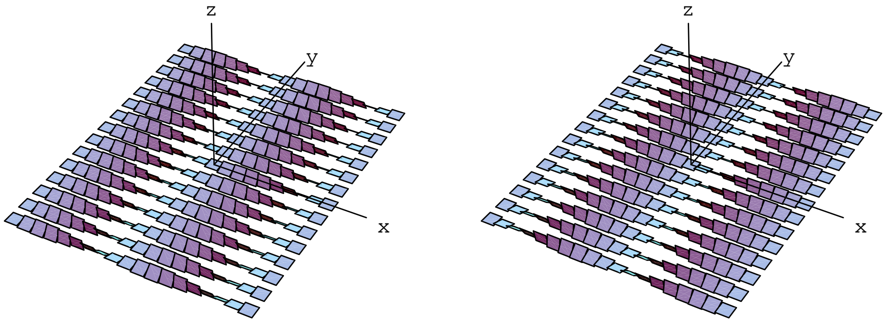


Figure 3: Standard overtwisted contact structures on $\mathbb{R}^3$.

Proof. We only need to check the relation on some coordinate patch $x^1, \cdots, x^3$. Let $\alpha = \alpha_i dx^i$ and $\beta = \beta_i dx^i$. The proof of Theorem 5.13 shows that $\alpha \wedge *\beta = g^{ij}\alpha_i\beta_i dV$. Since $g(\alpha^\sharp, \beta^\sharp) = g^{ij}\alpha_i\beta_i$, the claim follows. $\square$

On a Riemannian manifold, the *dot product* of two vector fields X, Y is defined as $X \cdot Y = g(X,Y)$. Then, from $F \cdot \nabla \times F dV = F^\flat \wedge (\nabla \times F)^\flat = F^\flat \wedge d(F^\flat)$ we see that it is not natural to define helicity for the vector field F. Instead, helicity should be defined for the 1-form $F^\flat$. Then helicity does not depend on the metric.

Definition 5.16 (Helicity) *[23, 50] Let α be a 1-form on a 3-manifold M. The* helicity *of α is defined as*

$$\mathscr{H}(\alpha) = \int_M \alpha \wedge d\alpha.$$

We can now make three observations. First, the above definition of helicity generalizes Definition 2.1: under the assumptions in Definition 2.1, $\mathscr{H}(\alpha^\sharp) = \mathscr{H}(\alpha)$. Second, from Definition 5.16, it follows that helicity is a purely topological measure which does not depend on any metric structure. This is quite easy to understand since helicity can be seen as a generalization of the *writhing number* defined for knots [16]. This writhing number is one measure of how knotted a knot is. Such a measure, if well defined, should not depend on the choice of metric. The third observation is the relation between helicity and contact structures: Frobenius theorem 5.8 states that $\ker \alpha$ is a contact structures if and only if it is constantly twisting, or equivalently, by Theorem 5.5, that $\alpha \wedge d\alpha$ is never zero. Therefore $\alpha \wedge d\alpha$ can be interpreted as a local measure of how much $\ker \alpha$ twists. (See e.g. Example 5.14.) In Section 2 we noted that the sign of $\mathbf{F} \cdot \nabla \times \mathbf{F}$ (the vector counterpart of $\alpha \wedge d\alpha$) indicates whether $\mathbf{F}$ is right-hand rotating or left-hand rotating. Since $\alpha \wedge d\alpha$ can never vanish in a transversally oriented contact structure, a transversally oriented contact structure can only have one type of rotation: either right-hand rotation or left-hand rotation. That is, a transversally oriented contact structure can not change from right-hand rotation to left-hand rotation while being non-integrable at every point.

We previously noted that contact structures possess a characteristic twisting. The above discussion shows that this twisting is further handed. A contact structure is either completely "right-handed" (with positive helicity), or completely "left-handed" (with negative helicity). This relation between contact geometry and helicity was the motivation for studying the helicity decomposition in Section 3.

5.4 Classification of contact structures

A surprising property of contact structures (of same dimension) is that they all locally look the same. This result is known as Darboux's theorem, and its interpretation is that all interesting information about contact structures is of global nature. The study of these global properties is called *contact topology* [13]. Typical problems in contact topology are, for instance, the classification of contact structures on different manifolds or the existence of periodic orbits in contact structures.

In this section we first define a contactomorphism; a mapping that preserves the contact condition. We then state Darboux's theorem, which shows the local invariance of contact structures. In Subsections 5.4.1-5.4.2 we show two methods to distinguish global properties of contact structures, and in Subsection 5.4.3 we list two classification results for contact structures on $\mathbb{R}^3$ and on the 3-torus.

Definition 5.17 *[12] Let ξ be a contact structure on a 3-manifold M, and let η be a contact structure on a 3-manifold N. The structures ξ and η are* contactomorphic *if there exists a diffeomorphism $f : M \to N$ such that $f_*\xi = \eta$. Then f is a* contactomorphism.

In the above definition f_* is the *push-forward* of the map $f : M \to N$ [37]. It maps vector fields on M to vector fields on N. The push-forward map naturally extends to planefields on M. If $\xi = \text{span}\{X, Y\}$, then $f_*\xi = \text{span}\{f_*X, f_*Y\}$. If $M = N$, then $\xi = \ker \alpha$ and $\eta = \ker \beta$ are contactomorphic if and only if $f^*\alpha = \lambda\beta$ for some non-vanishing function $\lambda : M \to \mathbb{R}$ [51]. Here f^* is the *pull-back* of f [37]. The pull-back of a function $f : M \to N$ transforms forms on N to forms on M.

Theorem 5.18 (Darboux's theorem) *[51] Let ξ and ξ' be contact structures on two 3-manifolds M and N. Then ξ and ξ' are locally contactomorphic.*

The above result states the following. If $x \in M$ and $y \in N$, then there exist some neighborhoods $U \subset M$ ($x \in U$) and $V \subset N$ ($y \in V$) and a diffeomorphism $f : U \to V$, such that $\xi|_U$ is contactomorphic to $\eta|_V$. Here, $\xi|_U$ is the *restriction* of ξ to U. Darboux's theorem, for instance, states that any contact structure on a 3-manifold is locally contactomorphic in an orientation preserving way to one of the standard contact structures on $\mathbb{R}^3$. This is quite different from Riemannian geometry, where it does *not* hold that every Riemannian metric is locally Cartesian [12]. The next example shows two contact structures on $\mathbb{R}^3$ which are globally contactomorphic to the standard structures on $\mathbb{R}^3$.

Example 5.19 (Cylinder symmetric standard structures on $\mathbb{R}^3$) If

$$\alpha'_\pm = dz \pm \frac{1}{2}\big(xdy - ydx\big)$$

on $\mathbb{R}^3$, then $\alpha'_\pm \wedge d\alpha'_\pm = \pm dx \wedge dy \wedge dz$, so $\ker \alpha'_+$ and $\ker \alpha'_-$ are contact structures with opposite orientations. These are called the *cylinder symmetric contact structures on* $\mathbb{R}^3$. In cylinder coordinates $\{r, \theta, z\}$, $\alpha'_\pm = dz \pm r^2 d\theta$. The structures $\ker \alpha'_+$ and $\ker \alpha'_-$ are plotted in Figure 4.

The pull-back of the mapping $f_\pm : (x, y, z) \mapsto (\pm x, y, -\frac{1}{2}xy \pm z)$ maps $xdy \pm dz$ to $\alpha'_\pm$. Hence $\ker \alpha'_\pm$ are contactomorphic to the standard structures on $\mathbb{R}^3$. $\Box$

5.4.1 Tight and overtwisted contact structures

In the classification of 3-dimensional contact structures, contact structures have divided into two classes: *overtwisted contact structures* and *tight contact structures*. Of these, tight contact structures seem to reveal more information about the topology of the underlying space, whereas overtwisted structures are more trivial [12]. For instance, on an arbitrary manifold, every homotopy class (see next subsection) has at least one overtwisted contact structure. It also holds that every 3-manifold has at least one overtwisted contact structure [12]. Also, on $\mathbb{R}^3$, there exists enumerable many overtwisted structures, but only two tight structures.

Physically, tight structures also seem to be of more interest. One conjecture states that tight structures are always related to physical solutions that minimize energy,

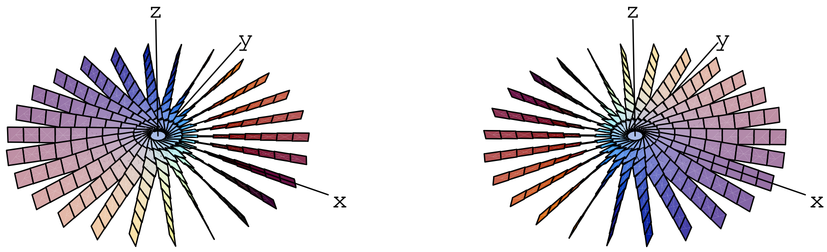


Figure 4: Cylinder symmetric contact structures on $\mathbb{R}^3$.

whereas the energy of overtwisted structures can always be reduced [11]. Any analogous division for contact structures in higher dimensions or for symplectic structures has not been found.

Definition 5.20 (Tight, Overtwisted) *Let ξ be a contact structure on a 3-manifold M^3. Then ξ is* overtwisted *if it has an* overtwisted disc $D \subset M^3$. *Such an overtwisted disc satisfies two properties: D is the image of an 2-disc $D_2 \subset \mathbb{R}^2$, such that the mapping $D_2 \to D$ is an embedding, and the boundary ∂D is tangential to ξ. If ξ is not overtwisted, it is* tight.

Definition 5.21 *[45] The mapping $f : N \to M$ between manifolds N and M is an* embedding, *if $f(N)$ is a smooth submanifold of M, and $f : N \to f(N)$ is a diffeomorphism.*

Theorem 5.22 *The contact structures in Example 5.14 are overtwisted, the contact structures in Examples 5.6 are tight.*

To prove the first claim, we show that a 2-disc can be mapped into $\mathbb{R}^3$ such that its image is tangential to the contact structure on the right side in Figure 3. One possible overtwisted disc is sketched in Figure 5. More precisely, it shows the projection of an overtwisted disc onto the xy-plane. The shaded rectangles represent the contact structure where it is perpendicular to the z-direction. In this figure, the two curves parallel to the x-axis are "horizontal", i.e., they are straight lines that have no z-component. The two curves parallel to the y-axis are straight lines moving "downhill" in the counter-clockwise direction. To make the curve into a closed curve, this downhill motion is compensated in the corners, where the curve moves "uphill" (in the counter-clockwise direction). It follows that the standard overtwisted contact structures are overtwisted. The proof that no such disc exists for the standard contact structures, i.e., that these structures are tight, is highly nontrivial and can be found in [52, 53].

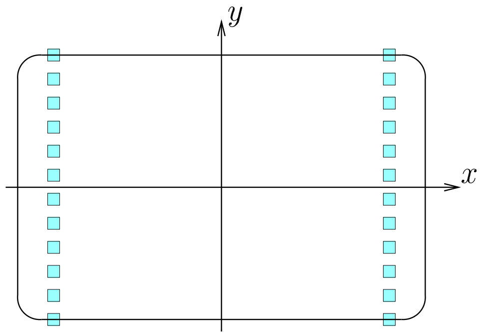


Figure 5: An overtwisted disc.

5.4.2 Homotopy classes for contact structures

Since all contact structures on 3-manifolds locally look the same, contact structures can only be distinguished by their global behavior. To illustrate this, let us consider an example. If we compare the standard structures in Figure 2 to the overtwisted structures in Figure 3, it is motivated to say that the overtwisted structures rotate more than the standard structures. However, this rotation can only be seen globally. If we let $z \to 0$ in $\cos(x)dz \pm sin(x)dy$, and use first order approximations for $\sin$ and $\cos$, we obtain $dz \pm xdy$; the standard structures on $\mathbb{R}^3$. Thus, up to a rotation in the yz-plane, the overtwisted structures locally look like the standard structures on $\mathbb{R}^3$. It is also of interest to note that in the limit process, the Beltrami condition is lost, but the contact condition is preserved. In other words, $dz \pm xdy$ is a contact form, but not a not a Beltrami 1-form (in the Cartesian metric). This result is compatible with an observation made in [11], where contact structures are described as structurally stable.

To distinguish the global properties of planefields we next present the division of planefields into *homotopy classes*. Intuitively, we define a mapping f that maps the normals of a planefield onto S^2 (the unit sphere in $\mathbb{R}^3$). By studying how S^2 is covered under this mapping, we can read off how much the planefield rotates.

For simplicity we only consider a planefield $\ker \alpha$ defined on $\mathbb{R}^3$ with Cartesian coordinates and metric. The mapping $f : \mathbb{R}^3 \to S^2$ is defined as follows. Let

x be a point in $\mathbb{R}^3$. Then $n_x = \alpha_x^\flat$ is normal to $\ker \alpha_x$. In $\mathbb{R}^3$, with Cartesian coordinates, we can project n_x to the origin with respect of the parallel transport. Let this projection be $\pi : T\mathbb{R}^3 \to T_0\mathbb{R}^3$. Then $f(x) = \frac{\pi(n_x)}{|\pi(n_x)|} \in S^2$. Since $\ker \alpha$ is a planefield, the Euclidian norm $|\pi(n_x)|$ does not vanish, and f is well defined. Further, since α is a smooth 1-form, $f : \mathbb{R}^3 \to S^2$ is a smooth mapping. We will say that f is the *induced mapping* for $\ker \alpha$.

Definition 5.23 (Homotopy) *The planefields ξ and η are in the same* homotopy class*, if there is a smooth mapping $g : \mathbb{R}^3 \times [0,1] \to S^2$ such that $g(x,0) : \mathbb{R}^3 \to S^2$ is the induced mapping for ξ, and $g(x,1) : \mathbb{R}^3 \to S^2$ is the induced mapping for η. We then say that ξ and η are* homotopic.

On an arbitrary manifold homotopy is defined in a similar way. However, then there is no unique way of transporting vectors for one tangent space to another. Mathematically, a *connection* must be chosen to transport vectors from one tangent space to another [45]. On a manifold the definition of homotopy classes is therefore unique only up to a choice of a connection. The actual numbering of the homotopy classes also depends on the trivialization of the tangent bundle.

All the contact structures we have presented so far have in common that their induced maps do not cover all of S^2. Such planefields are said to be of *homotopy class zero*. Similarly, if any point on S^2 is covered at least once, but all points are not covered twice, then the planefield is said to be of *homotopy class one*. Generally, if any point on S^2 is covered at least n times, but all points are not covered $n+1$ times, then the planefield is said to be of *homotopy class n*. It holds that two planefields in homotopy class n and m are homotopic if and only if $n = m$. The next example gives examples of contact structures in higher homotopy classes.

Example 5.24 (Contact structures in homotopy classes $1, 2, 3, \cdots$) Let us first define

$$\alpha_k = \cos(kr)dz + kr\sin(kr)d\theta,$$

where $k \in \mathbb{R}$ and $\{r, \theta, z\}$ are the standard cylinder coordinates. This structure rotates like the structures in Figure 3, but the plane of rotation is now $\{dz, d\theta\}$. In [12], $\ker \alpha_k$ is shown to be a contact structure.

Using $\ker \alpha_k$ and the standard cylinder symmetric structure in Example 5.19, we next construct an enumerable class of contact structures on $\mathbb{R}^3$, which are in the homotopy classes $1, 2, 3, \cdots$. The first two of these are plotted in Figure 6, and generally they are constructed as follows. Inside a cylinder (of radius r_0) the structures equals $\ker \alpha_k$. Further, k is chosen such that the planefield makes an integer number of turns for $r \in [0, r_0]$. (In Figure 6, the structures turn once and twice.) Outside a cylinder (at r_1 slightly larger than r_0), the structure equals $\ker dz + r'^2 d\theta$ (see Example 5.19). Here r' is possibly some reparametrization of r. In the region $r \in [r_0, r_1]$, it is possible to construct a contact structure that smoothly connects the two structures (see e.g. [54]). In the figures this region is left blank.

From the figures, it can be seen that these structures are overtwisted. (An overtwisted disc is $r = r_0$.) We also see that if the planefield turns n times inside the cylinder, then the structure is of homotopy class n. Here we have taken $n > 0$. It is convenient to say that the mirror images of the structures for $n = 1, 2, 3, \cdots$ are in homotopy classes $n = -1, -2, -3, \cdots$. The case $n = 0$ is excluded since then $\ker \alpha_k$ is not a contact structure. The standard overtwisted structures in Example 5.14 are in the homotopy class zero. It follows that on $\mathbb{R}^3$, there are two mirror symmetric overtwisted contact structures in every homotopy class. □

5.4.3 Contact structures on $\mathbb{R}^3$ and T^3

Theorems 5.26 and 5.27 give the classification of all contact structures on $\mathbb{R}^3$ and all tight structures on the 3-torus T^3 (the unit cube with opposite sides identified).

Definition 5.25 (Isotopy, Contact isotopy) *Let ξ_0 and ξ_1 be two plane fields on a 3-manifold M. Further, let $\psi_t : M \to M$ is a diffeomorphism for each $t \in [0, 1]$ such that, in addition, ψ_t is a smooth mapping with respect of t. If $\psi_{t*}\xi_0 = \xi_0$ when $t = 0$, and if $\psi_{t*}\xi_0 = \xi_1$ when $t = 1$, then the planefields ξ_0 and ξ_1 are* isotopic. *Further, if $\psi_{t*}\xi_0$ is a contact structure for all $t \in [0, 1]$, then the contact structures ξ_0 and ξ_1 are* contact isotopic.

Theorem 5.26 (Eliashberg) *[52] If ξ be a contact structure on $\mathbb{R}^3$, then ξ is isotopic to a contact structure in one of the below three classes:*

i) the standard structures given in Example 5.6.

ii) the standard overtwisted structures given in Example 5.14.

iii) the contact structures in homotopy classes $\pm 1, \pm 2, \pm 3, \cdots$ given in Example 5.24.

Theorem 5.27 (Giroux, Konda) *[11] All tight contact structures on T^3 are contact isotopic (up to the choice of a connection on T^3) to $\ker \beta_n$ for some integer $n > 0$, where*

$$\beta_n = \sin(2\pi n z)dx + \cos(2\pi n z)dy.$$

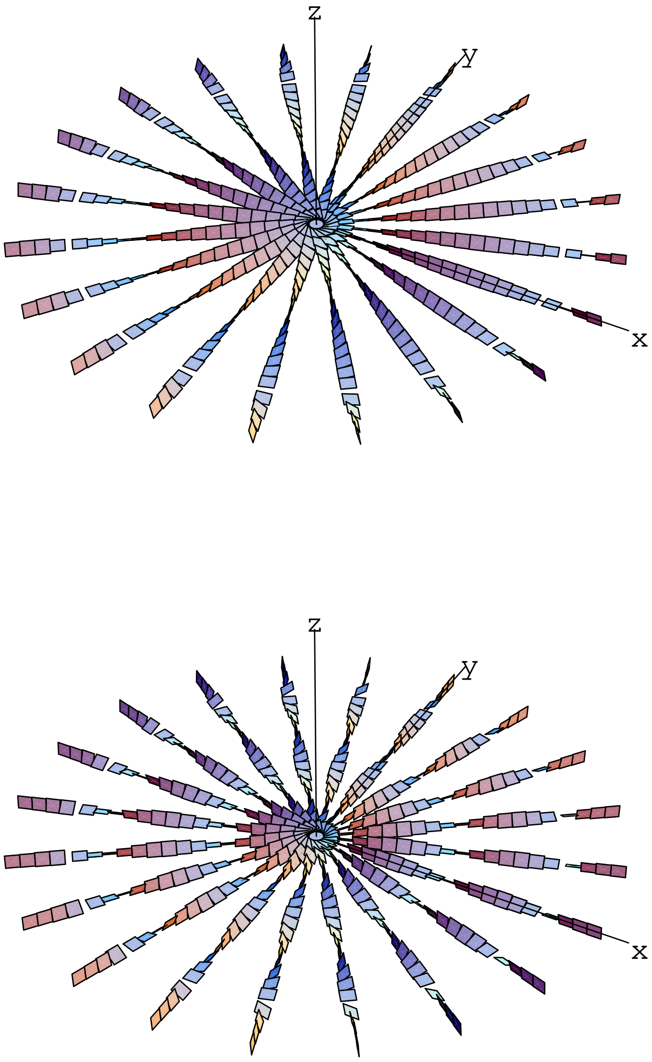


Figure 6: Two structures in homotopy classes 1 and 2..

The 3-torus also have overtwisted contact structures. One example is given by the so called ABC-flows. In numerical studies these show chaotic behavior [23, 50].

5.5 Carnot-Carathéodory metric

In this section we study curves that are always tangential to a contact structure. We will see that any two points on the underlying manifold can be connected by such a curve. Moreover, if the underlying space has a Riemannian metric, then a contact structure induces a natural metric onto the underlying space. This metric is the Carnot-Carathéodory metric.

Theorem 5.28 (Chow's connectivity theorem) *[55] Let $X_1, \ldots, X_m$ be smooth vector fields on a connected manifold M, such that successive Lie brackets of these fields span each tangent space T_pM for p in M. Then any two points in M can be joined by a piecewise smooth curve in M, where each piece is a segment of an integral curve of one of the fields in X_i.*

Corollary 5.29 *Let ξ be a contact structure on a connected 3-manifold M^3. Then any two points in M^3 can be connected by a piecewise smooth curve such that each component is tangential to the contact structure.*

Proof. Let p be a point in M^3. Since ξ is a contact structure, there exist vector fields X, Y such that $\operatorname{span}\{X, Y\} = \xi$, but $[X, Y] \notin \xi$ in some neighborhood U of p. Thus $\operatorname{span}\{X, Y, [X, Y]\} = TU$, and, by Chow's theorem, the claim holds in U. Since M^3 is connected, any two points can be connected by some parametrized curve. Since this curve can be covered by finitely many neighborhoods U as above, the claim follows. □

In the cylinder symmetric contact structure it is easy to check that any two points can be connected using purely geometric reasoning. There are two cases. If the z-components of the given points are equal, they can be connected to the z-axis. If the z components of the given points are not the equal, we can first move the points to the z-axis. Then there are infinitely many ways to connect the points. Figure 7 shows three possible curves.

Suppose ξ is a contact structure on a Riemannian 3-manifold M^3. Then we can use Corollary 5.29 to define a new metric on M^3. If we are given two points on M^3, then they can be connected by some curve tangential to ξ. Since M^3 has a Riemannian metric, we can measure the length of this curve. Further, if there are many ways to connect the two points, we can take the infinum of the lengths of all such curves.

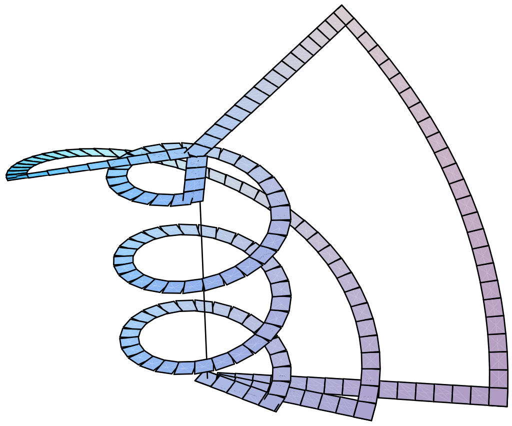


Figure 7: Connecting two points in the cylinder symmetric structure.

Definition 5.30 (Carnot-Carathéodory metric) *[55, 56] Let ξ be a contact structure on a Riemannian 3-manifold. The* Carnot-Carathéodory metric *between the points m, p in M is defined as*

$$\begin{aligned} d(m,p) &= \inf\Big\{ \sum_{i=1}^{M} l(\gamma_i) | \gamma_i \text{ smooth curve tangential to } \xi, \\ &\qquad \gamma_1\gamma_2\cdots\gamma_M \text{ connects } m \text{ and } p \Big\}, \end{aligned}$$

where $l(\gamma)$ is the Riemannian length of the curve γ.

It is possible to show that the Carnot-Carathéodory metric satisfies the axioms for a metric. However, the Carnot-Carathéodory metric is not a Riemannian metric. Due to the infinum in the definition, it is usually only possible to calculate an upper bound for the Carnot-Carathéodory metric.

As the name suggests, the Carnot-Carathéodory metric is related to thermodynamics. The relation is as follows. In Carathéodory's mathematical formalization of thermodynamics [37, 46, 57], it is possible to interpret quasistatic adiabatic processes as curves tangential to a contact structure. The Carnot-Carathéodory metric measures the length of these curves [13].

5.6 Symplectic geometry

The even dimensional dual theory to contact geometry is symplectic geometry. In this section we show how this geometry naturally emerges from the study of Hamilton's equations. We prove that so called *symplectic manifolds* are the only manifolds on which Hamilton's equations are well defined. These manifolds will be the natural generalization of phase-space, i.e., the space of location and momentum vectors in Hamilton mechanics [51, 58]. With this in view, it is easy to understand why symplectic geometry is always even dimensional; each particle has one location vector and one momentum vector of same dimension.

We begin by studying the local Hamilton's equation in $\mathbb{R}^{2n}$,

$$\begin{aligned} \dot{x}^i &= \frac{\partial H(x,y,t)}{\partial y^i}, \qquad i = 1,\dots,n && (60)\\ \dot{y}^i &= -\frac{\partial H(x,y,t)}{\partial x^i}. && (61) \end{aligned}$$

Here x^i are components of the location vectors, y^i are components of the momentum vectors, and $H(x,y,t)$ is a possibly time-dependent *Hamiltonian* which determines the dynamics of the system. Also, t is the time parameter, and $\dot{x}$ is the time derivative of x.

We next write Hamilton's equations with vector notation. For this reason, let z be the $2n$-column vector with components $x^1,\cdots,x^n,y^1,\cdots,y^n$, and let $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, where I and 0 are the $n\times n$ identity and zero matrices. Hamilton's equations then read

$$\dot{z} = J\nabla_z H(z,t), \tag{62}$$

where $\nabla_z H(z,t) = (\frac{\partial H}{\partial z_1},\cdots,\frac{\partial H}{\partial z_{2n}})^T$. We can now make an important observation. Namely, suppose we are given the vector field $X(z,t) = J\nabla_z H(z,t)$. Then the solution curve starting from an initial point $x(0), y(0)$ is completely determined by the equation $\dot{z} = X(z,t)$. We will assume that the equation has a solution, at least for some t [51]. If H is time-independent, then the solution curve is determined as the *flow* of X [37, 51]. The vector field $X(z,t)$ is the *Hamiltonian vector field*.

Intuitively, one can think of the phase-space as filled with a phase fluid. When a particle is dropped into the fluid, the fluid carries the particle forward by its flow. As this particle flows in phase space, the position and momentum vectors for all particles in the physical system can change [51].

5.6.1 Transformations in phase-space: symplectomorphisms

If we perform the change of coordinates $x^i \mapsto -y^i$ and $y^i \mapsto x^i$ in equations 60-61, we see that the equations are form invariant. Transformations with this property are called *symplectomorphisms* (see below). In this section we characterize such transformations and describe their geometry.

The most general transformation in phase space is a diffeomorphism $u : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $u = u(z)$. Under this transformation equation 62 transforms as

$$\begin{aligned} \dot{u} &= \Psi \dot{z} \\ &= \Psi J \nabla_z H(z(u), t) \\ &= \Psi J \Psi^T \nabla_u H(z(u), t) + \Psi J \Theta \frac{\partial}{\partial t} H(z(u), t), \end{aligned}$$

where Ψ is the Jacobian of the transformation, and Θ is the column vector with components $\frac{\partial t}{\partial u^1}, \cdots, \frac{\partial t}{\partial u^{2n}}$. Here u^i is the ith component of $u(z)$. Since Θ is identically zero, equation 62 is equivalent to $\dot{u} = \Psi J \Psi^T \nabla_u H(z(u), t)$.

Definition 5.31 *A $2n \times 2n$-matrix Ψ is a* symplectic matrix*, if $\Psi J \Psi^T = J$. If U and V are open sets in $\mathbb{R}^{2n}$, then a diffeomorphism $f : U \to V$ is a* symplectomorphism *if the Jacobian of f is a symplectic matrix.*

In other words, equation 62 is form invariant if and only if the transformation is a symplectomorphism.

Theorem 5.32 *Symplectic matrices satisfy the following properties:*

a) The determinant of a symplectic matrix equals one.

b) With standard matrix multiplication symplectic matrices form a group.

c) If $\Psi = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, where A, B, C, D are $n \times n$ matrices, then Ψ is symplectic if and only if

$$AD^T - BC^T = I, \quad AB^T = BA^T, \quad CD^T = DC^T.$$

d) If X and Y are real $n \times n$ matrices, then $U = X + iY$ is unitary if and only if $\begin{pmatrix} X & -Y \\ Y & X \end{pmatrix}$ is symplectic.

These properties are proved in [51] and only *a)* is non-trivial. It shows that a symplectic mapping is volume preserving. Property *b)* shows that a composition of symplectomorphisms is a symplectomorphism. It also shows that the inverse of a symplectomorphism is a symplectomorphism. The group of $2n \times 2n$ real symplectic matrices is denoted by $Sp(2n)$. Property *c)* shows that $Sp(2)$ is isomorphic to $SO(2)$. ($SO(n)$ are the real orthogonal $n \times n$ matrices whose determinant is one.) The next theorem further describes the geometry of symplectomorphisms. It shows that an arbitrary volume preserving diffeomorphism need not be a symplectomorphism.

Theorem 5.33 (Gromov) *[51] Let $B^{2n}(R) \subset \mathbb{R}^{2n}$ be the Euclidian ball with radius R and center 0, and let $Z^{2n}(r) = \{x_1, \cdots, x_n, y_1, \cdots, y_n | x_1^2 + y_1^2 \leq r^2\}$ be the* symplectic cylinder. *If $\phi : B^{2n} \to Z^{2n}$ is a symplectomorphism, then $R \leq r$.*

5.6.2 Symplectic manifolds

Definition 5.34 (Symplectic manifold) *Let M be a smooth even dimensional manifold. An atlas of M in which all transition-functions are symplectomorphisms, is a* symplectic atlas. *If a manifold M has a symplectic atlas, then M is a* symplectic manifold*.*

From the previous section it follows that Hamilton's equations are form invariant on a manifold, if and only if the manifold is symplectic. This shows that a symplectic manifold is a natural structure for Hamilton's equations, i.e., we have not imposed any external structure onto Hamilton's equations. We next prove the following mathematical characterization of symplectic manifolds. Usually, this is given as the definition of a symplectic manifold.

Proposition 5.35 *A manifold M^n is a symplectic manifold if and only if there exists a closed non-degenerate 2-form ω on M^n. Then (M^n, ω) is a* symplectic structure*, and ω is a* symplectic form*.*

Definition 5.36 *A 2-form ω on a n-manifold M^n is* non-degenerate*, if $\omega(X, Y) = 0$ for all vector fields X implies that $Y = 0$.*

Geometrically, the symplectic form enables us to measure directed area on the tangent space of the underlying manifold [17]. If a symplectic manifolds is defined as in Proposition 5.35, it follows that symplectic manifolds are even dimensional.

Theorem 5.37 *[51] If a manifold has a closed non-degenerate 2-form, then the manifold is even dimensional.*

The next theorem shows that symplectic manifolds are orientable.

Theorem 5.38 *[51] Let M^{2n} be a even-dimensional manifold, with a 2-form ω. Then ω is non-degenerate if and only if the $2n$-form $\omega^n = \omega \wedge \cdots \wedge \omega$ does not vanish at any point.*

Before we prove Proposition 5.35 we give some examples of symplectic structures.

Example 5.39 (Examples of symplectic structures)

a) Let S be an orientable Riemannian 2-surface. Then S is a symplectic manifold with a symplectic form given by the area form.

b) The *standard symplectic structure* on $\mathbb{R}^{2n}$ is given by

$$\omega = \sum_{i=1}^{n} dx^i \wedge dy^i,$$

where $x^i, \cdots, x^n, y^1, \cdots, y^n$ are coordinates for $\mathbb{R}^{2n}$. It is clear that ω is closed. In [51], it is shown that

$$\omega^n = n!(-1)^{\frac{n(n-1)}{2}} dx^1 \wedge \cdots \wedge dx^n \wedge dy^1 \wedge \cdots \wedge dy^n. \quad (63)$$

c) All manifolds are not symplectic. For instance, S^4 is not. If ω is a symplectic form on S^4, then ω is exact, since the second homology class of S^4 vanishes [45]. In other words, since ω is a closed 2-form, $\omega = d\alpha$ for some 1-form α, and $d(\omega \wedge \alpha) = \omega \wedge \omega$. Since $\omega \wedge \omega$ is a volume-form on S^4, Stokes theorem [45] implies that

$$0 \neq \int_{S^4} \omega \wedge \omega = \int_{\partial S^4} \omega \wedge \alpha.$$

Since S^4 has no boundary, the last integral vanishes, and S^4 can have no symplectic form. □

The proof of Proposition 5.35 is based on the symplectic version of Darboux's theorem and Lemma 5.41 stated below. The symplectic Darboux's theorem, which is also known as Darboux's theorem, has the same content as in contact geometry. It states that all symplectic structures on a $2n$-manifold are locally symplectomorphic to the standard structure on $\mathbb{R}^{2n}$. In consequence, all symplectic structures (of same dimension) are locally symplectomorphic.

Theorem 5.40 (Darboux's theorem, symplectic version) *[51] Let M^{2n} be a manifold with a closed non-degenerate 2-form ω. For any point $x \in M^{2n}$, there exists a chart U ($x \in U$) with coordinates $x^1, \cdots, x^n, y^1, \cdots, y^n$, such that on U*

$$\omega = \sum_{i=1}^{n} dx^i \wedge dy^i.$$

Lemma 5.41 *[51] Let U and V be two open sets in $\mathbb{R}^{2n}$ with coordinates $x^1, \cdots, x^n$, $y^1, \cdots, y^n$ and $\bar{x}^1, \cdots, \bar{x}^n, \bar{y}^1, \cdots, \bar{y}^n$. Further, let $\omega_1 = \sum_{i=1}^n dx^i \wedge dy^i$, and $\omega_2 = \sum_{i=1}^n d\bar{x}^i \wedge d\bar{y}^i$. Then a mapping $f : U \to V$ is a symplectomorphism if and only if $f^*\omega_2 = \omega_1$.*

Proof of lemma. It will be convenient to introduce coordinates $z^1, \cdots, z^{2n}$ for $x^1, \cdots, x^n$, $y^1, \cdots, y^n$ and coordinates $\bar{z}^1, \cdots, \bar{z}^{2n}$ for $\bar{x}^1, \cdots, \bar{x}^n, \bar{y}^1, \cdots, \bar{y}^n$. Then $\omega_1 = \sum_{i=1}^n dz^i \wedge dz^{i+n}$ and $\omega_2 = \sum_{i=1}^n d\bar{z}^i \wedge d\bar{z}^{i+n}$. If $\xi = \xi^i \frac{\partial}{\partial z^i}$ and $\eta = \eta^i \frac{\partial}{\partial z^i}$ are two arbitrary vector fields on U, then

$$\begin{aligned}
f^*\omega_2(\xi, \eta) &= \sum_{i=1}^{n} d(\bar{z}^i \circ f) \wedge d(\bar{z}^{i+n} \circ f)(\xi, \eta) \\
&= \sum_{i=1}^{n} \Psi^i_j \Psi^{i+n}_k dz^j \wedge dz^k(\xi, \eta) \\
&= \sum_{i=1}^{n} \Psi^i_j \xi^j \Psi^{i+n}_k \eta^k - \Psi^{i+n}_k \xi^k \Psi^i_j \eta^j \\
&= \sum_{i=1}^{n} (\Psi\xi)^i (\Psi\eta)^{i+n} - (\Psi\xi)^{i+n} (\Psi\eta)^i \\
&= \xi^T \Psi^T J \Psi \eta,
\end{aligned}$$

where Ψ is the Jacobian of the mapping $f : U \to V$, $\Psi^j_i = \frac{\partial(\bar{z}^j \circ f)}{\partial z^i}$. On the last two lines, we have identified ξ and η with $2n$-column vectors. For instance, $\xi = (\xi^1, \cdots, \xi^{2n})^T$, and $(\Psi\xi)^i$ is the element on the ith row of the column vector $\Psi\xi$. A similar calculation to the above shows that $\omega_1(\xi, \eta) = \xi^T J \eta$. The claim follows by setting $f^*\omega_2 = \omega_1$ for all ξ and η. □

Proof of Proposition 5.35. Suppose M^{2n} is a symplectic manifold, i.e., all transition functions of M^{2n} are symplectomorphisms. Further, suppose U and V are overlapping coordinate patches with local coordinates $\bar{x}^1, \cdots, \bar{x}^n, \bar{y}^1, \cdots, \bar{y}^n$ and $x^1, \cdots, x^n$, $y^1, \cdots, y^n$. Then $\omega_1 = \sum_{i=1}^n dx^i \wedge dy^i$, and $\omega_2 = \sum_{i=1}^n d\bar{x}^i \wedge d\bar{y}^i$ define two forms on U and V. Both ω_1 and ω_2 are closed, they are non-degenerate by equation 63, and by Lemma 5.41, they match on $U \cap V$. Hence the claim follows.

Conversely, suppose that there exists a closed non-degenerate 2-form ω on M^{2n}. By Darboux's theorem, every point of M^{2n} has a neighborhood U with local coordinates $x^1, \cdots, x^n, y^1, \cdots, y^n$ such that $\omega = \sum_{i=1}^n dx^i \wedge dy^i$. These neighborhoods give an atlas to M^{2n}. Since M^{2n} is a manifold, any atlas can be reduced to an atlas with enumerably many charts. All this means that M^{2n} has an atlas such that on each chart, ω is of the form $\omega = \sum_{i=1}^n dx^i \wedge dy^i$. Since ω is globally defined, the transition-functions must preserve ω. Hence, by Lemma 5.41, the transition-functions are symplectomorphisms. $\square$

5.6.3 Hamilton's equations on symplectic manifolds

We can now give a coordinate free formulation for Hamilton's equations on a symplectic manifold. Suppose M^{2n} is a symplectic manifold with a symplectic form $\omega \in \Omega^2(M^{2n})$ and a time-dependent Hamiltonian $H_t : M^{2n} \times \mathbb{R} \to \mathbb{R}$. Then the equation

$$dH = \iota_X \omega \tag{64}$$

fixes the Hamiltonian vector field X. (By non-degeneracy, only the zero vector field is mapped to zero by the mapping $X \mapsto \iota_X \omega$. Hence the mapping is invertible.) Furthermore, if $x^1, \cdots, x^n, y_1, \cdots, y_n$ are local coordinates given by Darboux's theorem, then $\omega = \sum dx^i \wedge dy_i$, and equation 64 is equivalent to the local expression for X, $X(z,t) = J\nabla_z H(z,t)$.

If H is time-independent, then the solution curve in phase-space from some initial value $z_0 \in M^{2n}$ is given by the flow of X. The case when H is time-dependent is considered in [51]. It is also possible to consider time as one coordinate in phase-space. See for instance [10].

5.6.4 Symplectic structure of the cotangent bundle

We next prove Proposition 5.44, which shows that for an arbitrary manifold, its cotangent bundle is always a symplectic manifold. This stems from the fact that the cotangent bundle carries a canonical 1-form known as the *Poincaré 1-form*. This gives the cotangent bundle enough structure to make it into a to symplectic manifold. This is quite different from the tangent bundle, where no such canonical structure has been found [37].

Suppose M is an arbitrary manifold. Its *cotangent bundle* T^*M is the space of linear mappings $TM \to \mathbb{R}$, i.e., one-forms on M. If $x^1, \cdots, x^n$ are local coordinates on a coordinate patch $U \subset M$, then $(dx^1)_r, \cdots, (dx^n)_r$ form a basis for T_r^*U when $r \in U$. That is, if $\alpha \in \Omega_r^1 U$, then $\alpha = \sum_i \alpha_i(r)(dx^i)_r$ for some functions α_i. It follows that $x^1, \cdots, x^n, \alpha_1, \cdots, \alpha_n$ are local coordinates for T^*U. In particular, T^*M is an even dimensional manifold. If V is another coordinate patch overlapping U with coordinates $\bar{x}^1, \cdots, \bar{x}^n$, then α transforms as $\alpha = \alpha_i dx^i = \alpha_i \frac{\partial x^i}{\partial \bar{x}^j} d\bar{x}^j = \bar{\alpha}_j d\bar{x}^j$, where $\bar{\alpha}_j = \alpha_i \frac{\partial x^i}{\partial \bar{x}^j}$.

Proposition 5.42 *Let M be an n-dimensional manifold. Then T^*M carries a coordinate independent 1-form $\alpha \in \Omega^1(T^*M)$. On a coordinate patch $U \subset T^*M$ with local coordinates $x^1, \cdots, x^n, \alpha_1, \cdots, \alpha_n$ (as above), the expression for α is $\alpha = \alpha_i dx^i$.*

Proof. Suppose that V is another coordinate patch overlapping U with coordinates $\bar{x}^1, \cdots, \bar{x}^n, \bar{\alpha}_1, \cdots, \bar{\alpha}_n$ (also as above). Then the transition function from U to V is $x^i \mapsto x^i(\bar{x}^1, \cdots, \bar{x}^n)$ and $\alpha_j \mapsto \alpha_i \frac{\partial x^i}{\partial \bar{x}^j}$. In particular, x^i does not depend on $\bar{\alpha}_1, \cdots, \bar{\alpha}_n$, and $\alpha_i dx^i = \alpha_i \frac{\partial x^i}{\partial \bar{x}^j} d\bar{x}^j = \bar{\alpha}_j d\bar{x}^j$. □

Definition 5.43 *The form α in Proposition 5.42 is called the* Poincaré 1-form.

Proposition 5.44 *The cotangent bundle T^*M of an arbitrary n-manifold M is a symplectic manifold with a symplectic form given by $d\alpha$, where α is the Poincaré 1-form $\alpha \in \Omega^1(T^*M)$.*

Proof. Since $dd = 0$, $d\alpha$ is closed, and we only need to check that $d\alpha$ is non-degenerate on T^*M. If $x^1, \cdots, x^n, \alpha_1, \cdots, \alpha_n$ are local coordinates on T^*M, then $d\alpha = \sum_{i=1}^n dx^i \wedge d\alpha_i$, which, by equation 63, is non-degenerate. □

Proposition 5.44 shows that for an arbitrary manifold, Hamilton's equations can always be formulated on its cotangent bundle. In $\mathbb{R}^{2n}$ the interpretation is that the momentum vectors are actually momentum covectors. For instance, if x_i are the location coordinates, and y_i are momentum coordinates in phase-space $\mathbb{R}^3 \times \mathbb{R}^3$, then the Poincaré form equals $\sum y_i dx_i$; the momentum covector. In traditional vector notation, this is heuristically seen as follows. If U is the potential energy for a particle, then $U = -\mathbf{F} \cdot d\mathbf{r}$, where $\mathbf{F}$ is the force vector for the particle. Hence $\mathbf{F}$ is a 1-form; integrating $\mathbf{F} \cdot d\mathbf{r}$ over a path gives the drop in potential of the particle. The relation between force and momentum is $\mathbf{F} = \frac{d}{dt}\mathbf{p}$. Hence the momentum vector $\mathbf{p}$ should actually be treated as a 1-form [59].

5.7 Relations between contact and symplectic geometry

The next proposition shows how symplectic structures can be generated from contact structures.

Proposition 5.45 *[51] Let* $\ker \alpha$ *be a contact structure on a* 3*-manifold* M*. Then* $d(e^\theta \alpha)$ *is a symplectic form on the* 4*-dimensional manifold* $M \times \mathbb{R}$*, where* θ *is the coordinate on* $\mathbb{R}$*. (Here* α *is written as a form on* $M \times \mathbb{R}$*.)*

Proof. We have $\omega = d(e^\theta \alpha) = e^\theta(d\theta \wedge \alpha + d\alpha)$. Thus,

$$\omega \wedge \omega \quad = \quad e^{2\theta}(2d\theta \wedge \alpha \wedge d\alpha + d\alpha \wedge d\alpha).$$

Since $\alpha \wedge d\alpha$ is never zero and since $d\alpha \wedge d\alpha$ does not contain differentials of θ, the claim follows. $\square$

There are also other relations between contact and symplectic geometry. See for instance references [51, 60].

6 Contact geometry from Helmholtz's equation

In Section 4.4.1 we showed that there is a connection between symplectic geometry and electromagnetism. In this section we take solutions to Helmholtz's equation $\nabla \times (\nabla \times \mathbf{E}) = k^2\mathbf{E}$ and show that the decomposed fields $\mathbf{E}_\pm$ induce contact structures. This suggests that the geometry of electromagnetism is related to both symplectic and contact geometry.

To calculate the decomposed fields from a time-harmonic solutions to Helmholtz's equation $\mathbf{E}$, we shall use equation 9. This equation was used as a model for the helicity decomposition. In the time domain the decomposed fields are then

$$\mathbf{E}_\pm = \frac{1}{2}\Re\left\{\big(\mathbf{E} \pm \frac{1}{k}\nabla \times \mathbf{E}\big)e^{-i\omega t}\right\}. \tag{65}$$

The advantage of using this formula is that it is local. We can therefore apply it to solutions which are not necessarily in L^2_{curl}. (We will, for instance, study contact structures for planewaves.) If $\mathbf{E}$ is a solution to Helmholtz's equation, then $\mathbf{E}_\pm$ are Beltrami fields. If they, in addition, do not vanish at any point, then they induce two contact structures, $\ker{(\mathbf{E}_+)^\flat}$ and $\ker{(\mathbf{E}_-)^\flat}$. In this section, we will always use the Cartesian metric. We therefore make no distinction between vector fields and 1-forms.

This section is organized as follows. In Section 6.1 we derive a local invariance result for Helmholtz's equation. Essentially this result states that any solution to Helmholtz's equation can locally be transformed into any other solution. This is an immediate consequence of Darboux's theorem for contact structures. In Section 6.2 we study contact structures induced by known solutions to Helmholtz's equation. We show that an arbitrary planewave induces two contact structures; one corresponding to the right-hand circulary polarized component and one corresponding to the left-hand circulary polarized component. We also see that for these structures the Carnot-Carathéodory metric seems to describe the path traversed by a ray of light. We also show that solutions in wave-guides induce contact structures.

6.1 Local invariance of Helmholtz's equations

Suppose we have two solutions $\mathbf{E}$ and $\mathbf{E}'$ to the Helmholtz's equation in some region Ω of $\mathbb{R}^3$. By equation 65, these induce four Beltrami fields $\mathbf{E}_\pm$ and $\mathbf{E}'_\pm$. For this section, we shall assume that none of these fields vanish at any point in Ω. Then, by Theorem 5.13, they induce four contact structures. By Darboux's theorem, we know that any two contact structures are *locally* contactomorphic.

Thus, the contact structures induced by $\mathbf{E}_+$ and $\mathbf{E}_-$ are locally contactomorphic to the contact structures induced by $\mathbf{E}'_+$ and $\mathbf{E}'_-$. In addition, since the volume forms $(\mathbf{E}_\pm)^\flat \wedge d(\mathbf{E}_\pm)^\flat$ and $(\mathbf{E}'_\pm)^\flat \wedge d(\mathbf{E}'_\pm)^\flat$ have the same orientation, these contactomorphisms are both orientation preserving. By adding a possible scaling to these contactomorphisms, we can construct mappings $f_\pm$ as in the diagram below.

$$\begin{array}{ccccc} \mathbf{E} = & \mathbf{E}_+ & + & \mathbf{E}_- \\ & \Big\downarrow f_+ & & \Big\downarrow f_- \\ \mathbf{E}' = & \mathbf{E}'_+ & + & \mathbf{E}'_- \end{array}$$

This means that if we have two solutions $\mathbf{E}$ and $\mathbf{E}'$ to Helmholtz's equation whose decomposed fields do not vanish, then locally $\mathbf{E}$ can be transformed into $\mathbf{E}'$. From this result we can make two observations. First, to transform a solution to Helmholtz's equation into another solution, one needs, in general, two mappings; one for $\mathbf{E}_+$ and $\mathbf{E}'_+$, and one for $\mathbf{E}_-$ and $\mathbf{E}'_-$. Second, the above result states that all solutions to Helmholtz's equation are, in some sense, similar to each other. One interpretation is that the contact structures for $\mathbf{E}_+$ and $\mathbf{E}_-$ contain the necessary twisting for the field to radiate, and locally there are only two ways to twist; right twist and left twist.

The present result can also be compared to the *Riemann mapping theorem*. It states that any two simply connected regions in the complex plane (which are not the entire complex plane) can analytically be transformed into each other [61]. This is a powerful tool in electrostatics, which can be used to transform solutions to Laplace's equation from one region to another region.

6.2 Contact structure from solutions to Helmholtz's equation

6.2.1 Planewave

The planewave is the most simple time dependent solution to Maxwell's equations. In this section we show that a planewave induces two contact structures; one from the right-hand circulary polarized (RCP) component and one from the left-hand circulary polarized (LCP) component.

A general planewave solution to Maxwell's equations is of the form

$$\mathbf{E}(z,t) = \Re\left\{\mathbf{A}e^{i(kz-\omega t)}\right\},$$

where $k = \omega\sqrt{\epsilon\mu}$, and $\mathbf{A}$ is a complex constant vector with no z-component. Using equation 8, the decomposed components are then

$$\mathbf{E}_\pm(z,t) \quad = \quad \frac{1}{2}\Re\left\{(\mathbf{A} \pm i\mathbf{u}_z \times \mathbf{A})e^{i(kz-\omega t)}\right\}. \tag{66}$$

Since $\mathbf{E}$ is a solution to Helmholtz's equation, we know that $\nabla \times \mathbf{E}_\pm = \pm k \mathbf{E}_\pm$. If we write $\mathbf{A}_\pm = \frac{1}{2}(\mathbf{A} \pm i\mathbf{u}_z \times \mathbf{A})$, $\mathbf{A}_r = \Re\{\mathbf{A}\}$, and $\mathbf{A}_i = \Im\{\mathbf{A}\}$), then

$$\begin{aligned}\Re\{\mathbf{A}_\pm\} &= \frac{1}{2}\big(\mathbf{A}_r \mp \mathbf{u}_z \times \mathbf{A}_i\big),\\ \Im\{\mathbf{A}_\pm\} &= \frac{1}{2}\big(\mathbf{A}_i \pm \mathbf{u}_z \times \mathbf{A}_r\big),\end{aligned}$$

and $\mathbf{u}_z \times \Re\{\mathbf{A}_\pm\} = \pm\Im\{\mathbf{A}_\pm\}$. We then have that $|\Re\{\mathbf{A}_\pm\}|^2 = |\Im\{\mathbf{A}_\pm\}|^2$ and $\Re\{\mathbf{A}_\pm\} \cdot \Im\{\mathbf{A}_\pm\} = 0$, so the decomposed fields are circulary polarized planewaves. We also have that

$$\mathbf{E}_\pm(z,t) \cdot \nabla \times \mathbf{E}_\pm(z,t) = \frac{\pm k}{8}|\mathbf{A} \pm i\mathbf{u}_z \times \mathbf{A}|^2.$$

In other words, the helicity densities for the decomposed fields are constant and proportional to the energy densities of the decomposed fields. Thus, in general, a planewave induces two contact structures; one for the RCP component and one for the LCP component.

We can now write equation 66 as

$$\begin{aligned}\mathbf{E}_\lambda(z,t) &= \Re\{\mathbf{A}_\pm\}\cos(kz-\omega t) - \Im\{\mathbf{A}_\pm\}\sin(kz-\omega t)\\ &= \Re\{\mathbf{A}_\pm\}\cos(kz-\omega t) \mp \mathbf{u}_z \times \Re\{\mathbf{A}_\pm\}\sin(kz-\omega t).\end{aligned}$$

For instance, if we set $\mathbf{A} = \mathbf{u}_x$, we obtain the fields in Example 2.2. If follows that the contact structures induced by $\mathbf{E}_\pm$ look like the standard overtwisted contact structures in Figure 3, i.e., the contact planes constantly rotate around the direction of propagation. We also see that t does not modify this behavior. To simplify the analysis, we set $t = 0$. We then see that if an RCP (or LCP) planewave passes trough two points, then the path given by the Carnot-Carathéodory metric between these points is the straight line connecting the points. Moreover, the Carnot-Carathéodory metric gives the Cartesian length between these points. Thus, for planewaves the Carnot-Carathéodory metric describes the path traversed by a ray of light.

In Section 5.4.1 we noted that there is a hypothesis which says that tight contact structures are always related to physical solutions that minimize energy. If we assume that the contact structures induced by $\mathbf{E}_\pm$ are defined on $\mathbb{R}^3$, then the structures are overtwisted and the hypothesis does not hold. However, from a physical point of view, planewaves are most naturally defined on the 3-torus. Then the structures become tight and there is no contradiction.

6.2.2 Refraction of a planewave

A basic result in electromagnetics is that when a planewave encounters an infinite dielectric half-space $z < 0$, then part of the wave is reflected from the half-space, and part of the wave is transmitted into the half-space. The angle of reflection is always equal to the angle of incidence. However, the angle of the transmitted wave depends on both the permittivity of the half-space and the polarization of the incident wave [62]. This problem can be solved by decomposing the incident electric planewave into two linearly polarized components $\mathbf{E}_t$ and $\mathbf{E}_p$ as in Figure 8. For these, one can show that the reflected and transmitted fields are linearly polarized planewaves. Moreover, the direction of propagation for the transmitted planewaves are in general different for the two different polarizations. This is illustrated in Figure 8.

In the previous section we saw that any planewave can be decomposed into two circulary polarized components. We can therefore conclude that when a RCP planewave encounters the half-space, four circulary polarized planewaves are transmitted into the media: two RCP waves and two LCP planewaves. This is shown in Figure 9. We will here not consider the refracted wave. It is either RCP or LCP. In particular, this means that the dielectric boundary couples the $+$- and $-$-components in the helicity decomposition.

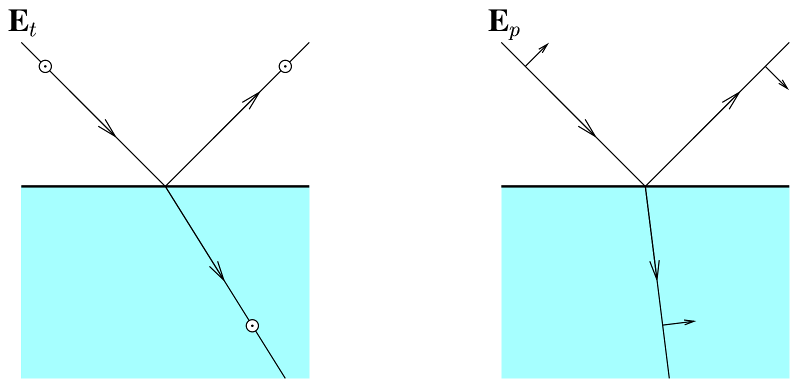


Figure 8: Refraction from the dielectric half-space $z < 0$.

In the previous section we saw that for planewaves, the Carnot-Carathéodory metric described the path traversed by the wave. Let us now consider the Carnot-Carathéodory metric for one transmitted RCP component in Figure 9. Let us first take two points in Figure 9 (one with $z > 0$ and the other with $z < 0$) such that the RCP wave passes trough both points. If the Carnot-Carathéodory metric would describe the path the RCP wave traverses, then then the path given by the

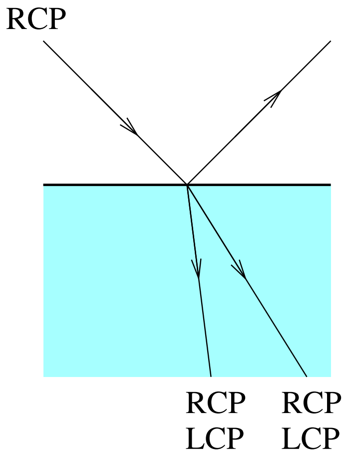


Figure 9: Refraction of a RCP planewave

Carnot-Carathéodory metric should match with the broken line drawn in Figure 9 between the points. This path is tangential to the contact structures induced by the RCP waves (excluding the plane $z = 0$) and hence an admissible path for the Carnot-Carathéodory metric. The problem, however, is that we do not know if there exists another path which is also tangential to the contact structure, but has a shorter length. Unfortunately, it seems to be quite difficult to prove that no such path exists. It would, however, seem very plausible that the shortest distance would be a piecewise straight line. If this is indeed the case, then at least for planewaves, the Carnot-Carathéodory metric would correctly describe the path of a ray of light. In particular, since the RCP and LCP components induce two different contact structures and thus two different Carnot-Carathéodory metrics, these would take into account the different scattering behaviors for different polarizations. Then the path traversed by an RCP or LCP planewave would be a geodesic of the Carnot-Carathéodory metric for the corresponding contact structure.

6.2.3 Rectangular and circular waveguides

Waveguides are metallic structures that are used to guide electromagnetic waves from one point to another. In this section we show that solutions to Helmholtz's equations in these structures also induce contact structures. Our approach in this section will be completely geometrical. We will plot cross-sections of planefields induced by different solutions and see that these planefields show the constant twisting characteristic for contact structures. This shows that contact structures is not something peculiar to only planewaves, but contact structures also exist in more complicated solutions to Helmholtz's equation.

Traditionally solutions in waveguides are divided into two classes: TE and TM solutions. In TE solutions the electric field has no z-component, and in TM solutions the magnetic field has no z-component. Here, the z-axis points in the direction of propagation, i.e., the direction of the waveguide. These solutions exist in different modes, which are enumerated by two integers m and n, and written as TE_{mn} and TM_{mn}.

On the next pages, the $+$-component of the electric field is plotted for different solutions to waveguides. The mathematical expressions for these solutions are taken from [63]. These are all 2π periodic in the z-direction. They are, however, only plotted for $z = 0, \frac{1}{3}\frac{\pi}{2}, \frac{2}{3}\frac{\pi}{2}$, and $\frac{\pi}{2}$ since these plots show the basic twisting behavior for the planefield. All the plots are plotted for $t = 0$. Also, we only plot the $+$-component since the $-$-component is symmetrical; it simply twists with opposite helicity.

Figures 10-25 show the $+$-components of the electric field for the TE_{01}, TE_{11}, TM_{11}, TE_{21} solutions in a rectangular waveguide. To make the plots easier to read, each plot is shown from two different angles. Figures 10-13 show the $+$-component of the electric field for the TE_{01} mode. The expression for these fields are

$$\begin{aligned}\mathbf{E}_{\pm} &= \frac{\pm 1}{2\sqrt{1+\pi^2}}\Big(\frac{1}{\pi}\cos z \sin \pi y\, \mathbf{u}_y - \sin z \cos \pi y\, \mathbf{u}_z\Big) \\ &+ \frac{1}{2\pi}\sin z \sin \pi y\, \mathbf{u}_x.\end{aligned}$$

Using Mathematica it follows that

$$\mathbf{E}_{\pm}\cdot\nabla\times\mathbf{E}_{\pm} = \pm\frac{1}{8\pi^2\sqrt{1+\pi^2}}\big((1-\cos 2\pi y) + \pi^2(1-\cos 2z)\big).$$

Here $x \in (0,1)$ and $y \in (0,1)$ define the interior of the waveguide.

This mathematically shows that the decomposed components of the electric field for the TE_{01} solution induce two contact structures. Figures 14-25 show the $+$-components for other solutions in a rectangular waveguide. From these figures, we see that TE and TM solutions are somehow symmetrical. For instance, the TE_{11} solution is obtained from the TM_{11} solution by shifting the solution in the xy-plane.

Figures 26-37 show planefields for a cylinder symmetric waveguide. Here, we observe the same behavior as in the rectangular case. Only here, the TE solution seem to be identical to the TM solution with a phase-shift in the z direction. This would suggest that (at least from a theoretical point of view) it is more natural to divide the fields inside a waveguide into $+$-solutions and $-$-solutions. The advantage of such a division would be that it would not be based on Cartesian coordinates (see Section 1). Instead, a $+/-$ division would divide solutions in a waveguide into two sets of solutions which propagate independently of each other. This division thus has a physical interpretation. We could also say that the $+/-$ division represents the internal division of the fields in a waveguide whereas the TE/TM division is based on our Cartesian view of electromagnetism.

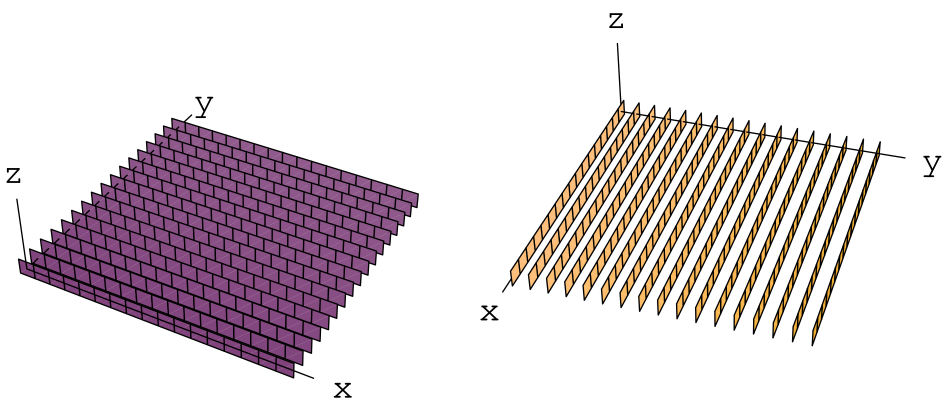


Figure 10: TE_{01} $+$-field at $z = \frac{0}{3}\frac{\pi}{2}$.

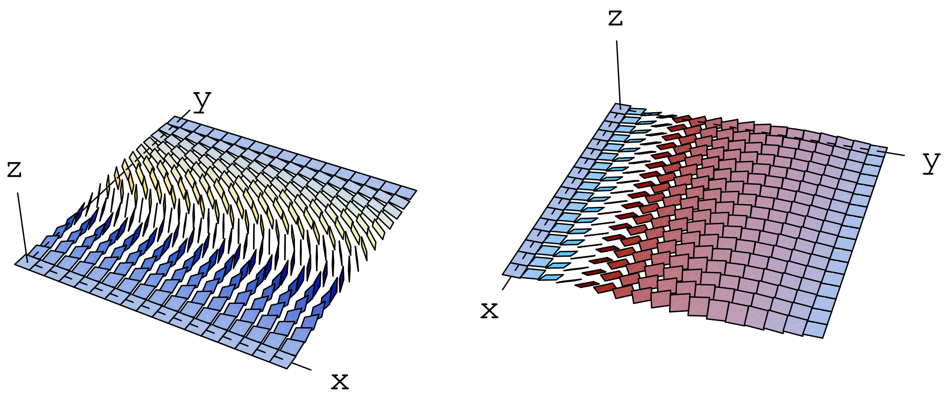


Figure 11: TE_{01} $+$-field at $z = \frac{1}{3}\frac{\pi}{2}$.

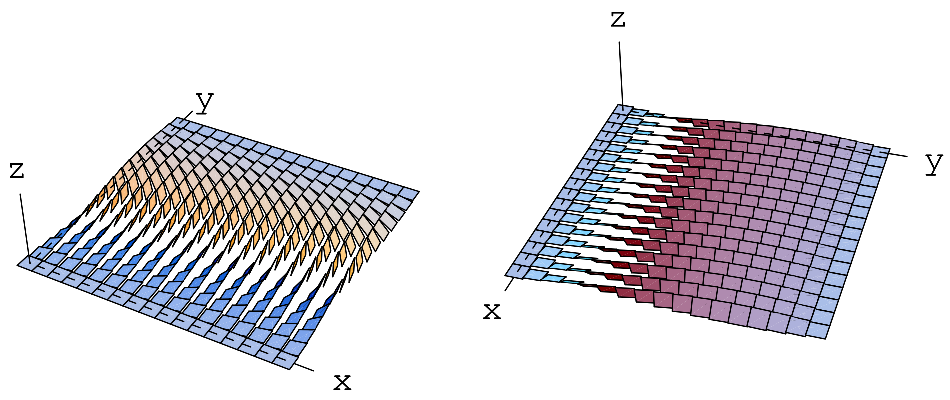


Figure 12: TE_{01} $+$-field at $z = \frac{2}{3}\frac{\pi}{2}$.

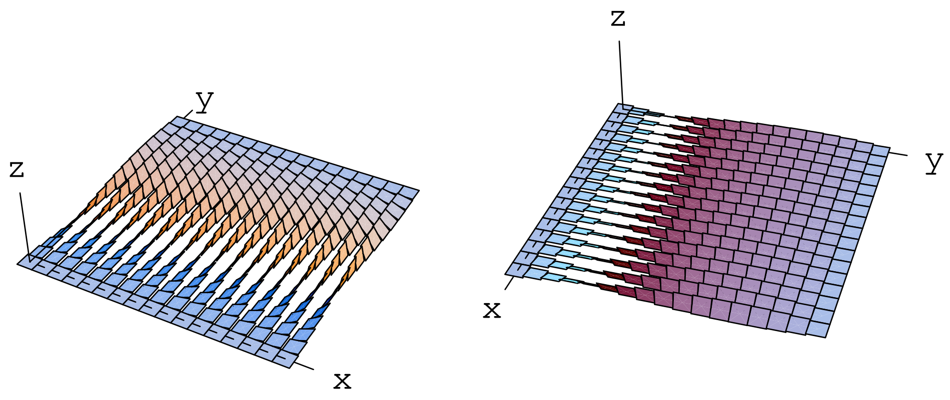


Figure 13: TE_{01} $+$-field at $z = \frac{3}{3}\frac{\pi}{2}$.

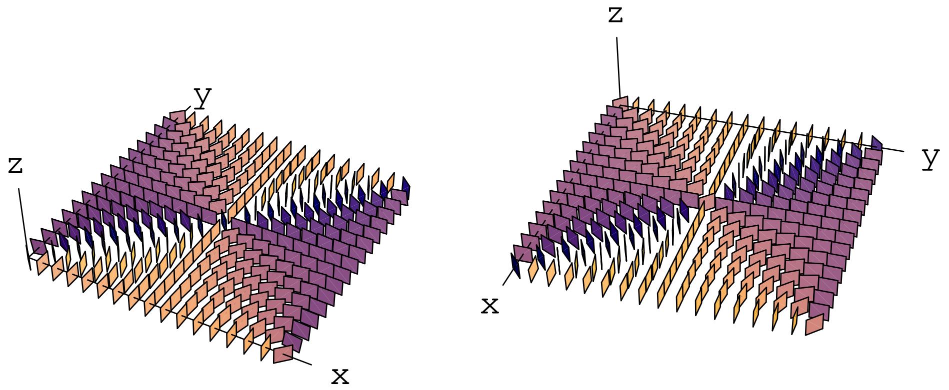


Figure 14: TE_{11} +-field at $z = \frac{0}{3}\frac{\pi}{2}$.

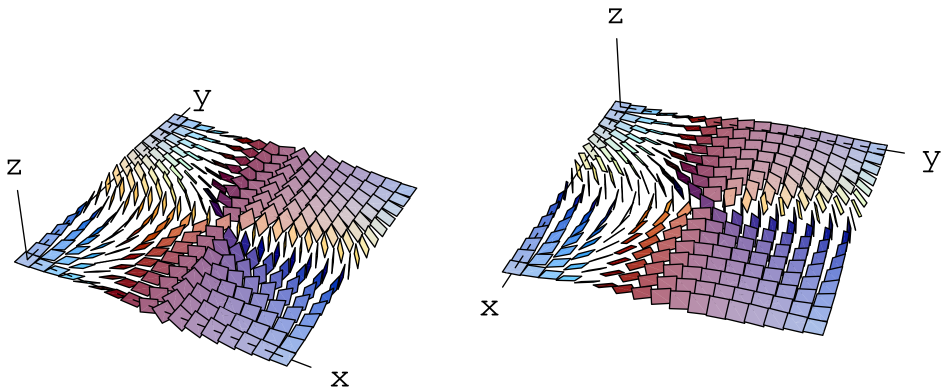


Figure 15: TE_{11} +-field at $z = \frac{1}{3}\frac{\pi}{2}$.

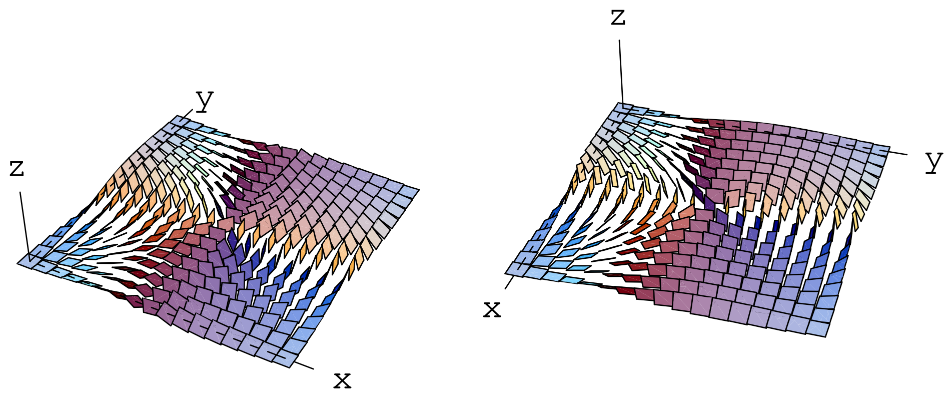


Figure 16: TE_{11} +-field at $z = \frac{2}{3}\frac{\pi}{2}$.

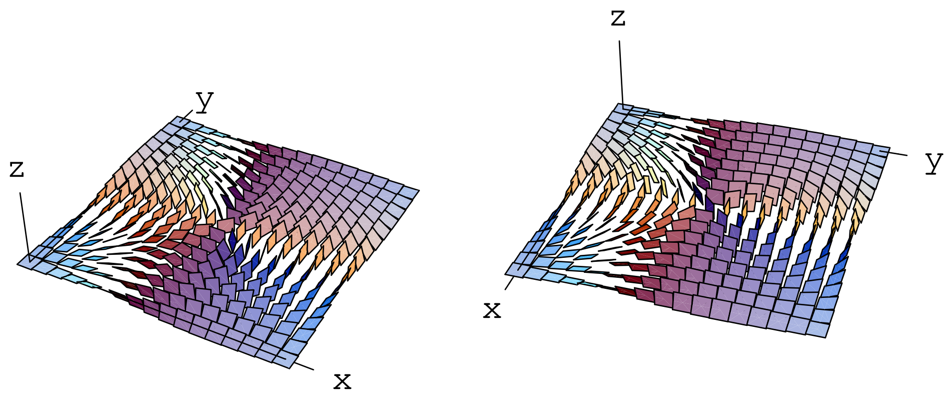


Figure 17: TE_{11} +-field at $z = \frac{3}{3}\frac{\pi}{2}$.

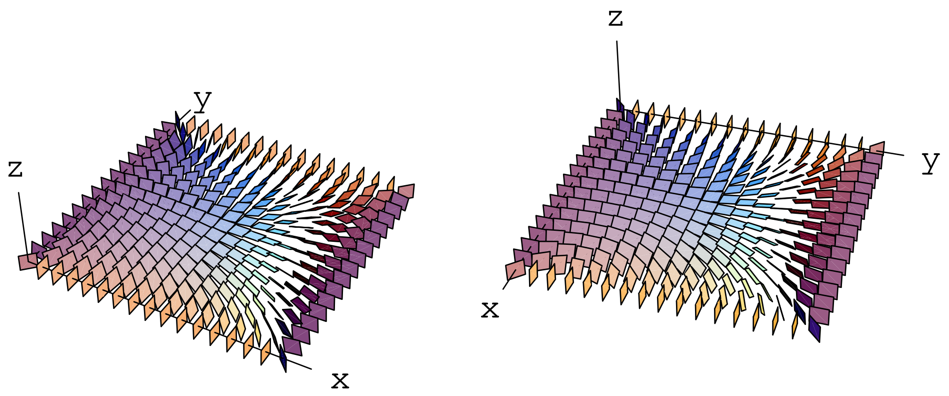


Figure 18: TM_{11} +-field at $z = \frac{0}{3}\frac{\pi}{2}$.

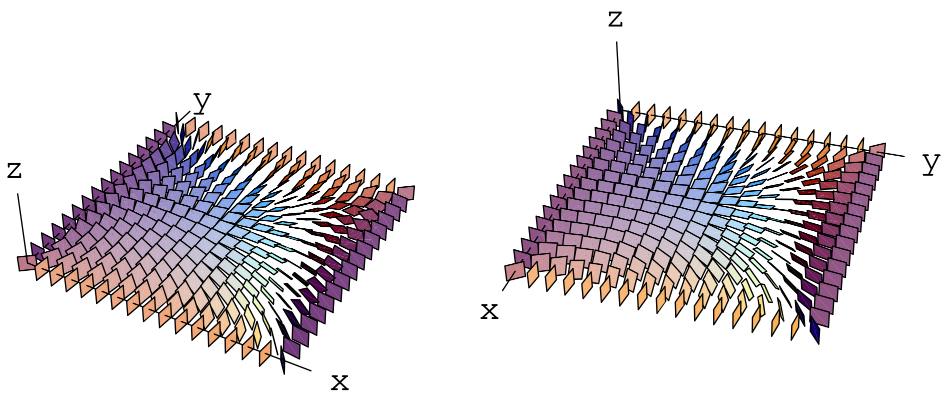


Figure 19: TM_{11} +-field at $z = \frac{1}{3}\frac{\pi}{2}$.

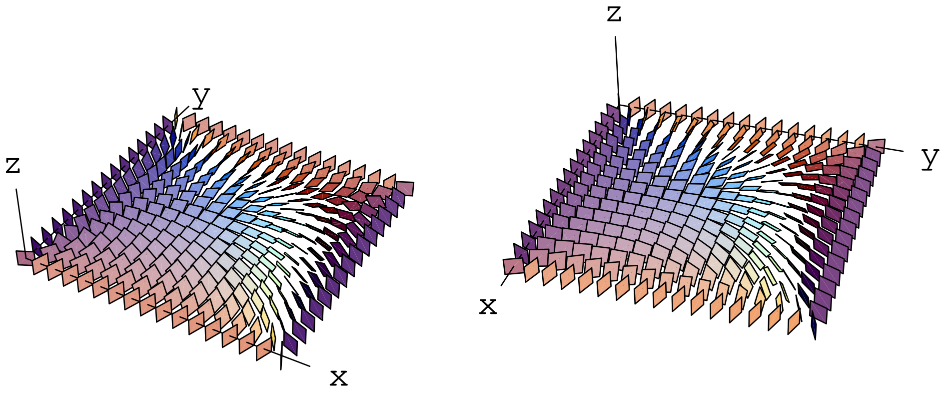


Figure 20: TM_{11} $+$-field at $z = \frac{2}{3}\frac{\pi}{2}$.

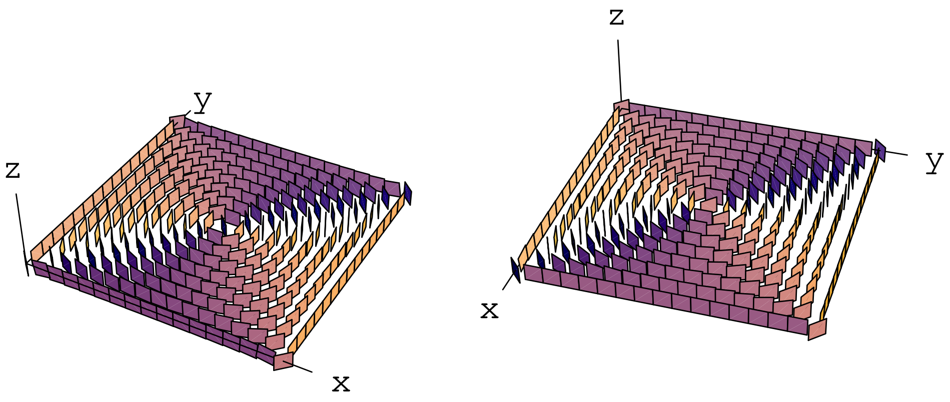


Figure 21: TM_{11} $+$-field at $z = \frac{3}{3}\frac{\pi}{2}$.

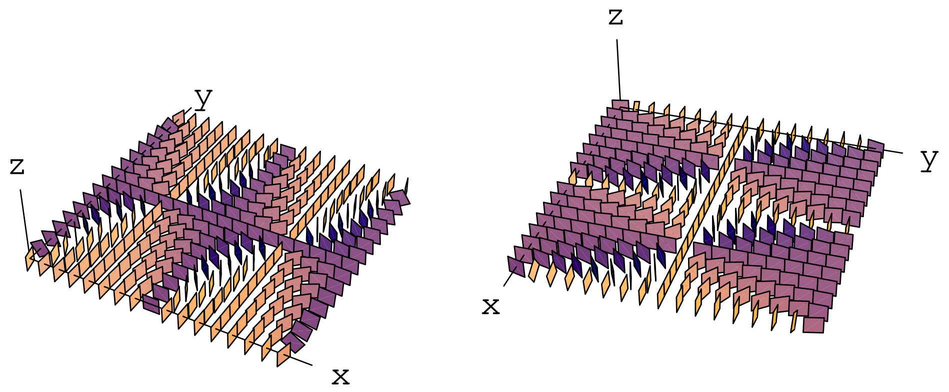


Figure 22: TE_{21} $+$-field at $z = \frac{0}{3}\frac{\pi}{2}$.

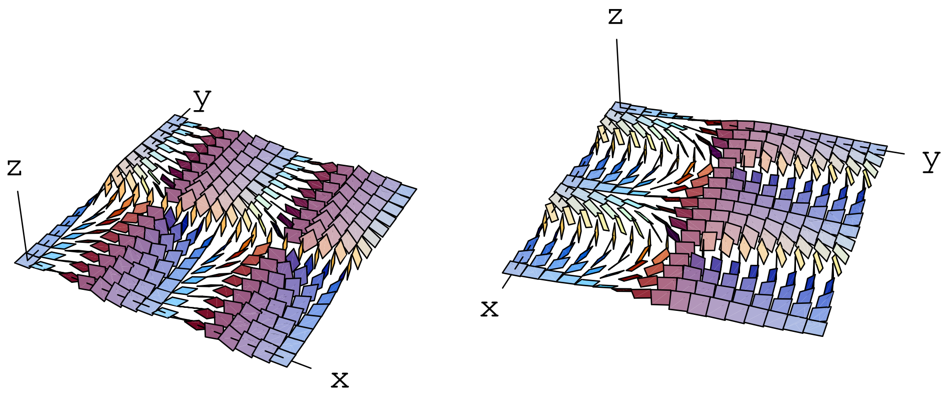


Figure 23: TE_{21} $+$-field at $z = \frac{1}{3}\frac{\pi}{2}$.

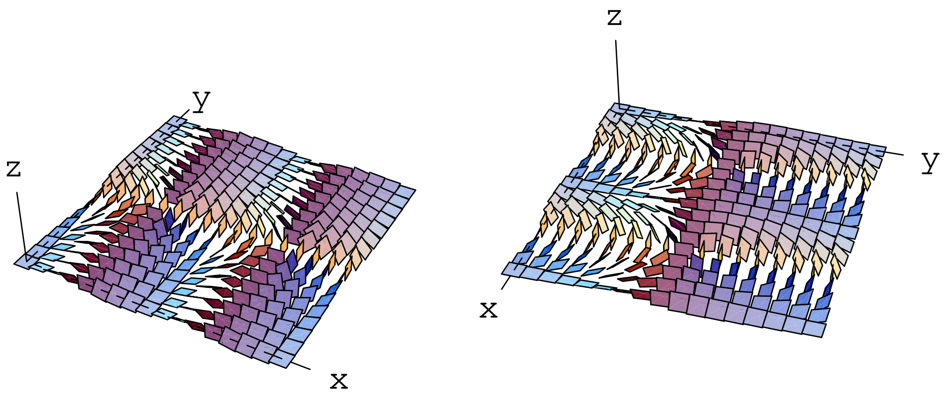


Figure 24: TE_{21} +-field at $z = \frac{2}{3}\frac{\pi}{2}$.

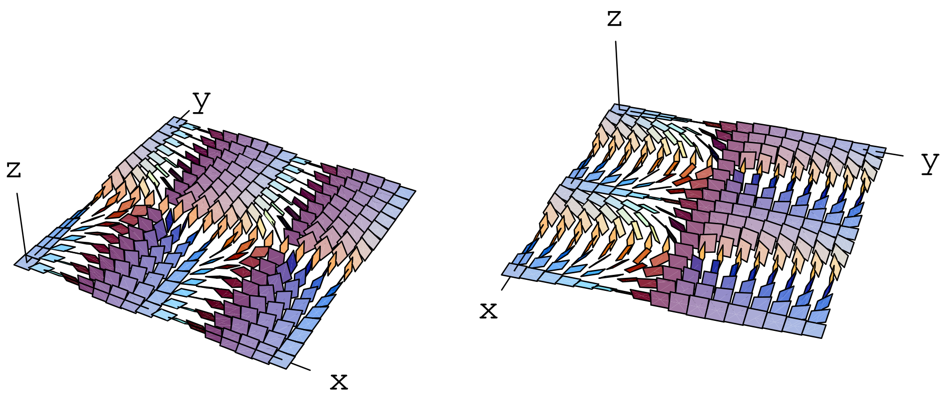


Figure 25: TE_{21} +-field at $z = \frac{3}{3}\frac{\pi}{2}$.

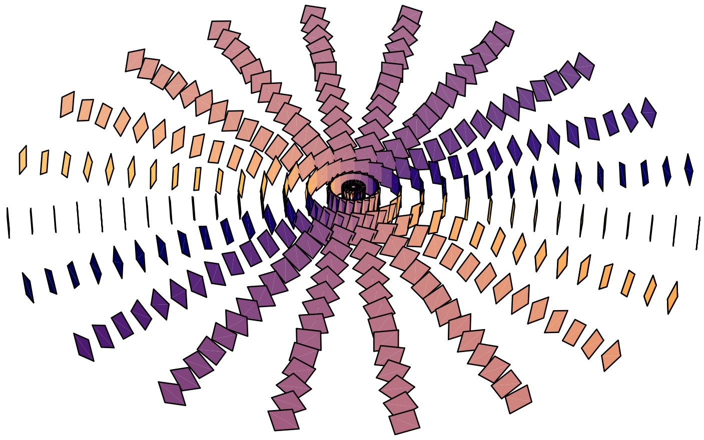


Figure 26: TE_{01} +-field at $z = \frac{0}{3}\frac{\pi}{2}$.

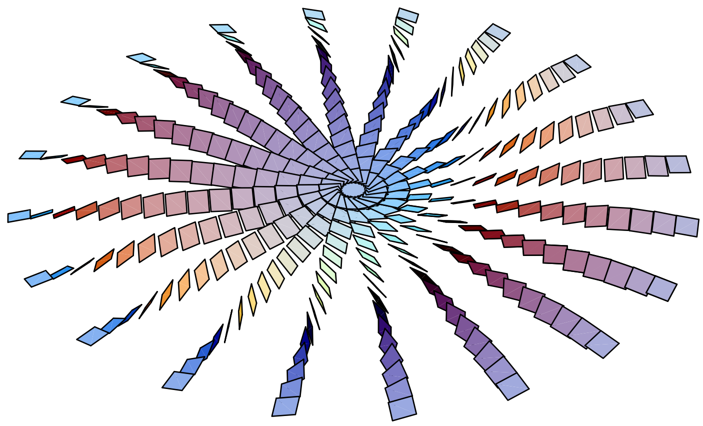


Figure 27: TE_{01} +-field at $z = \frac{1}{3}\frac{\pi}{2}$.

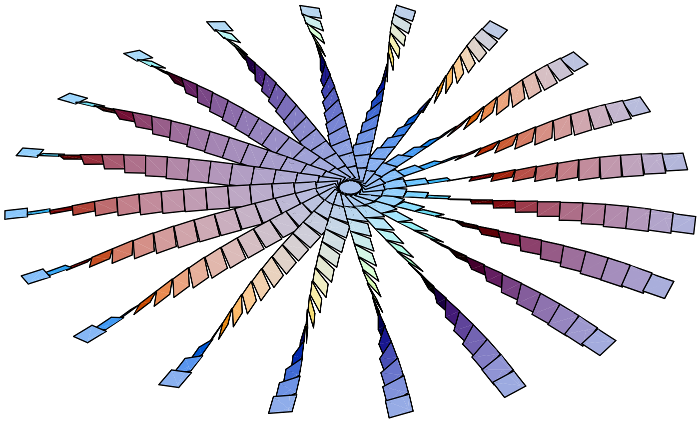


Figure 28: TE_{01} +-field at $z = \frac{2}{3}\frac{\pi}{2}$.

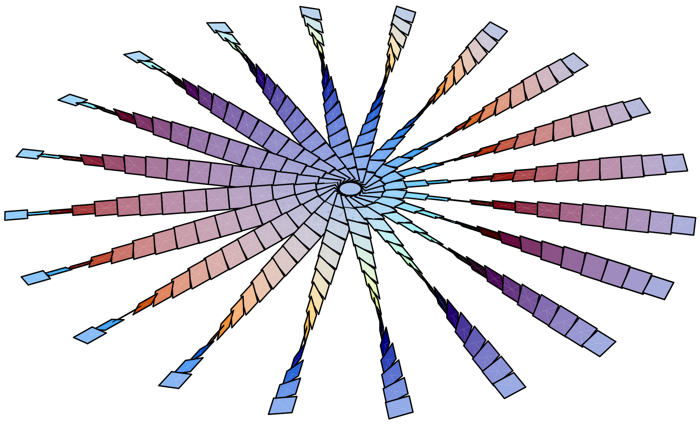


Figure 29: TE_{01} +-field at $z = \frac{3}{3}\frac{\pi}{2}$.

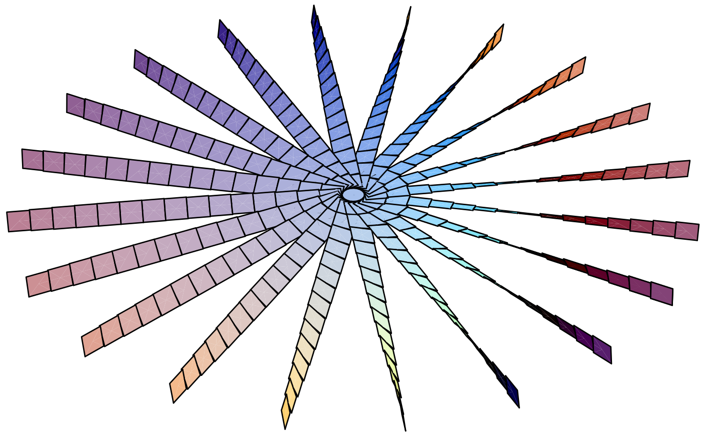


Figure 30: TM_{01} +-field at $z = \frac{0}{3}\frac{\pi}{2}$.

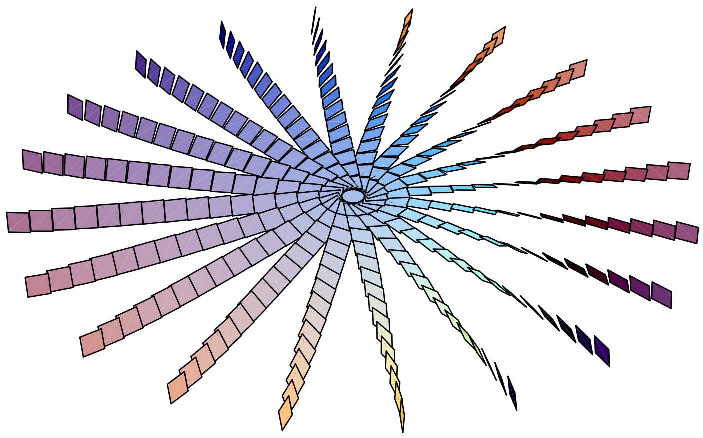


Figure 31: TM_{01} +-field at $z = \frac{1}{3}\frac{\pi}{2}$.

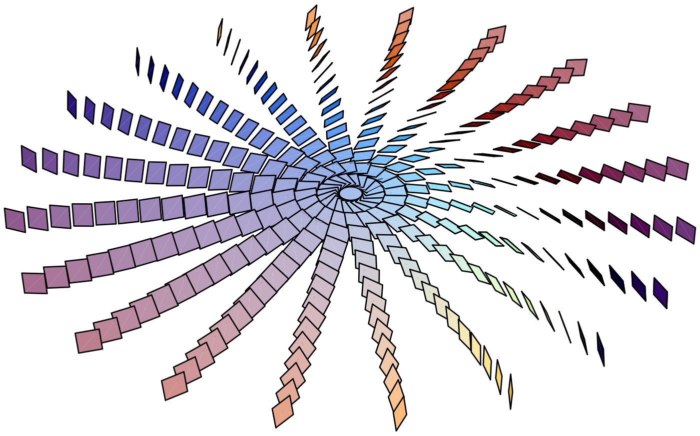


Figure 32: TM_{01} +-field at $z = \frac{2}{3}\frac{\pi}{2}$.

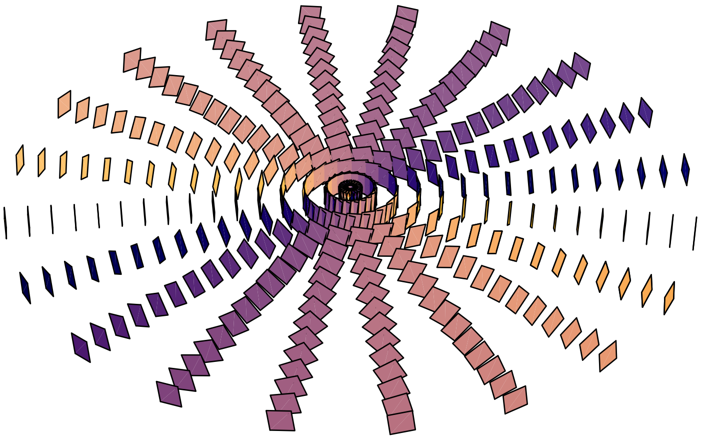


Figure 33: TM_{01} +-field at $z = \frac{3}{3}\frac{\pi}{2}$.

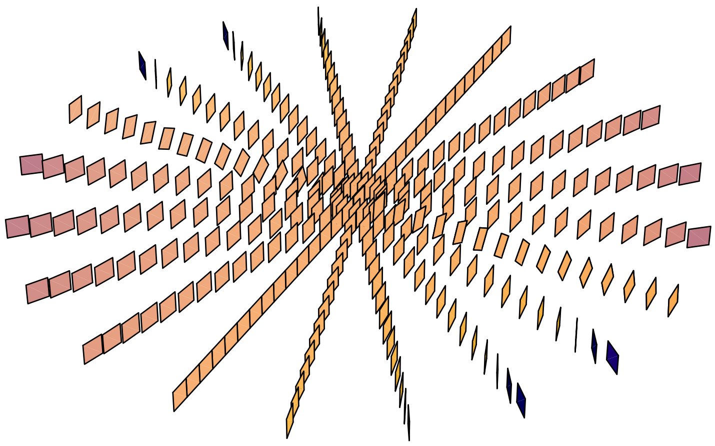


Figure 34: TE_{11} +-field at $z = \frac{0}{3}\frac{\pi}{2}$.

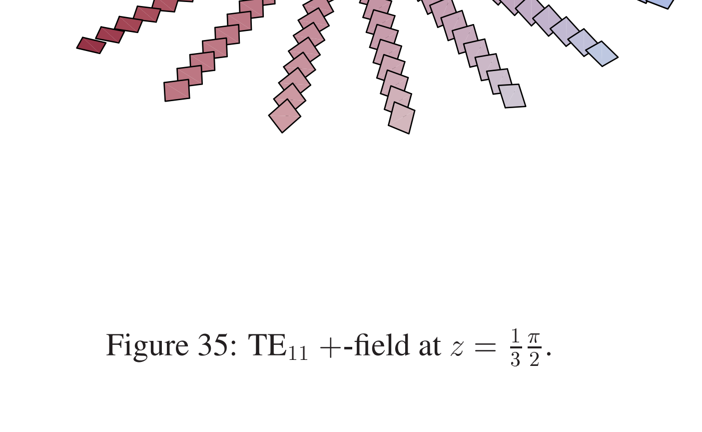


Figure 35: TE_{11} +-field at $z = \frac{1}{3}\frac{\pi}{2}$.

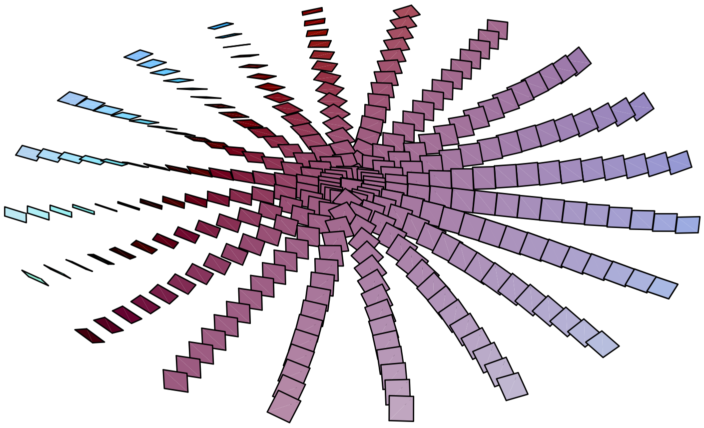


Figure 36: TE_{11} +-field at $z = \frac{2}{3}\frac{\pi}{2}$.

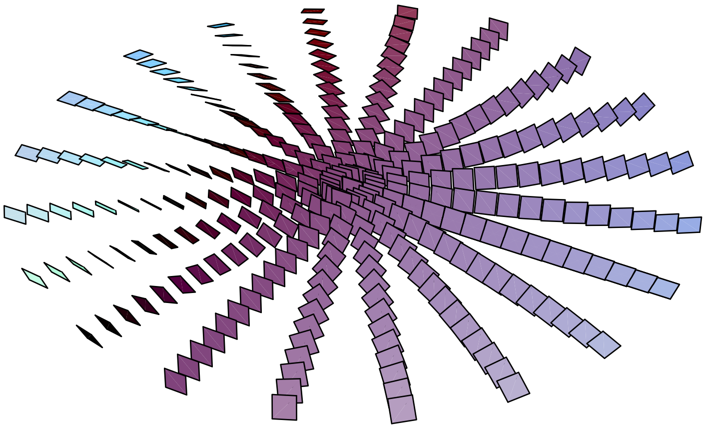


Figure 37: TE_{11} +-field at $z = \frac{3}{3}\frac{\pi}{2}$.

7 Conclusions

In this work we have applied a known mathematical decomposition from fluid mechanics to electromagnetics. We have also shown that this decomposition can be derived as a generalization of the Bohren decomposition and a refinement of the Helmholtz's decomposition. The key property of this decomposition is that it commutes with the curl operator. This property enables us to decompose Maxwell's equations in arbitrary media. The main advantage of these decomposed Maxwell's equations is that they give a much more detailed view of electromagnetics. In particular, they show that the decomposed components of the $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{H}$ fields are governed by three *uncoupled* sets of equations. We have also seen that by formulating the constitutive equations for these decomposed fields, it is possible to directly describe the handed behavior of the media. Depending on the complexity of the media, these constitutive equations might or might not couple the components of the decomposed fields.

In this work we have also studied contact and symplectic geometry and their relation to electromagnetics. Since contact and symplectic geometry has been found in numerous other areas of physics, this study is highly motivated. In this work, we have shown that known solutions to Helmholtz's equation induce contact structures. However, from the present work, we can not say whether an arbitrary solution to Maxwell's equations induces contact structures. The problem is that the helicity decomposition only assures that, say, for the electric field $\mathbf{E}$, we have $\int_{\mathbb{R}^3} \mathbf{E}_+ \cdot \nabla \times \mathbf{E}_+ \geq 0$. To prove that $\ker \mathbf{E}_+^\flat$ is a contact structure, one should be able to conclude that $\mathbf{E}_+ \cdot \nabla \times \mathbf{E}_+ > 0$. Since the helicity decomposition is based on the Fourier transform, it can be very difficult to prove such local properties for the decomposed fields. Probably the most simple way to gain further insight into this problem, would be to perform numerical experiments.

However, if, indeed, the decomposed fields in Maxwell's equations would always induce contact structures, it would be a very attractive result since it would give more "structure" to electromagnetism. If one can always assume that a solution splits into three components, and two of these would be contact structures, one can make much more assumptions, and possibly derive quite general results for solutions to Maxwell's equations. For instance, since contact geometry has been studied as a mathematical branch, there are many results, which could be applied directly to electromagnetism. As an example, we used Darboux's theorem in Section 6.1 to derive a local invariance result for solutions to Helmholtz's equations. By similar argumentation, we could use Darboux's theorem to show that all solutions to Maxwell's equations locally look like the standard contact structure. This could possibly be used to design a FEM solver for Maxwell's equations, which would take into account the internal structure of the solution. For instance, if we compare the decomposed solutions in Figures 10-25 to the the figures of the standard structure in Figure 3, we see that they are very similar. That would suggest that in such a solver, one would not need too many elements to model the solution. However, how the 0-field should be modeled in such a solver is not quite clear. One approach would be to disregard it. Since its curl vanishes, it does not radiate. Therefore, one could expect that it does not modify the qualitative properties of the solution such as, for example, the radiation pattern of an antenna, which usually is what is sought.

In the present work we have only studied non-relativistic electromagnetism. Our approach has, in fact, been somewhat unnatural since we have studied time dependent solutions only when time is constant. The next step would be to study the geometry of electromagnetism in four dimensions, i.e., in spacetime. That would probably involve symplectic geometry. One approach would be to incorporate time into the contact structures studied in this work using some connection between contact and symplectic geometry. Another approach would be to derive a similar decomposition as given in Proposition 3.2, using the Hodge operator in spacetime, which is an anti-involution, i.e., it satisfies $** = -id$.

The main motivations for studying contact and symplectic geometry in electromagnetism is that these structures are purely topological. In other words, they do not require an external structure such as a Riemannian metric. Another feature, which is very interesting is that both contact and symplectic structures induce their own internal "Hodge operators", i.e., mappings $\Omega^p(M^n) \to \Omega^{n-p}(M^n)$ [64]. In this work we have not studied these mappings. However, it is quite possible that using these mappings, one could formulate the constitutive equations. If that would be possible, it would yield an almost topological formulation for electromagnetism. In such a formulation, the only metrical dependence would be due to the helicity decomposition. An alternative approach would be to treat the decomposed fields as fundamental quantities of electromagnetism. If one further assumes that these are contact structures, and the the constitutive equations could be written using the induced Hodge operators, that would yield a completely topological formulation for electromagnetism.

A Dyadic algebra

A *dyad* or a *dyadic* is a linear mapping that maps a vector in $\mathbb{R}^3$ to another vector in $\mathbb{R}^3$. The difference, however, between dyadic algebra and standard matrix algebra is that dyadic algebra provides a powerful algebra for manipulating these mappings, or dyadics, without choosing a basis in $\mathbb{R}^3$. For instance, using certain dyad products, it is possible to give coordinate free expressions for the inverse and determinant of a dyad [5]. This usually makes calculations with dyadic algebra much more simpler than writing down the mappings in Cartesian coordinates.

Historically, dyads were introduced in 1884 by J.W. Gibbs in the same pamphlet as vector algebra [5]. A modern introduction to dyadic algebra and its application to electromagnetics can be found in [5]. In this Appendix we give a short introduction to the parts of dyad algebra, which we use in Section 3 to derive the helicity decomposition.

A dyadic is written as $\overline{\overline{\mathbf{A}}}$, $\overline{\overline{\mathbf{B}}}$ using two over-lines. The most general dyadic is a finite sum of the form

$$\overline{\overline{\mathbf{A}}} = \sum_{i=1}^{N} \mathbf{a}_i \mathbf{b}_i,$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are possibly complex vectors in $\mathbb{R}^3$. For this dyadic, we define the following operations.

a) If γ is a complex number, then $\gamma \mathbf{A} = \sum_{i=1}^{N} (\gamma \mathbf{a}_i) \mathbf{b}_i$. The *complex conjugate* of $\overline{\overline{\mathbf{A}}}$ is defined as $\overline{\overline{\mathbf{A}}}^* = \sum_{i=1}^{N} \mathbf{a}_i^* \mathbf{b}_i^*$.

b) If $\overline{\overline{\mathbf{B}}} = \sum_{i=N+1}^{M} \mathbf{a}_i \mathbf{b}_i$ is another dyadic, then $\overline{\overline{\mathbf{A}}} + \overline{\overline{\mathbf{B}}} = \sum_{i=1}^{M} \mathbf{a}_i \mathbf{b}_i$.

c) The *dot product* between the dyadic $\overline{\overline{\mathbf{A}}}$ and a possibly complex vector $\mathbf{x}$ yields a vector defined as

$$\begin{aligned} \mathbf{x} \cdot \overline{\overline{\mathbf{A}}} &= \sum_{i=1}^{N} (\mathbf{x} \cdot \mathbf{a}_i) \mathbf{b}_i, \\ \overline{\overline{\mathbf{A}}} \cdot \mathbf{x} &= \sum_{i=1}^{N} \mathbf{a}_i (\mathbf{b}_i \cdot \mathbf{x}). \end{aligned}$$

From these expressions it follows that $\overline{\overline{\mathbf{A}}}$ represents the same mapping in $\mathbb{R}^3$ as the 3×3 matrix $\sum_{i=1}^{N} \mathbf{a}_i \mathbf{b}_i^T$, where $\mathbf{a}_i$ and $\mathbf{b}_i$ are interpreted as column vectors in some orthonormal basis. (In dyadic algebra row and column vectors are not distinguished.)

d) The *cross product* between the dyadic $\overline{\overline{\mathbf{A}}}$ and a possibly complex vector $\mathbf{x}$ yields another dyadic defined as

$$\begin{aligned}\mathbf{x}\times\overline{\overline{\mathbf{A}}} &= \sum_{i=1}^{N}(\mathbf{x}\times\mathbf{a}_i)\mathbf{b}_i,\\ \overline{\overline{\mathbf{A}}}\times\mathbf{x} &= \sum_{i=1}^{N}\mathbf{a}_i(\mathbf{b}_i\times\mathbf{x}).\end{aligned}$$

e) The *dot product* between the dyadic $\overline{\overline{\mathbf{A}}}$ and $\overline{\overline{\mathbf{B}}} = \sum_{i=1}^{N'}\mathbf{a}'_i\mathbf{b}'_i$ yields the dyadic

$$\begin{aligned}\overline{\overline{\mathbf{A}}}\cdot\overline{\overline{\mathbf{B}}} &= \Big(\sum_{i=1}^{N}\mathbf{a}_i\mathbf{b}_i\Big)\cdot\Big(\sum_{j=1}^{N'}\mathbf{a}'_j\mathbf{b}'_j\Big)\\ &= \sum_{j=1}^{N'}\sum_{i=1}^{N}(\mathbf{b}_i\cdot\mathbf{a}'_j)\mathbf{a}_i\mathbf{b}'_j.\end{aligned}$$

Since matrix algebra is associative, it follows from *c)*, that the dot product between dyadics is associative.

f) The *transpose* of $\overline{\overline{\mathbf{A}}}$ is the dyadic $\overline{\overline{\mathbf{A}}}^{\mathrm{T}} = \sum_{i=1}^{N}\mathbf{b}_i\mathbf{a}_i$. Then $\overline{\overline{\mathbf{A}}}^{\mathrm{T}}\cdot\mathbf{x} = \mathbf{x}\cdot\overline{\overline{\mathbf{A}}}$ and $\mathbf{x}\cdot\overline{\overline{\mathbf{A}}}^{\mathrm{T}} = \overline{\overline{\mathbf{A}}}\cdot\mathbf{x}$. Also, $(\overline{\overline{\mathbf{A}}}\cdot\overline{\overline{\mathbf{B}}})^{\mathrm{T}} = \overline{\overline{\mathbf{B}}}^{\mathrm{T}}\cdot\overline{\overline{\mathbf{A}}}^{\mathrm{T}}$.

In $\mathbb{R}^3$ one can prove that the most general dyadic can always be written as $\overline{\overline{\mathbf{A}}} = \sum_{i=1}^{3}\mathbf{a}_i\mathbf{b}_i$, i.e., $N \le 3$ [5]. It should be clear, that for an arbitrary dyadic $\overline{\overline{\mathbf{A}}}$, the $\mathbf{a}_i$:s and $\mathbf{b}_i$:s are not uniquely determined. This is similar to vector algebra: there is no unique way of writing a vector as the sum of other vectors [5]. We say that two dyadics $\overline{\overline{\mathbf{A}}}$ and $\overline{\overline{\mathbf{B}}}$ are equal ($\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{B}}}$), if they represent the same mapping in $\mathbb{R}^3$, i.e., if $\overline{\overline{\mathbf{A}}}\cdot\mathbf{x} = \overline{\overline{\mathbf{B}}}\cdot\mathbf{x}$ for all vectors $\mathbf{x}$.

The next example shows that, in general, the dot product between a vector and a dyad is not associative.

Example A.1 We have

$$\begin{aligned}(\overline{\overline{\mathbf{A}}}\cdot\mathbf{x})\cdot(\overline{\overline{\mathbf{B}}}\cdot\mathbf{y}) &= (\mathbf{x}\cdot\overline{\overline{\mathbf{A}}}^{\mathrm{T}})\cdot(\overline{\overline{\mathbf{B}}}\cdot\mathbf{y})\\ &= \mathbf{x}\cdot(\overline{\overline{\mathbf{A}}}^{\mathrm{T}}\cdot\overline{\overline{\mathbf{B}}})\cdot\mathbf{y},\end{aligned}$$

and

$$(\overline{\overline{\mathbf{A}}}\cdot\overline{\overline{\mathbf{B}}})\cdot\mathbf{x} = \overline{\overline{\mathbf{A}}}\cdot(\overline{\overline{\mathbf{B}}}\cdot\mathbf{x}),$$

but

$$(\overline{\overline{\mathbf{A}}} \cdot \mathbf{x}) \cdot \overline{\overline{\mathbf{B}}} \neq \overline{\overline{\mathbf{A}}} \cdot (\mathbf{x} \cdot \overline{\overline{\mathbf{B}}}).$$

These results follow by setting $\overline{\overline{\mathbf{A}}} = \mathbf{a}_i\mathbf{b}_i$ and $\overline{\overline{\mathbf{B}}} = \mathbf{a}_j\mathbf{b}_j$ and using linearity. □

The *identity dyadic* is denoted by $\overline{\overline{\mathbf{I}}}$ and satisfies $\overline{\overline{\mathbf{I}}} \cdot \overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{I}}}$. If $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ is an orthonormal real basis for $\mathbb{R}^3$, then $\overline{\overline{\mathbf{I}}} = \mathbf{u}_1\mathbf{u}_1 + \mathbf{u}_2\mathbf{u}_2 + \mathbf{u}_3\mathbf{u}_3$, and $\mathbf{a} \cdot \overline{\overline{\mathbf{I}}} = \overline{\overline{\mathbf{I}}} \cdot \mathbf{a} = \mathbf{a}$. Further, $(\mathbf{b} \times \overline{\overline{\mathbf{I}}}) \cdot \mathbf{a} = \sum_{i=1}^{3} \mathbf{b} \times \mathbf{u}_i(\mathbf{u}_i \cdot \mathbf{a}) = \mathbf{b} \times \mathbf{a}$. This shows that the dyad $\mathbf{b} \times \overline{\overline{\mathbf{I}}}$ represents the mapping $\mathbf{x} \mapsto \mathbf{b} \times \mathbf{x}$. If $\mathbf{b} = \sum_{i=1}^{3} b_i\mathbf{u}_i$, then the matrix corresponding to $\mathbf{b} \times \overline{\overline{\mathbf{I}}}$ in $\mathbb{R}^3$ is

$$\begin{pmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{pmatrix}. \tag{67}$$

This matrix is anti-symmetric. We also have $(\mathbf{b} \times \overline{\overline{\mathbf{I}}})^{\mathrm{T}} = -\mathbf{b} \times \overline{\overline{\mathbf{I}}}$. This is seen by writing out $\overline{\overline{\mathbf{I}}}$ as above and using the relation $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A})$ twice.

The *zero dyadic* maps all vectors to zero. We denote it by 0 and not by $\overline{\overline{\mathbf{0}}}$.

Example A.2 (Examples of dyadic operators) Suppose that $\mathbf{u}$ is a (real) unit vector, and $\mathbf{a}$ is an arbitrary (real) vector in $\mathbb{R}^3$.

a) The dyad $\overline{\overline{\mathbf{A}}} = \mathbf{uu}$ is the projection onto $\mathbf{u}$, i.e., the vector $(\mathbf{uu}) \cdot \mathbf{a}$ is the component of $\mathbf{a}$ parallel to $\mathbf{u}$.

b) The dyad $\overline{\overline{\mathbf{B}}} = \overline{\overline{\mathbf{I}}} - \mathbf{uu}$ is the projection of a vector onto the plane orthogonal to $\mathbf{u}$, i.e., the vector $(\overline{\overline{\mathbf{I}}} - \mathbf{uu}) \cdot \mathbf{a}$ has no component parallel to $\mathbf{u}$.

c) The dyad $\overline{\overline{\mathbf{C}}} = \overline{\overline{\mathbf{I}}} - 2\mathbf{uu}$ reflects a vector trough the plane orthogonal to $\mathbf{u}$.

With dyadic algebra it is easy to prove a number of properties for these operators. For instance, we have that $\overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{A}}}$, $\overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{B}}} = 0$, $\overline{\overline{\mathbf{A}}} + \overline{\overline{\mathbf{B}}} = \overline{\overline{\mathbf{I}}}$, $\overline{\overline{\mathbf{B}}} \cdot \overline{\overline{\mathbf{B}}} = \overline{\overline{\mathbf{B}}}$, $\overline{\overline{\mathbf{B}}} \cdot \overline{\overline{\mathbf{C}}} = \overline{\overline{\mathbf{B}}}$, and $\overline{\overline{\mathbf{C}}} \cdot \overline{\overline{\mathbf{C}}} = \overline{\overline{\mathbf{I}}}$. These are, of course, all geometrically clear. However, the main advantage with dyadic algebra is that these relations can be proved without fixing coordinates in $\mathbb{R}^3$. □

The *inverse* of a dyad $\overline{\overline{\mathbf{A}}}$ (when defined) is the dyadic $\overline{\overline{\mathbf{A}}}^{-1}$ for which $\overline{\overline{\mathbf{A}}} \cdot \overline{\overline{\mathbf{A}}}^{-1} = \overline{\overline{\mathbf{A}}}^{-1} \cdot \overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{I}}}$. The *determinant* of a dyad is also defined. Although it is possible to define it using certain dyad products [5], for our purpose, it is easiest to simply define it as the determinant of the corresponding matrix. Suppose $\overline{\overline{\mathbf{A}}}$ is a dyad, and $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ is an orthonormal basis. Then the matrix $A_{ij} = (\overline{\overline{\mathbf{A}}} \cdot \mathbf{u}_j) \cdot \mathbf{u}_i$ (i row, j column) represent the same mapping as $\overline{\overline{\mathbf{A}}}$; for a column vector $\mathbf{x}$, $\overline{\overline{\mathbf{A}}} \cdot \mathbf{x} = A \cdot \mathbf{x}$. We can then define the *determinant* of the dyad $\overline{\overline{\mathbf{A}}}$ as $\det \overline{\overline{\mathbf{A}}} = \det A$. Since the determinant is invariant of the choice of basis, $\det \overline{\overline{\mathbf{A}}}$ does not depend on $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$.

We say that a dyad is a *rotational dyad*, if $\overline{\overline{\mathbf{R}}}^{\mathrm{T}} = \overline{\overline{\mathbf{R}}}^{-1}$, $\det \overline{\overline{\mathbf{R}}} = 1$, and $\overline{\overline{\mathbf{R}}}$ is real ($\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{R}}}^*$). For such dyads, we have the following result.

Lemma A.3 *[5, 65] Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two possibly complex vectors in $\mathbb{R}^3$, and let $\overline{\overline{\boldsymbol{R}}}$ be a rotational dyad. Then*

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= (\overline{\overline{\boldsymbol{R}}} \cdot \boldsymbol{u}) \cdot (\overline{\overline{\boldsymbol{R}}} \cdot \boldsymbol{v}), \\ \overline{\overline{\boldsymbol{R}}} \cdot (\boldsymbol{u} \times \boldsymbol{v}) &= (\overline{\overline{\boldsymbol{R}}} \cdot \boldsymbol{u}) \times (\overline{\overline{\boldsymbol{R}}} \cdot \boldsymbol{v}). \end{aligned}$$

Proof. The first claim follows directly from $\overline{\overline{\mathbf{R}}} \cdot \mathbf{u} = \mathbf{u} \cdot \overline{\overline{\mathbf{R}}}^{\mathrm{T}}$ and Example A.1. To prove the second claim, we first fix some orthonormal coordinates $\mathbf{u}_1, \mathbf{u}_1, \mathbf{u}_3$ for $\mathbb{R}^3$. (An alternative proof using dyadic algebra can be found in [5].)

We denote the i:th coordinate of a vector $\mathbf{a}$ by either a^i or a_i. Then $(\mathbf{u} \times \mathbf{v})_i = \frac{1}{2}\varepsilon_{ikl} u^k v^l$ with Einstein's summing convention. However, since we do not distinguish between row and column vectors, the indices will not always be upper and lower. The components in the $\overline{\overline{\mathbf{R}}}$ dyad are indexed as above, $R_{ij} = (\overline{\overline{\mathbf{R}}} \cdot \mathbf{u}_j) \cdot \mathbf{u}_i$. Then

$$\begin{aligned} ((\overline{\overline{\mathbf{R}}} \cdot \mathbf{u}) \times (\overline{\overline{\mathbf{R}}} \cdot \mathbf{v}))_m &= \frac{1}{2}\varepsilon_{ikm} R_{ij} u^j R_{kl} v^l \\ &= \frac{1}{2}\varepsilon_{ikn} R_{ij} R_{kl} R_{np} R^{\mathrm{T}}_{pm} u^j v^l \\ &= \frac{1}{2}\varepsilon_{jlp} R_{mp} u^j v^l \\ &= (\overline{\overline{\mathbf{R}}} \cdot (\mathbf{u} \times \mathbf{v}))_m, \end{aligned}$$

where we have used the relations $R_{np} R^{\mathrm{T}}_{pm} = \delta_{nm}$, $R^{\mathrm{T}}_{pm} = R_{mp}$, $\varepsilon_{ikn} R_{ij} R_{kl} R_{np} = \det \overline{\overline{\mathbf{R}}} \varepsilon_{jlp}$, and $\det R = 1$. □

References

[1] R.A. Hegstrom, D.K. Kondepudi, *The Handedness of the Universe*, Scientific American, January 1990, 98-105.

[2] M. Gardner, *The ambidextrous universe*, Allen Lane, The Penguin Press, 1967.

[3] I. Hargittai, C.A. Pickover, (ed.), *Spiral symmetry*, World Scientific, 1992.

[4] R. Osserman, *A survey of minimal surfaces*, Van Nostrand Reinhold mathematical studies 25, 1969.

[5] I.V. Lindell, *Methods for Electromagnetic Field Analysis*, Clarendon Press, 1992.

[6] I.V. Lindell, A.H. Sihvola, S.A. Tretyakov, A.J. Viitanen, *Electromagnetic Waves in Chiral and Bi-Isotropic Media,* Artech House, 1994.

[7] K.F. Warnick, R.H. Selfridge, D.V. Arnold, *Teaching electromagnetic field theory using differential forms*, IEEE Transactions on education, Vol. 40, No. 1, Feb. 1997, 53-68.

[8] A. Bossavit, *On the notion of anisotropy of constitutive laws. Some implications of the "Hodge implies metric" result*, COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, Vol. 20, No. 1, 2001, 233-239.

[9] A. Gross, G.F. Rubilar, *On the derivation of the spacetime metric from linear electrodynamics*, arXiv:gr-qc/010301v2, 13 Jul. 2001.

[10] V. Guillemin, S. Sternberg, *Symplectic techniques in physics*, Cambridge University Press, 1984.

[11] J. Etnyre, R. Ghrist, *Contact topology and hydrodynamics. I. Beltrami fields and the Seifert conjecture*, Nonlinearity 13 (2000), no. 2, 441–458, preprint dg-ga/9708011 accessed.

[12] J. B. Etnyre, *Introductory Lectures on Contact Geometry*, preprint math.SG/0111118, November 2001.

[13] H. Geiges, *A Brief History of Contact Geometry and Topology*, Expositiones Mathematicae, Vol. 19, 2001, 25-53.

[14] M. Chen, *On the geometric structure of thermodynamics*, Journal of Mathematical Physics, Vol. 40, Nr. 2, Feb. 1999.

[15] R. Gulliver, *The heat-flow method in contact geometry*, Proceedings of Symposium on Geometric Analysis and Applications, Centre for Mathematics and its Applications, Canberra, Australia 2001, (http://www.ima.umn.edu/˜gulliver/papers/gaa.ps accessed August 2002)

[16] J. Cantarella, D. DeTurck, H. Gluck, M. Teytel, *Influence of Geometry and Topology on Helicity*, (www.math.uga.edu/˜cantarel/research/papers/influence/deturck.pdf accessed August 2002).

[17] M.J. Gotay, J.A. Isenberg, *The symplectization of science*, Gazette des Mathématiciens, Vol. 54, 59-79, 1992, (http://www.math.hawaii.edu/˜gotay/Symplectization.pdf accessed August 2002)

[18] R. McLachlan, R. Quispel, *Six Lectures on the Geometric Integration of ODEs*, July 1998, (http://www.massey.ac.nz/˜rmclachl/ accessed August 2002).

[19] I. Saitoh, Y. Suzuki, N. Takahashi, *The Symplectic Finite Difference Time Domain Method*, IEEE Transactions on magnetics, Vol. 37, No. 5, Sept. 2001.

[20] J.J. Duistermaat, *Fourier integral operators*, Courant Institute of Mathematical Sciences, New York University, 1973.

[21] A. Grigis, J. Stöstrand, *Microlocal Analysis for Differential Operators*, London Mathematical Society, Lecture Notes Series 106, Cambridge University Press, 1994.

[22] I.V. Lindell, F. Olyslager, *Duality in electromagnetics*, Helsinki University of Technology, Electromagnetics Laboratory Report Series, Report 331, Espoo, April 2000.

[23] V.I. Arnold, B.A. Keshin, *Topological Methods in Hydrodynamics*, Applied Mathematical Sciences 125, Springer, 1998.

[24] P.M. Bellan, *Spheromaks, A practical application of Magnetohydrodynamic Dynamos and Plasma Self-Organization*, Imperial College Press, 2000.

[25] P.R. Baldwin, G.M. Townsend, *Complex Trkalian fields and solutions to Euler's equations for the ideal fluid,* Physical Review E, Vol. 51, Nr. 3, 1995, 2059-2068.

[26] K. H. Tsui, *Force-free field model of ball lightning*, Physics of plasmas, Vol. 8, No. 3, March 2001, 687-689.

[27] A. Lakhtakia, *Beltrami fields in Chiral Media*, World Scientific, 1994.

[28] D. Reed, *Foundational Electrodynamics and Beltrami fields*, in Advanced Electromagnetism: Foundations, Theory and Applications, edited by T.W. Barret, D.M. Grimes, World Scientific, 1995.

[29] C.F. Bohren, D.R. Huffman, *Absorption and scattering of light by small particles*, John Wiley & Sons, 1983.

[30] A. Serdyukov, I. Semchenko, S. Trekyakov, A.Sihvola, *Electromagnetics of bi-anisotropic materials. Theory and applications*, Gordon and Breach science publishers, 2001.

[31] H.E. Moses, *Eigenfunctions of the curl operator, rotationally invariant Helmholtz theorem, and applications to electromagnetic theory and fluid mechanics*, SIAM Journal of Applied Mathematics, Vol. 21, No. 1, July 1971.

[32] F. Waleffe, *The nature of triad interactions in homogeneous turbulence*, Physics of Fluids A, Vol. 4, Nr. 2, February 1992.

[33] M.A. MacLeod, *The spherical curl transform of a linear force-free magnetic field*, Journal of Mathematical Physics, Vol. 39, No. 3, 1998.

[34] M.A. MacLeod, *A new description of force-free magnetic fields*, Journal of Mathematical Physics, Vol. 36, No. 6, 1995.

[35] P. Constantin, A. Majda, *The Beltrami spectrum for incompressible fluid flows*, Communications in mathematical physics, Vol 115, pp. 435-456, 1998.

[36] J. Bognár, *Indefinite inner product spaces*, Springer-Verlag, 1974.

[37] T. Frankel, *Geometry of physics*, Cambridge University press, 1997.

[38] K. Araki, K. Suzuki, K. Kishida, S. Kishiba, *Multiresolution approximation of the vector fields on T^3*, preprint math-ph/9904015, 16 April 1999.

[39] E. Kreyszig, *Introductory functional analysis with aplications*, John Wiley & Sons, 1978.

[40] F.G. Friedlander, *Introduction to the theory of distributions*, Cambridge University Press, 1982.

[41] G.B.Folland, *Real analysis*, John Wiley & Sons, 1999.

[42] E.M.Stein, G.Weiss, *Introduction to Fourier analysis on Euclidian spaces*, Princeton University Press, 1971.

[43] J. Van Bladel, *A discussion on Helmholtz' theorem*, Electromagnetics, vol. 13, pp. 95-110, 1993

[44] P. Hillion, *Electromagnetic waves in linear media*, Journal of Physics A: Mathematical and general, Vol. 28, 1995, 2647-2657.

[45] I. Madsen, J. Tornehave, *From Calculus to Cohomology*, Cambridge University press, 1997.

[46] B. Schutz, *Geometrical methods of mathematical physics*, Cambridge University press, 1980.

[47] W.M. Boothby, *An introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, 1986.

[48] P. Lounesto, *Clifford Algebras and Spinors*, London Mathematical Society Lecture Note Series 286, Cambridge University Press, 2001.

[49] W. Thirring, *A Course in Mathematical Physics 2: Classical Field Theory*, Springer-Verlag, 1978.

[50] G.E. March, *Helicity and Electromagnetic Field Topology*, in Advanced Electromagnetism: Foundations, Theory and Applications, edited by T.W. Barret, D.M. Grimes, World Scientific, 1995.

[51] D. McDuff, D. Salamon, *Introduction to Symplectic Topology*, Clarendon Press, 1997.

[52] Y. Eliashberg, *Classification of contact structures on* 3*-manifolds*, International Mathematics Research Notices, 1993, No. 3, 87-91.

[53] D. Bennequin, *Entrelacements et équation de Pfaff*, Astérisque, Vol. 107-108, 1983, 87-161.

[54] C. Meckert, *Forme de contact sur la somme connexe de deux variétés de contact de dimension impaire*, Université de Grenoble. Annales de l'Institut Fourier, Vol. 32, Nr. 3 (1982), 251-260.

[55] A. Bellaïche, J-J. Risler (ed.), *Sub-Riemannian Geometry*, Birkhäuser, 1996.

[56] J. Mitchell, *On Carnot-Carathéodory metrics*, Journal of Differential Geometry, Vol. 21, 1985, 35-45.

[57] W. Burke, *Applied differential geometry*, Cambridge University press, 1985.

[58] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, 1980.

[59] B. Jancewicz, *Answer to question 55: Are there pictorial examples that distinguish covariant and contravariant vectors?*, preprint gr-gc/9807044, July 1998.

[60] Y. Eliashberg, *Symplectic topology in the nineties*, Differential Geometry and its Applications, Vol. 9, 1998, 59-88.

[61] R.A. Silverman, *Introductory complex analysis*, Dover publications, 1972.

[62] D.K. Cheng, *Fundamentals of Engineering Electromagnetics*, Addison-Wesley Publishing Company, 1993.

[63] I. Lindell, *Aaltojohtoteoria*, Wave-guide theory (in Finnish), Otatieto 583, 1997.

[64] J.V. Beltran, *Star calculus on Jacobi manifolds*, Differential Geometry and its applications, 16 (2002), 181-198.

[65] R. Feynman, *Six not-so-easy pieces: Einstein's relativity, Symmetry, and Space-time*, Addison-Wesley, 1997.