TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Janne Keränen

Towards Computational Electromagnetics in Spacetime

Julkaisu 953 • Publication 953

Tampere 2011

Tampereen teknillinen yliopisto. Julkaisu 953
Tampere University of Technology. Publication 953

Janne Keränen

Towards Computational Electromagnetics in Spacetime

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Sähkötalo Building, Auditorium S1, at Tampere University of Technology, on the 11th of March 2011, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2011

ISBN 978-952-15-2528-5 (printed)
ISBN 978-952-15-2555-1 (PDF)
ISSN 1459-2045

Abstract

In this work, we developed a geometric computational method for electromagnetic wave problems. The method unifies the spatial and temporal discretizations and produces a four-dimensional spacetime computational scheme, which treats spatial and time dimensions equally.

This new method seeks primarily to develop the current geometric methods, i.e., the finite integration technique and the cell method. We introduce some improvements to these methods and, additionally, develop tools for further development. For these, we present a couple of challenge problems and largely solve them with our method in this thesis.

There are several reasons why we focus on the mathematical structures underlying the spacetime and electromagnetic models. For example, first, four-dimensional fields cannot be modeled with elementary vector algebra, and, second, a functional computational method is usually based on a solid mathematical foundation. From the background of the theory of relativity, we develop a model for spacetime. Because this model includes an indefinite metric, all metric-dependent concepts must agree with this non-Riemannian indefinite metric. Consequently, we utilize concepts new to electromagnetic computation. And because our target is electromagnetic computation, we choose a model that agrees well with computation: we model electromagnetic fields with cochains.

In the end, we introduce a computational method that makes use of a pair of four-dimensional Lorentzian orthogonal meshes. In developing it, we concentrate on the constitutive relations, boundary conditions, and gauging methods and implement the method and demonstrate it with several examples.

Preface

The research leading to this thesis was carried out between years 2004 and 2010 within the unit of Electromagnetics at the Department of Electronics, Tampere University of Technology. This thesis is a continuation of various electromagnetic waves-related research projects conducted at the unit of Electromagnetics over the last decade. The main aim of these projects has been the development of current computational methods.

This thesis project was quite challenging, because it combined computational electromagnetics with the theory of relativity—which I was mainly unfamiliar with before the thesis—and because there was relatively scarce prior research on this subject. Unsurprisingly, a few severe problems were confronted during this research; the two most severe obstructions halted the progress for nearly two years and were finally solved only just within the last year. For this reason, the thesis had to be finalized quite rapidly and many originally planned research subjects had to be omitted. Despite these limitations, I am quite happy about the final stage of this thesis.

This work was funded by the Graduate School of Applied Electromagnetism and the TUT Graduate School.

I want to give thanks to all the folks who have assisted or supported me during this work. My greatest gratitude goes to Professor Lauri Kettunen, my supervisor and the head of our unit. He has given me the opportunity to concentrate fully on my research and, further, he has created (among others) an encouraging atmosphere in our unit. Additionally, special thanks are due to Dr. Timo Tarhasaari, who has advised me during the whole thesis project and helped me to work out some of the hardest questions in this work. Thanks to Dr. Timo Lepistö for proofreading this thesis and Dr. Jukka Uusitalo for implementing the interpolators for several figures.

The unit of Electromagnetics has been a great place to work in; I want to thank every present and former workmate from there. Especially I want to thank Saku Suuriniemi, Aki Korpela, Jari Kangas and Pasi Raumonen, and all my roommates: Emmi Koljonen, Tero Kortelainen, and Olli Särkkä.

Last but not least, I wish to thank my fiancée Katri for all her support during the last six years.

Contents

List of symbols

Chapter 2

t_i	time instant of primal mesh variables
$t_{i+\frac{1}{2}}$	time instant of dual mesh variables
Δt	time-step
∂_t	time derivative

Chapter 3

$[\mathbf{A}]_{ij}$	ij^{th} element of matrix $\mathbf{A}$
c	speed of light
o, p	events
t	time coordinate
x, y, z	spatial coordinates
τ	proper time

Chapter 4

A	atlas
$\mathfrak{b}$	pointwise magnetic flux, 2-form
$\mathcal{B}_p(M)$	space of p-boundaries
$\mathcal{B}_p(K)$	space of finite p-boundaries
$\mathcal{B}^p(M)$	space of p-coboundaries
$\mathcal{C}$	mapping from p-forms to p-cochains
$C_p(M)$	space of polyhedral p-chains
$C_p^\flat(M)$	space of flat $p-$chains
$C_\times(M)$	chain complex
$C_p(K)$	space of finite p-chains
$C_\times(K)$	finite chain complex

$C^p(M)$	space of polyhedral p-cochains
$C^p_\flat(M)$	space of flat p-cochains
$C^\times(M)$	cochain complex
$C^p(K)$	space of finite p-cochains
$C^\times(K)$	finite cochain complex
$\tilde{C}_p(M)$	space of outer oriented polyhedral p-chains
$\tilde{C}^\flat_p(M)$	space of outer oriented flat p-chains
$\tilde{C}_\times(\tilde{K})$	outer oriented finite chain complex
$\tilde{C}^\times(M)$	twisted cochain complex
$\tilde{C}^p(\tilde{K})$	space of finite twisted p-cochains
$\tilde{C}^\times(\tilde{K})$	finite twisted cochain complex
d	exterior operator for cochains
d	exterior derivative for differential forms
$\mathbf{d}_p$	real array representative of d for p-cochains
$\mathscr{d}$	pointwise electric flux, twisted 2-form
det	determinant
dt	1-form associated with the time direction, i.e., $(dt\|e_0) = 1$
$\mathscr{e}$	pointwise electric field, 1-form
$\mathcal{E}$	electric field 1-cochain
e_0	basis unit vector of a Lorentzian manifold in the time direction
$e_1, \ldots, e_n$	a standard basis for T^xM (see section 4.1.3)
$[\boldsymbol{e}_i]_j$	j^{th} element on real array $\boldsymbol{e}_i$ (i.e., $\boldsymbol{e}_i \in \mathbb{R}^n$)
$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$	a standard basis for $\mathbb{R}^n$ (see section 4.1.3)
$\mathrm{extr}(\sigma, v, t)$	extrusion of a flat chain Σ along vector field v according to t
$\mathcal{F}$	mapping from p-cochains to p-forms
$F^p(M)$	space of p-dimensional differential forms
$\mathscr{h}$	pointwise magnetic field, twisted 1-form
i_v	contraction by vector field v
$I^+(x), I^-(x)$	chronological future and past (see section 4.1.7)
$J^+(x), J^-(x)$	causal future and past (see section 4.1.7)
K	cellular mesh complex
$\tilde{K}$	outer oriented cellular mesh complex
M	manifold
M_τ	purely spatial submanifold of M
n	dimension of the manifold
ori	orientation comparison

O_r	an orientation of a manifold
P	submanifold
$\mathbb{R}^n$	Euclidian n-space
$\mathbb{R}^n_+$	half space of $\mathbb{R}^n$ (see definition 4.1)
$\mathbb{R}^n_1$	Minkowski n-space
$\mathbb{R}^n_\nu$	semi-Euclidian space
$\{s\}$	p-vector corresponding to simplex s
$\mathfrak{S}_k s$	a sequence of simplex subdivisions of s
t_P	trace operator for cochains for a submanifold P
TM	tangent bundle
T_pM	p-vector bundle
T^xM	tangent space
T^x_pM	space of p-dimensional multivectors
T_pM	p-covector bundle
T^1_xM	space of covectors
T^p_xM	space of p-dimensional multicovectors
$\tilde{T}^x_pM$	space of p-dimensional outer oriented multivectors
$\tilde{T}^p_xM$	space of p-dimensional twisted multicovectors
u	velocity field of an observer
U	an open subset of $\mathbb{R}^n$ or $\mathbb{R}^n_+$
v_x	vector field v at x, i.e., $v_x \in T^xM$
V	vector space
$\mathcal{Z}_p(M)$	space of p-cycles
$\mathcal{Z}_p(K)$	space of finite p-cycles
$\mathcal{Z}^p(M)$	space of p-cocycles
α	a curve or a real coefficient in a formal sum
γ	a (finite) curve on manifold
γ'	velocity vector of curve
ϵ_i	signature of scalar product
ν	index of a scalar product
μ_M	canonical measure on manifold (see section 4.3.4)
$\mu_{\mathbb{R}^n}$	Lebesque measure on $\mathbb{R}^n$
σ	a polyhedral cell, an unit n-vector, a global orientation, or electric conductivity
$\tilde{\Sigma}(M)$	an outer oriented chain relative to the global orientation of M
$\tilde{\Sigma}(M_\tau)$	an outer oriented chain relative to the global orientation of M_τ

$\mathbf{\Sigma}$	real array representative of a chain Σ
ϕ	a map between two manifolds
ϕ_*	pushforward by ϕ
ϕ^*	pullback by ϕ
χ, η	coordinate charts
$\Psi_t(\Sigma)$	transportation of a chain Σ along vector field v according to t
ω_x	a differential form ω at x, i.e., $\omega_x \in T_x^1 M$
$\{\Omega_i^p\}$	basis of $C_p(M)$ corresponding to a p-cell basis of $C_p(M)$ (see equation (4.31))
$\coprod$	disjoint union
$\star$	Hodge operator
$\langle \cdot, \cdot \rangle$	a metric (or scalar product) on a smooth manifold
$\| \cdot \|$	norm of a vector or a p-vector
$\| \cdot \|_m$	mass norm for chains and cochains
$\| \cdot \|_\flat$	flat norm for chains and cochains
∂, ∂_p	boundary operator (see section 4.3.3)
$\boldsymbol{\partial}_p$	real array representative of ∂_p
∂M	boundary of a manifold
$\exists_1$	there exists a unique
$\wedge$	exterior product of multivectors
$\ll$	chronological precedence (see section 4.1.7)
$<$	strict causal precedence (see section 4.1.7)
$\leq$	causal precedence (see section 4.1.7)

Chapter 5

a	pointwise electromagnetic potential, 1-form
$\boldsymbol{a}$	magnetic 1-form potential
A	electromagnetic potential cochain, four-potential
$\mathcal{A}$	three-dimensional magnetic (vector) potential cochain
$[A]$	coset of an ungauged potential in $C_\flat^1(\Omega, \Gamma)/i(\mathcal{Z}_\flat^1(\Omega, \Gamma))$
$\mathcal{B}$	magnetic flux cochain
$\mathcal{B}_p^\flat(\Omega, \Gamma)$	space of relative boundaries modulo Γ
$\mathcal{B}_\flat^p(\Omega, \Gamma)$	space of relative coboundaries modulo Γ
$C_p^\flat(\Gamma)$	space of flat p-chains on Γ
$C_p^\flat(\Omega, \Gamma)$	relative flat p-chains modulo boundary Γ
$C_\flat^p(\Gamma)$	space of flat p-cochains on Γ
$C_p^\flat(\Omega, \Gamma)$	relative flat p-cochains modulo Γ
$\mathcal{D}$	electric flux cochain

$D^+(S_0)$	future domain of dependence of S_0
$D^-(S_0)$	past domain of dependence of S_0
f	pointwise electromagnetic field, 2-form
F	electromagnetic field cochain, Faraday cochain
g	pointwise electromagnetic excitation, electromagnetic displacement, twisted 2-form
G	Maxwell cochain, electromagnetic excitation
$\mathcal{H}$	magnetic field cochain
j	pointwise electric charge-current, twisted 3-form
j	electric current density, twisted 2-form
J	electric charge-current cochain
$\mathcal{J}$	electric current cochain
J_S	hypersurface charge-current cochain
$\mathcal{J}_S$	surface current cochain
m	magnetization (twisted) 1-form
m	magnetization-polarization density, twisted 2-form
M	magnetization-polarization cochain
p	momentum, electric polarization (twisted 2-form)
p_0	energy
p	momentum-energy
$\frac{\partial p}{\partial \tau}$	four-force
$\mathcal{P}$	electric charge cochain
$\mathcal{P}_S$	surface charge cochain
v_1	the purely spatial part of v
v_2	the part of v in the time direction
$\mathcal{Z}_p^\flat(\Omega, \Gamma)$	space of relative cycles modulo Γ
$\mathcal{Z}_\flat^p(\Omega, \Gamma)$	space of relative cocycles modulo Γ
Γ	a part of boundary $\partial\Omega$
Γ_{boundary}	timelike (i.e., spatial) part of boundary $\partial\Omega$
Γ_{final}	final boundary part of $\partial\Omega$
Γ_{initial}	initial boundary part of $\partial\Omega$
ρ	electric charge density, twisted 3-form
ϕ	electric scalar potential cochain
φ	electric scalar potential differential form (0-form)
Φ	time integral of ϕ, i.e., a scalar potential impulse cochain
Ω	domain
$\star_M$	Hodge operator with material parameters taken into account (see equation (5.20))

$\cong$	isomorphism
$\oplus$	direct sum of spaces

Chapter 6

A'	part of A in $C^1(K, L_F)$, i.e., the free part of A
A_{L_F}	part of A on Γ_F, i.e., the part fixed by boundary conditions
$\boldsymbol{A}$	section 6.1: real array representative of A in $\mathbb{R}^{\kappa_1}$ section 6.3: part of $\boldsymbol{A}'$ in edge-cotree (belongs to $\mathbb{R}^{\kappa_1(T_1^*)}$)
$\boldsymbol{A}'$	real array representative of A' in $\mathbb{R}^{\kappa_1^A}$
$\boldsymbol{A}_{L_F}$	real array representative of A_{L_F} in $\mathbb{R}^{\kappa_1(L_F)}$
$\mathcal{B}^p(K, L)$	space of relative finite coboundaries modulo L
$\mathcal{C}_f$	finite mapping from p-forms to p-cochains (see equation (6.15))
$C_p(L_F)$	space of finite p-chains on Γ_F
$C^1(K, L_F)$	relative 1-cochain space for the free part of A, i.e., for A'
$C^2(K, L_F)$	relative 2-cochain space for the free part of F, i.e., for F'
$\tilde{C}_p(\tilde{K}, \tilde{L})$	relative p-chain space for the part of G that is not fixed by G-boundary condition, and has a one-to-one connection with $C_p(K)$
$\tilde{C}_p(\tilde{L})$	space of outer oriented finite p-chains on $\partial\Omega$
$\tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$	relative twisted cochain space for the free part of G, i.e., for G''
$\mathbf{C}$	variable side incidence matrix (see equation (6.41))
$\mathbf{C}'$	equation side incidence matrix (see equation (6.40))
$\tilde{\mathbf{d}}_p$	real array representative of d for $\tilde{C}^p(\tilde{K})$
F'	part of F in $C^2(K, L_F)$, i.e., the free part of F
F_{L_F}	part of F on Γ_F, i.e., the part fixed by boundary conditions
$\mathcal{F}_f$	finite mapping from p-chains to p-forms (see equation (6.16))
$\boldsymbol{F}$	real array representative of F in $\mathbb{R}^{\kappa_2}$
G'	part of G in $\tilde{C}^2(\tilde{K}, \tilde{L})$, i.e., in the interior of the domain
G''	part of G in $\tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$, i.e., the free part of G
$G_{\tilde{L}}$	part of G on $\partial\Omega$, i.e., the part that is known by G-boundary condition, or that can be zeroed otherwise
$\boldsymbol{G}$	real array representative of G in $\mathbb{R}^{\tilde{\kappa}_2}$
$\boldsymbol{G}'$	part of $\boldsymbol{G}$ in $C_p(\tilde{K}, \tilde{L})$ (belongs to $\mathbb{R}^{\kappa_2}$)
$\boldsymbol{G}_{\tilde{L}}$	real array representative of $G_{\tilde{L}}$
$\mathbf{H}_M$	diagonal material matrix (see equation (6.29))
$\boldsymbol{J}$	real array representative of J in $\mathbb{R}^{\tilde{\kappa}_3}$
$\boldsymbol{J}'$	modified source current (see equation (6.42))

$\tilde{K}$	dual (cellular) mesh complex
$\tilde{K}_{\text{eq}}$	dual mesh complex for the instances of the Maxwell-Ampère law, i.e., instances concerning the free part of G not on Γ_{final}
$\tilde{K}_F$	dual mesh complex for such a part of G that is not fixed by F-boundary condition via constitutive relation
$\tilde{K}_G$	set of cells where G is known in advance
L_F	cellular mesh complex on Γ_F, $L_F \subset K$
$\tilde{L}$	dual cellular mesh complex on $\partial\Omega$, $\tilde{L} \subset \tilde{K}$
$\tilde{L}_G$	dual cellular mesh complex on Γ_G, $\tilde{L}_G \subset \tilde{K}$
$\tilde{L}'_G$	boundary cell complex in $\tilde{K}_F$, where G is known by the G-boundary condition or where G can be zeroed (on Γ_{final})
$\mathbf{P}_{\text{eq}}$	projection matrix from the indices of dual 3-cells of $\tilde{K}$ to the indices of 3-cells in $\tilde{C}_3(\tilde{K}_{\text{eq}}) \setminus \tilde{C}_3(\tilde{L}_G)$
$\mathbf{P}_A$	projection matrix from $\boldsymbol{A}$ to $\boldsymbol{A}'$
$\mathbf{P}_G$	projection matrix from $\boldsymbol{G}$ to $\boldsymbol{G}'$
$\mathbf{P}_{T_1^*}$	projection matrix that chooses cotree edge parts from $\boldsymbol{A}'$
$\mathbf{Q}_A$	projection matrix from $\boldsymbol{A}$ to $\boldsymbol{A}_{L_F}$
$\mathbf{Q}_G$	projection matrix from $\boldsymbol{G}$ to $\boldsymbol{G}_{\tilde{L}}$
T_p	spanning p-tree modulo L
T_p^*	p-cotree modulo L
$\mathcal{Z}^p(K, L)$	space of relative finite cocycles modulo L
Γ_F	part of $\partial\Omega$ for the boundary condition of F
Γ_G	part of $\partial\Omega$ for the boundary condition of G
κ'	coefficient function for the finite constitutive relation (see (6.22))
κ	coefficient function for the finite constitutive relation with orientations taken care (see equation (6.26))
κ_i	number of i-cells in K
κ_1^A	number of 1-cells in the basis of $C^1(K, L_F)$, i.e., where A' lies
κ_1^{eq}	number of dual 3-cells in the basis of $\tilde{C}_3(\tilde{K}_{\text{eq}}) \setminus \tilde{C}_3(\tilde{L}_G)$, i.e., number of equations in the A-formulation before gauging
$\kappa_1(L_F)$	number of 1-cells in L_F
$\kappa_1(T_1^*)$	number of 1-cells in the coedge-tree T_1^*
$\kappa_1(\tilde{T}_3)$	number of 1-cells in the dual volume-tree $\tilde{T}_3$
$\tilde{\kappa}_i$	number of i-cells in $\tilde{K}$
$\tilde{\boldsymbol{\partial}}_p$	real counterpart of the boundary operator of $\tilde{K}$

Chapter 1

Introduction

My secondary-school physics text book says: "*the universe consists of matter and electromagnetic radiation*". This thesis deals with the latter.

Besides the functioning of the universe, most human activities are based on electromagnetic waves, such as most of information technology. Study of electromagnetic waves is thus vital for the development of technology, and more broadly, for a better understanding of our surroundings. This thesis examines a modern way of researching electromagnetic waves—by numerical computation. The idea is to do computation in a new way: directly in the four-dimensional scope, where space and time are combined into one unity, spacetime. The motivation lies in developing current computation methods with new ideas originating from computation made in spacetime. Consequently and hopefully, some serious limitations and downsides of the current methods could thereby be circumvented.

Electromagnetic wave computation methods are divided into two main branches: first, the time-harmonic methods and, second, the more general-purpose time-stepping methods. The former are adequate for a large number of wave problems, but often a non-time-harmonic method is necessary, e.g., with transient problems. We concentrate on the general non-time-harmonic case.

Time-stepping means that electromagnetic fields are computed only in a finite number of moments, called *time-steps*. For each step, fields are computed from previous field values with the aid of Maxwell and constitutive laws. Because a full field problem is solved in each step, the finite element method is excessively laborious for non-time-harmonic computation. Consequently, difference-based approaches, such as the *finite integration technique* [88] and the *cell method* [81, 83, 2], are more common. We call them both by the common name *geometric methods*.

In the geometric methods, space and time discretizations are essentially

done differently. Spatial connections are done with integral type equations, and temporal derivatives are approximated with time differentials that lead to leap-frog type time-stepping. Here the focus of this thesis differs from the traditional methods: the spatial and the temporal domain are combined into a part of a four-dimensional spacetime, and all approximations are similar. The formulation and thus the element meshes are fully four-dimensional and exploit the non-Euclidian geometry of spacetime.

For spacetime computational electromagnetics, we must build a model for spacetime and necessary structures to model field quantities. Because spacetime models arise from the theory of relativity, we must take a brief look at general relativity. A major topic in this work, therefore, is to collect and partly develop the mathematics behind spacetime modeling and the electromagnetics in spacetime.

We do not intend to transfer all computation into spacetime once and for all, but to discover new aspects and ideas for wave computation. Geometric methods are known to have some weaknesses and limitations, solutions to which have been sought for quite some time. We try to start our examination from a fresh point of view and overcome some of the obstacles based on ideas originating from spacetime computation. However, because of the breadth of the topic, we cannot solve all the problems, but at least we can introduce some tools to develop numerical computation and concentrate on some selected challenge problems.

The spacetime perspective for the geometric methods is not entirely new and has been adopted, e.g., in [48, 82, 74]. In fact, these three publications were the starting point of this work (and later also [68]). However, the purpose of this thesis is to go beyond these works and use a strictly four-dimensional formulation. A parallel method for the geometric methods, called the *discontinuous Galerkin method*, have already been used in true spacetime formulations [57]. However, our method have some advantages compared with this method; especially, discontinuous Galerkin lacks continuity properties intrinsic to our method.

This thesis has been greatly influenced by H. Whitney's *Geometric Integration Theory* [91] and the subsequent publications of algebraic topology applied to electromagnetics, e.g., see [34, 47, 48, 73, 74, 75, 82], and nearly all the recent publications of Alain Bossavit. Additionally, for the relativistic point of view, the main information sources have been the *Gravitation* of Misner, Thorne, and Wheeler [51], the *Spacetime Physics* of E. Taylor, and J. Wheeler [77], and the *Semi-Riemannian geometry* of B. O'Neill [53].

Chapters two and three provide a detailed background on this work, both in computational electromagnetics and theory of relativity. First, chapter two examines the geometric methods comprehensively and introduces some

challenge problems to illustrate the possibilities spacetime computation could offer. Some of the problems are examined in depth in the subsequent chapters. Secondly, chapter three introduces the theory of relativity without advanced mathematics. The chapter provides the motivation for our choosing the particular mathematical structures to model spacetime.

The fourth chapter builds the mathematical structures needed in the rest of the thesis. Because the geometry of our spacetime model is somewhat unfamiliar to engineers in electromagnetics, the chapter gives a comprehensive view of manifold theory and metric-dependent structures. Furthermore, because of the divergent metric, many well-known mathematical concepts must be rethought and redefined in order to generalize them into spacetime. This thesis delves extensively into the norms and topologies of chain and cochain spaces in spacetime.

Next, chapter five makes use of the above structures to construct a model for the electromagnetic field problem in spacetime. The following chapter introduces numerical methods for solving the problems and demonstrates the algorithms in two-, three-, and four-dimensional spacetimes with various examples. The last chapter discusses and concludes on the findings, including possible solutions to the challenge problems.

Each chapter and section begins with a synopsis of the contents. Note that the language is then somewhat imprecise and terms are used undefined, if they are well-known in electromagnetics or mathematics.

Because of lack of time and space, we were forced to exclude some relevant subjects. For example, we have omitted topologically complicated cases; i.e., we do not deal with domains with holes, tunnels, and such, because we forbid (relative) nonbounding cycles. However, most of the structures used are already feasible for topologically complex domains. Second, we have also excluded convergence and error bound analysis of our numerical method. These analyses would have been excessive hard without proper norms for cochain spaces.

Chapter 2

Background: numerical methods

If we want to develop current numerical methods, we need to know their present state and limitations. First, we briefly look at the history of numerical electromagnetic computation and along the way review the linkages between the various methods.

As mentioned in the introduction, we concentrate on time-dependent problems and omit the frequency domain variants of the methods. This follows from the fact that spacetime computation is inherently linked with time domain computation, because the frequency domain methods suppress the time direction with sinusoidal-form time symmetry.

Second, we discuss the geometric methods in detail and first review their current state and limitations. In this work, we aim to develop these methods further and find a possible solution to some of their limitations. In the last section, we introduce some challenge problems for our method. Some of the problems are solved in this thesis, and for the rest we develop tools for solving them.

2.1 Review of numerical methods

The *finite element method* (FEM) with nodal elements was invented in the 1960's in solid mechanics and introduced to electromagnetics in about 1968 by P. Silvester [66]. Because the nodal-element-based method lacked a tangential continuity of fields, results were imprecise with three-dimensional problems. For example, the method introduced so-called spurious modes [17]. The introduction of edge elements—and later the rest of the Whitney elements—finally solved these problems [12, 52]. In the edge-element FEM, the degrees

of freedom are located on the edges of a simplicial mesh, and the continuity requirements are satisfied naturally.

At the same time, yet independently, Yee-like schemes were developed. The first time-domain difference method was introduced by K. Yee in 1966 [92]; the method was dubbed the *Finite-difference time-domain* (FDTD) by A. Taflove in 1980 [71]. The method has two staggered grids Euclidean orthogonal to each other. In fact, the grids may be interpreted as four-dimensional staggered grids divisible into a staggered spatial grid and a leap-frog-scheme in the time direction. The variables are located at the center of the edges or facets of the two spatial meshes.

A method parallel to FDTD was introduced by T. Weiland in 1977 [87], called the *finite integration technique* (FIT). Later in 1996 in [88], Weiland formulated FIT with integrated variables over edges or facets. Like FDTD, FIT was originally restricted to two cuboidal orthogonal meshes. However, with certain constraints, FIT can be somewhat generalized to other type of meshes. For temporal variations of variables, the basic version of FIT uses a difference-based leap-frog time-stepping method identical to FDTD. However, at a fixed time instant, the variables are integrated over spatial domains, in contrast to FDTD, where the field components are dealt pointwisely. Nonetheless, in a pair of orthogonal staggered grids, FIT and FDTD become numerically identical [88]. However, conceptually they are so different that we cannot justify their being equal. In contrast, FIT and the *cell method*, introduced by Tonti [82, 46], are concordant. In other perspective, the cell method is a generalization of FIT. Unwilling to take sides on the subject, we simply call them both *geometric methods*.

Even if the edge-element FEM and the geometric methods seem distinct, they are, in fact, similar methods. The first order FEM (and naturally also FDTD) can be written with the same formalism as the geometric methods, which differ only in their constitutive matrices. In FEM, we have a mass matrix and a stiffness matrix, and in the geometric methods with an orthogonal pair of meshes, a diagonal material matrix. In fact, a Yee-like method could be introduced by using a mass matrix [8], but the method becomes non-explicit, because at each time-step we must solve a system of linear equations. [1] In an orthogonal hexahedral grid, the mass matrix can be lumped into a diagonal matrix [59], but in a tetrahedral mesh mass lumping fails [11].

Both FEM and the geometric methods are used with low and high frequency problems. However, FEM suits better static or low frequency cases,

[1] The non-explicit nature of FEM is the main reason why (continuous) Galerkin methods are not included in the geometric methods here. This distinction is quite artificial from the mathematical point of view, but from the engineering point of view the distinction by the numerical efficiency is justifiable.

whereas the geometric methods are ideal for high frequency wave problems. The merit of FEM over the geometric methods is its better accuracy of constitutive laws (and more detailed energy-based convergence results); however, this comes with a penalty in numerical efficiency, inherited from the non-diagonal constitutive matrices. This is not a drawback with low frequency problems, where a linear equation must be solved only once, or, at worst, a few times. However, with wave problems, where a matrix equation must be solved at each time-step, the raw efficiency of the geometric methods becomes decisive. The lost accuracy can always be compensated for with a denser mesh than in FEM.

Remark 2.1. However, there is a third alternative between FEM and geometric methods: the *discontinuous Galerkin method* (DG) [32]. The DG method resembles FEM, but has local basis functions, which leads to a (block-) diagonal mass matrix. Thus, this induces an explicit method for wave problems, similar to the geometric methods, but has more freedom with mesh element shapes. However, DG lacks the intrinsic continuity and conservation properties of the geometric methods.

Since the DG method have been extended to a true spacetime formulation (e.g., see [57]), it can be seen as an alternative to our method. Due to the limitations of the DG method, we look for a second alternative.

Because this work deals with a geometric method in spacetime, we focus on wave problems. However, even if the strengths of the method lie with high frequency problems, it can also be used with low frequency problems. Since there are methods better suited for low frequencies, we discontinue the topic here and concentrate on electromagnetic wave problems.

2.2 Geometric methods

From one point of view, our main topic is the merging of spatial and temporal discretizations into a combined spacetime discretization. Hence we must discuss how discretizations are performed in the traditional geometric methods. Additionally, because we aim to develop the geometric methods, we must discuss the limitations of its present-day versions and briefly review earlier efforts to solve the limitations.

2.2.1 Spatial discretization

Geometric methods of electromagnetic wave problems use as computation variables the integrals of electric and magnetic fields over the edges and

facets of a mesh. For example, a pointwise electric field is integrated over the edges of a mesh, whereas magnetic flux density is integrated over the faces of the same mesh, see figure 2.1. We denote by e_i the integral values of the

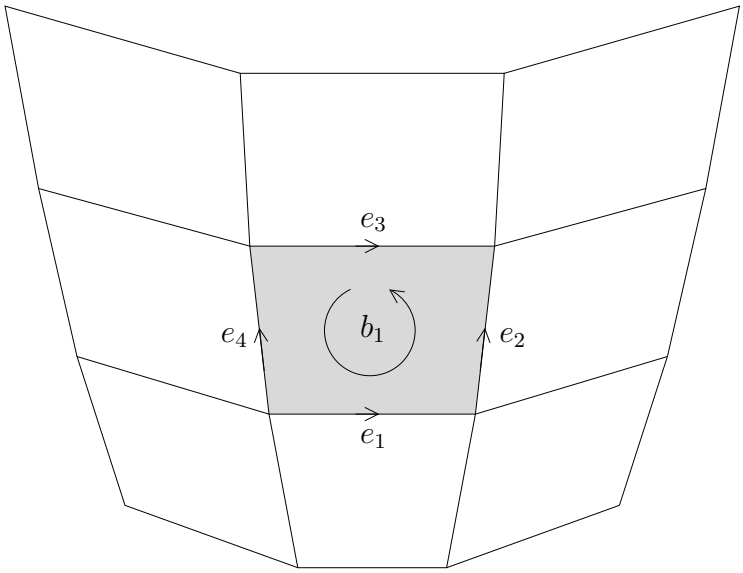


Figure 2.1: Part of a primal mesh: the electric field variables e_i are located on the edges and the magnetic flux variables b_j on the facets. A facet and its boundary edges are oriented according to the arrows. The Faraday law connects b_1 on the facet with e_i's on its boundary.

electric field over enumerated edges, similarly by b_j the magnetic fluxes. We can apply the Faraday law to each facet and its boundary and get—for the case in figure 2.1—the equation

$$e_1 + e_2 - e_3 - e_4 = -\partial_t b_1. \tag{2.1}$$

The signs derive from the orientation of the facet and its boundary—and how they match. This equation contains no approximations; it is only an instantiation of the Faraday law on this facet. If we collect all e_i-variables into a real array $\mathbf{e}$ and all b_i-variables into $\mathbf{b}$, and collect the instances of the Faraday law on each facet into one relation, we get

$$\mathbf{Ce} = -\partial_t \mathbf{b}, \tag{2.2}$$

where $\mathbf{C}$ is an *incidence matrix*, or a *curl matrix*. $\mathbf{C}$ includes the signs of e_i's and thus contains only values ± 1 and 0. The equations of type (2.2) are called *network equations* [7] (or *topological equations* [82] or *Maxwell's grid*

equations [88]), because they depend only on the topology of the mesh and not on metric properties.

All the above quantities are related to inner oriented geometrical objects; i.e., e_i depends on the direction of the edge and b_j on the positive direction of circulation on the facet. For the quantities in the Ampère law, we need a second mesh where the edges and facets are outer oriented; i.e., a facet is oriented by its crossing direction and an edge by the circulation of its transversal plane. The first mesh is called the *primal mesh* and the outer oriented one the *dual mesh*. Magnetic field density is integrated over the edges of the dual mesh, and electric flux density and electric current density over the dual facets—the integral values are denoted by h_k, d_l, and j_l, respectively. For a dual facet, we can write the Ampère law in the same form as (2.1). If we collect h_k's, d_l's, and j_l's into the real arrays $\mathbf{h}$, $\mathbf{d}$ and $\mathbf{j}$, the Faraday law on the dual facets can be written as

$$\tilde{\mathbf{C}}\mathbf{h} = \partial_t \mathbf{d} + \mathbf{j}, \tag{2.3}$$

where $\tilde{\mathbf{C}}$ is the curl matrix of the dual mesh. Like equation (2.2), this one depends only on the topology of the mesh and not on any metric properties.

The constitutive relations are metric-dependent and contain approximations. In the geometric methods, we choose meshes that are dual with each other, i.e., every dual edge pierces exactly one primal facet, and each primal edge pierces a dual facet. The edge direction must be complementary to the facet direction, and, in the basic version, they are assumed to be *orthogonal*. Figure 2.2 shows an *e*-*d*-relation. We assume that at the piercing point, the permittivity is ε_i. We do not know the fields' pointwise values at that point but only their integrals over the edge or facet. We must approximate and take the average of the fields on the edge or facet; in the notation of figure 2.2, we get the pointwise approximation e_i/l_i for the electric field and $b_j/\tilde{A}_i$ for the magnetic flux and thus the finite constitutive relation

$$\mathbf{d} = \mathbf{M}_\varepsilon \mathbf{e}, \tag{2.4}$$

where $\mathbf{M}_\varepsilon$ is a diagonal *material matrix*, with

$$[\mathbf{M}_\varepsilon]_{ii} = \frac{\varepsilon_i \tilde{A}_i}{l_i}. \tag{2.5}$$

For the *b*-*h* pair we get a similar result,

$$\mathbf{b} = \mathbf{M}_{\frac{1}{\mu}} \mathbf{h}, \tag{2.6}$$

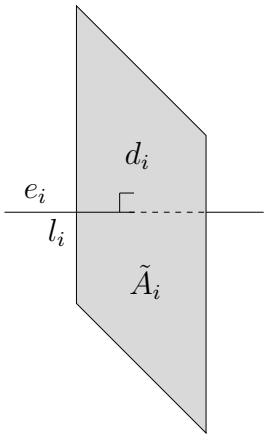


Figure 2.2: A primal edge and its dual facet in a pair of dual meshes. Because the primal edges and the dual facets are connected one-to-one, they can be enumerated with the same indices. Additionally, the edge-facet pairs are orthogonal. The area of the dual facet is $\tilde{A}_i$, and the length of the edge is l_i.

where $\mathbf{M}_{\frac{1}{\mu}}$ is also diagonal, with

$$[\mathbf{M}_{\frac{1}{\mu}}]_{ii} = \frac{A_i}{\mu_i \tilde{l}_i}, \tag{2.7}$$

where A_i is the area of the i^{th} primal facet and $\tilde{l}_i$ the length of the corresponding dual edge; μ_i is the permeability at their crossing point.

If the edge-facet pairs are not orthogonal, the material matrices cannot be diagonal, provided we want to keep the method as accurate as in the orthogonal case. In the non-orthogonal case, matrices can be constructed in multiple ways (see, e.g., [8, 14, 21, 46, 63]).

2.2.2 Temporal discretization

In electromagnetic wave problems, the fields are time-dependent. Traditionally, time-dependence is dealt with by assuming that the field variables $\mathbf{e}$, $\mathbf{b}$, $\mathbf{h}$, $\mathbf{d}$, and $\mathbf{j}$ are time-dependent real arrays. In most cases, we can assume that the geometry and material parameters are constant in time and that the meshes are thus independent of time, as in figure 2.3.

In the above, discretization is performed only spatially; i.e., the time derivatives in equations (2.2) and (2.3) assume that field variables are continuous (and differentiable) functions of time. A standard solution to time dis-

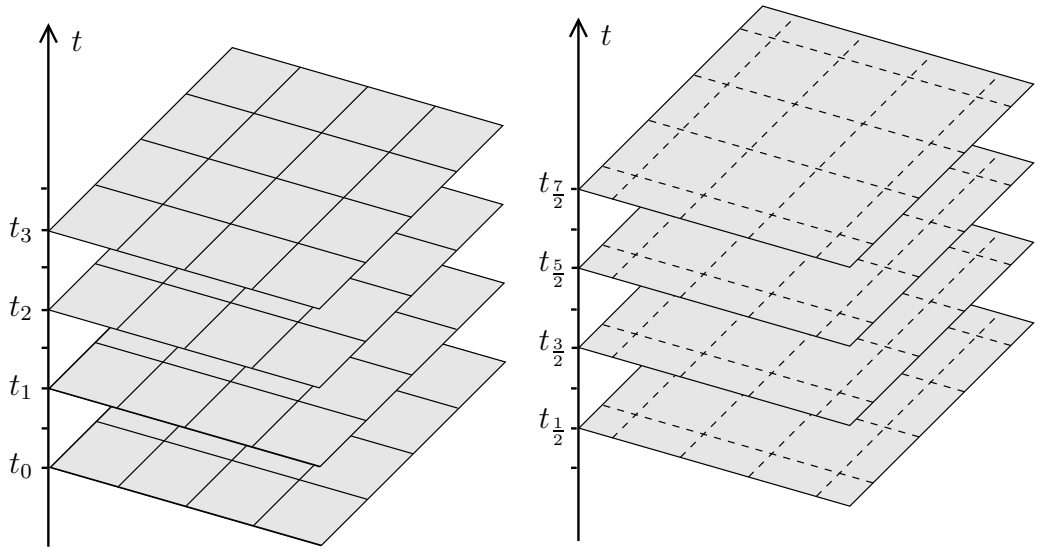


Figure 2.3: Primal and dual mesh at different time instants; the meshes are independent of time. For the primal mesh, facet variables are computed at primal time instants t_i, $i = 0, \ldots, N$, and for the dual mesh, at dual time instants $t_{i+\frac{1}{2}}$, $i = 0, \ldots, N$. Primal and dual edge variables are computed at intervals between corresponding time instants. Because the dual mesh instants are in the center of the primal time intervals, we have a staggered pair of time-grids.

cretization in all geometric methods is the leap-frog-scheme[2] , in which variables are computed at a finite set of *time instants* and at *time intervals* between the instants. The time difference between two sequential time instants is assumed to be constant; the constant is called the *time-step* and denoted by Δt. The time instants of primal mesh variables are denoted by $t_i = i\Delta t$ and the time intervals between them by $\Delta t_{i+\frac{1}{2}}$, with $i = 0, \ldots, N$. Dual mesh variables, however, are computed in the middle of primal time instants at $t_{i+\frac{1}{2}} = (i + \frac{1}{2})\Delta t$ or at dual time intervals Δt_i, again with $i = 0, \ldots, N$ (see figure 2.3). Notice that the center of a primal time interval coincides with a dual instant, and similarly the center of a dual interval and a primal instant agree. Consequently, primal instants and dual intervals are indexed with integers and dual instants and primal intervals with half-indexes.

Motivation for the above pair of staggered time-grids is in the leap-frog

[2]Of course, there are many more elaborate timestepping schemes than the standard leap-frog-scheme—e.g., the Newmark-Θ scheme from [60, 21]. However, the leap-frog-scheme is enough for the background information for this thesis, because we end up to a spacetime method resembling the leap-frog timestepping.

integration method. In a primal spatial mesh, the magnetic field is located at time instants and the electric field at time intervals; the real array variables at these instants and intervals are denoted by $\mathbf{b}^k$ and $\mathbf{e}^{k+\frac{1}{2}}$. In a dual mesh, $\mathbf{h}^{k+\frac{1}{2}}$ and $\mathbf{j}^{k+\frac{1}{2}}$ are located at dual instants and $\mathbf{d}^k$ at the dual time interval. The leap-frog-scheme approximates the time time-derivatives of the Faraday and Ampère laws (2.2) and (2.3) with a central difference; i.e., we have

$$\mathbf{b}^{k+1} = \mathbf{b}^k - \Delta t \mathbf{C}\mathbf{e}^{k+\frac{1}{2}}, \tag{2.8}$$

$$\mathbf{d}^{k+\frac{1}{2}} = \mathbf{d}^{k-\frac{1}{2}} + \Delta t(\mathbf{C}^T\mathbf{h}^k - \mathbf{j}^k), \tag{2.9}$$

where the electric current $\mathbf{j}^k$ is a source term known in advance. The above equations together with constitutive relations (2.4) and (2.6) forms the *leap-frog-scheme*. The scheme is shown graphically in figure 2.4. We use update

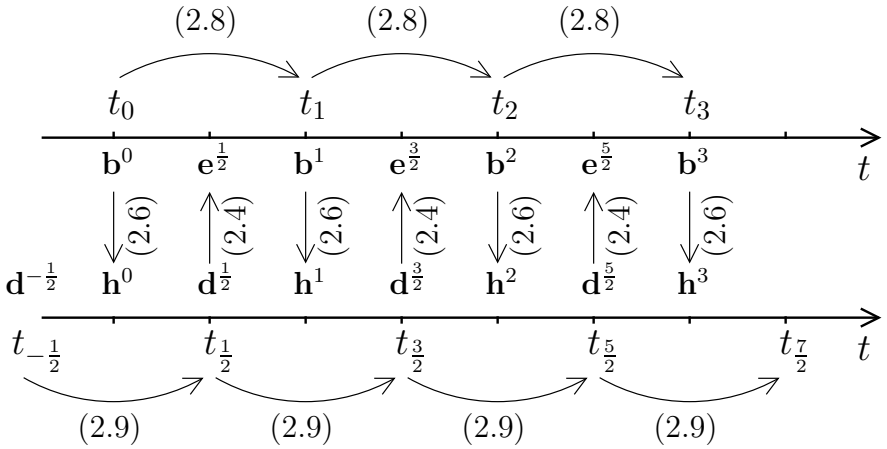


Figure 2.4: Leap-frog-scheme. The values of $\mathbf{b}$ and $\mathbf{d}$ are updated by turns with Faraday and Ampère laws (2.8) and (2.9). The constitutive relations (2.4) and (2.6) are used to transform primal variables into dual ones and vice versa. The arrays $\mathbf{b}^0$ and $\mathbf{d}^{-\frac{1}{2}}$ are known by initial conditions, and $\mathbf{j}$ is a known source term.

equations (2.8) and (2.9) by turns and assume that $\mathbf{b}^0$ and $\mathbf{d}^{-\frac{1}{2}}$ are known by initial conditions; by constitutive relation (2.6) we get $\mathbf{h}^0$. We can use $\mathbf{d}^{-\frac{1}{2}}$ and $\mathbf{h}^0$ to solve $\mathbf{d}^{\frac{1}{2}}$ by (2.9). Then, we get $\mathbf{e}^{\frac{1}{2}}$ from constitutive relations (2.4) and, after that, $\mathbf{b}^1$ can be solved from (2.8). Now we are where we started and can restart the process with (2.6) and (2.9) (see figure 2.4).

To keep the method stable, we must choose a time-step shorter than the upper limit given by the Courant condition [72, 88]

$$\Delta t \leq \frac{\sqrt{\epsilon\mu}}{\sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2}}}, \tag{2.10}$$

where Δx, Δy, and Δz are the grid sizes in the spatial dimensions.

2.2.3 Extensions and limitations

We described above only the most elementary version of the geometric methods. The past three decades have witnessed a huge number of different generalizations and extensions, which cannot all be discussed here, but we choose a couple that are germane to this thesis. In the following, we demonstrate two limitations of the basic method and make an effort to overcome them.

Conformal FIT

As explained above, the geometric methods are explicit only with a pair of orthogonal meshes. Thus for a numerically efficient method, the standard solution is to use Cartesian grids. Other types of orthogonal meshes have been suggested [9, 41], but the forms of possible orthogonal space-fillers are quite restricted. Hence in most cases, the Cartesian grids are used.

Cuboidal orthogonal meshes are not perfect for modeling curved boundaries or material interfaces because of the so-called *staircase problem*, which means that those surfaces must be approximated with a rough staircased mesh surface. Additionally, since many models have tiny details in some parts of the domain and large empty spaces in other parts, mesh size should be variable from point to point.

Fortunately, several attempts have been made to solve the above problems. In the *subgridding method* [80, 54], the mesh is locally subdivided into smaller cuboidal elements and the stairs are diminished. Because we must use interpolations in the layer between the coarse and the dense mesh, this naturally causes inaccuracy to the method. Additionally, if the standard leap-frog-scheme is used, by (2.10) a smaller time-step must be used in the whole mesh, which makes the method less effective.

A second alternative is the *conformal method*, where mesh elements can be freely divided into two parts of different material. On the boundary surface, one material can be a *perfect electric conductor*. The constitutive laws are applied by stressing the material parameters by their relative lengths or areas in the element [42, 79, 94]. In the boundary case, this effectively shortens the edge sizes of the elements, and a smaller time-step is thus needed, once again. Downsizing the time-step could be avoided with the so-called *uniformly stable conformal scheme*, where the time-step does not need to be reduced; instead, the material matrix becomes non-diagonal in the boundary regions [95]. With some extra approximations, this non-diagonality could we avoided [96].

Local time-stepping

By the Courant condition (2.10), the maximal time-step is bounded from above by the size of the smallest element in the whole mesh. In the basic leap-frog-scheme, this same time-step size is used in the whole mesh area. Thus the use of locally dense meshes becomes numerically inefficient, because the number of needed time-steps grows unnecessarily. A locally dense mesh needs, additionally, either extra approximations, with subgridding methods, or superfluous amount degrees of freedom in places where they are not required.

The above problem with the time-step has been tackled with local time-stepping methods. For example, subgridding schemes use a small time-step in the dense mesh and a large one in other regions [80]. Of course, this creates extra difficulties on the boundary of different time-step regions. On these boundaries, fields are computed at different time instants on the different sides of the boundary. To connect them, we need to interpolate the field values of one time-grid into another—and possibly in the reverse direction. Such interpolation, naturally, adds to approximation errors and instability (e.g., see [15]).

2.3 Challenge problems

In our method, four three-dimensional electromagnetic fields and fluxes are combined into two four-dimensional fluxes. Thus for computation, we build a four-dimensional spacetime mesh, where the temporal dimension does not differ from the spatial ones. Additionally, we use metric-independent network equations with the four-dimensional problem, similar to (2.2) and (2.3) in the three-dimensional case. There is nothing specific about the time direction in the network equations either; the uniqueness of the temporal direction centers around the constitutive laws, i.e., the metric.

We settle on a method that closely resembles a static three-dimensional formulation but lies in a four-dimensional space. The difference is embedded in the metric and in the concept of orthogonality inherited from the metric. Hence we solve a single system equation and gain a solution to the fields at a finite set of locations and time instants. The spacetime constitutive law is based on the orthogonality of the facets in two meshes. However, in spacetime the concept of orthogonality differs markedly from the Euclidian and gives us more freedom to construct our method.

If a spacetime mesh is constructed by extruding a three-dimensional mesh in the time direction, our method is almost identical with current geometric

methods. However, this mesh is only a special case of general spacetime meshes, and we try to exploit this freedom and create a generalization of the geometric methods.

In the following, we present some problems which are tricky for standard geometric methods, but which could be solved with our method. We return to their solvability at the latest in the final summary. Unfortunately, due to lack of time inherited from problems mentioned in the Preface, we cannot solve all of them during this thesis.

Local time-stepping and the locally dense mesh

As mentioned above, standard geometric methods require some extra approximations by interpolation, if we use a locally dense mesh or local time-stepping, because their possible forms of mesh elements is considerably restricted. Could a pair of spacetime meshes with spacetime orthogonality allow more freedom and enable use of locally dense meshes and a time-step size that varies from point to point according to mesh size? Then the mesh would be spacetime adaptive.

If we can build a proper spacetime mesh, would it still be interpreted three-dimensionally, or should it be truly four-dimensional?

Lightcone mesh

Consider a sphere wave from an impulse source in a vacuum, i.e., a single wavefront travelling outwards from one point in space. In spacetime language, the four-dimensional cone-shape of the wavefront is called the lightcone. Because on the outside and inside of the lightcone fields vanish, we can build a mesh only around the lightcone and drop the number of degrees of freedom to a fraction of the original. Thus we would have considerably fewer operations in the solution process than with a non-lightcone mesh. From a three-dimensional point of view, this computation would be done in a mesh that moves along the wavefront, and where no mesh would exist in front of and behind the wavefront—i.e., like in the above case, the mesh would be spacetime adaptive.

If this method would give us a result for a point source, we could solve a more general source term by using Green's functions. Of course, this assumes that the source is far from the boundaries or material interfaces. Then lightcone meshes would make for more efficient computation.

This challenge problem is quite academic, but it demonstrates the opportunities the spacetime geometric method could provide. Additionally, lightcone meshes are quite easy to implement, as we see in chapter 6.

Moving geometry

If the geometry of the problem changes enough as a function of time, the mesh must also change as a function of time. In the three-dimensional perspective, this is a troublesome task. However, in the spacetime perspective, a moving geometry should not pose extra problems. Because the time direction does not essentially differ from the spatial directions, the mesh could change as a function of time, as it can also change along, e.g., the x-direction. Thus a problem with moving geometry is a truly four-dimensional problem and should be easy to solve with a spacetime computation method.

A rotating machine, e.g., a rotating radar, antenna, or electric motor, is an example of a moving geometry. These rotations should be easy to solve with a spacetime geometric method. However, if the angular velocity is constant and the whole geometry rotates, the rotation forms a helical symmetry in spacetime. By this symmetry, the problem could be turned into a three-dimensional problem, but not necessary into a classical electromagnetic field problem.

Parallelization

Because the time-stepping scheme reduces to a solution of a single system of linear equations, the parallelization of the scheme becomes the parallelization of a problem in numerical linear algebra. Thus could standard linear algebra parallelization packages be used instead of a specific parallelization scheme made for the time-stepping?

Relativistic computational electromagnetism

If our spacetime model could include curvature, we could do computational electromagnetics in the presence of strong gravitational effects, a definite advantage in some cosmological or black hole models. However, from the engineering point of view, we cannot yet envisage applications of these models in the near future.

Chapter 3

Origin of spacetime: relativity

We aim to do computational electromagnetics in a four-dimensional spacetime. For foundation of this computation, we need a mathematical model for spacetime. Fortunately, we do not have to develop one from scratch, because in the last century a proper model was already developed in the field of modern theory of relativity, and a separate branch of mathematics has grown to aid that development. We only need to gain a brief understanding of the basics of the theory of general relativity and pick the results and models we need. Relativity also gives us an intuition and motivation as to how things are dealt with in spacetime modeling. Our emphasis is not on the theory itself but on the mathematical structures behind the model of spacetime.

Let us briefly introduce relativity, because we cannot assume an electrical engineer to be familiar with the subject. Especially, let us look for the motivation for using the manifold theory in general relativity, because the need for manifolds—or the need for a variety of coordinate charts—springs from the physical theory itself.

General relativity is a theory of gravity—of how bodies react in gravitational interactions. In the global perspective, general relativity is somewhat complicated, but, fortunately, locally the physics of gravity is simple. Locally, the universe is gravitation-free for an inertial observer[1]; i.e., in a small enough room and small enough elapsed time we—in free fall—cannot detect the effects of gravitation. This local theory, e.g., the theory of relativity in the absence of gravitation, is called *special relativity*. [77]

As a consequence of the above facts, the universe can be locally modeled by special relativity. On the other hand, the manifold is a mathematical structure that models the regions of space (or spacetime) so that they locally resemble a simple space. We can immediately see that this simple space in the

[1] An observer in free-fall.

case of spacetime manifolds is a Minkowski space, the mathematical model of spacetime in special relativity. Hence we must begin with the basics of special relativity to understand why we decide to use the particular mathematical model of spacetime we are introducing in the next chapter. In the end, we briefly introduce general relativity to get a global picture of the spacetime model.

Further, because this chapter stands for a motivation for mathematical structures and appears before the introduction of any delicate mathematical structures, we keep the mathematics as simple as possible. This chapter is heavily influenced by *Spacetime Physics* by E. Taylor and J. Wheeler [77].

3.1 Gravitation-free spacetime

Special relativity discards the notions of absolute space and time familiar from Newtonian mechanics by stating that distance and time depend on the observer. This character, which makes some quantities relative to the observer, also gives the name, i.e., relativity. While making some quantities—somewhat surprisingly—relative, the theory also makes some others absolute, i.e., the same for all observers, especially the speed of light in a vacuum. In addition, relativity yields the equivalence of matter and energy.

3.1.1 Basic definitions

To speak about gravitation or relativity—or even about the concept of spacetime—we need some basic definitions. Some definitions in this chapter may be somewhat imprecise or informal, because we try to keep the mathematics as simple as possible. If necessary, we will redefine notions in a more formal way in the following chapters.

What is spacetime? Eventually, we choose to model it with a Lorentzian manifold but, for the time being, we are satisfied with a definition that is adequate for special relativity; i.e., we build a model for a gravitation-free spacetime. The gravitation-free spacetime is a four-dimensional space, which combines the three spatial dimensions and one temporal dimension. In other words, if we have a spacetime frame, a point in spacetime can be represented by three space coordinates and one time coordinate.

A spacetime point—or location—is called the *event*. For example, an event can be an explosion of a firecracker, the collision of two particles, or the emission of a flash of light from an atom or from a spark plug firing.

The determination of the space and time coordinates of an event in one frame is not as straightforward as it seems. From basic physics we know

that information has a limited traveling speed. For this reason, it is not easy to detect which events occur at the same time. To get the idea, think of constructing a cubical latticework of sticks with a clock sitting at every intersection of crossing sticks. These clocks remain in their fixed locations (from the point of view of this frame) and measure time at these fixed places. The clocks can be synchronized by choosing one of them as reference clock and by sending a flash of light from that clock. By the known speed of light and the known distances of the clocks from the reference, all the other clocks can be synchronized with the reference clock. Now an event can be located by the coordinates of the nearest clock in the latticework and by the time the clock is showing, i.e., by four numbers (x, y, z, t) describing the coordinates in that frame.

In the language of the above latticework, an *observer* is the person who collects the data from all the recording clocks of one frame. A more practical way to think is that the observer is the name of the traveller going through spacetime who determines one spacetime frame at every instant in the process. Two events occur in the same place in that frame, if the events are located at the same distance and direction from the observer, but possibly at a different time.[2] If the observer moves relative to some other frame, the meaning of being in the same place differs between these two frames.

3.1.2 Spacetime diagram

We already noticed that we must somehow visualize the spacetime notions. For comprehensive visualization, we should have four-dimensional graphics, which, however, is quite impossible for humans to perceive.[3] We limit our visualization to two-dimensional examples, which have only one spatial dimension and one time dimension. Basically, this means that everything is assumed to happen on one line in space. On most occasions, this is enough for us.

A two-dimensional spacetime drawing is called the *spacetime diagram*—or the *spacetime map*. It resembles a geographic map with the south-north and west-east dimensions, but now one of the directions is time instead of a spatial direction. The location and time of an event in the diagram come from one frame, i.e., from one latticework of clocks. If we change from one frame to another, we get a different diagram.

[2] An observer measures or observes, but does not see because of the finite speed of light that messes up a direct visual observation.

[3] Several attempts have been made, e.g., the Schlegel diagram [62] and the net of a polychoron [64].

There are some established conventions for drawing spacetime diagrams. One event is chosen as the reference event, and thus it is given a zero position in space and zero time. Hence it is located at the origin of the spacetime map. Further, it is customary to choose spatial location as the horizontal axis and time as the vertical axis. Figure 3.1 shows an example of a spacetime map and events placed in the same place or at the same time.

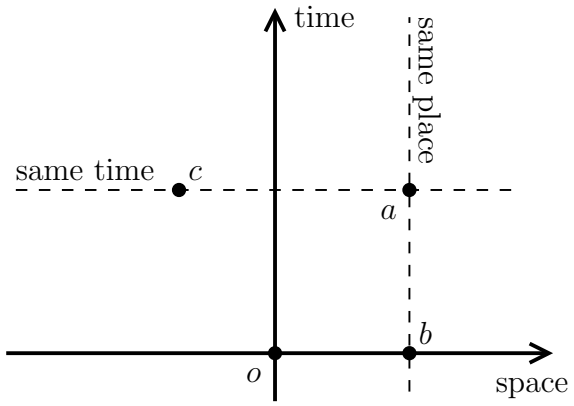


Figure 3.1: The spacetime diagram. The spatial axis is horizontal and the time axis vertical; the reference event o is located at the origin. In this frame, events a and b occur in the same place, and events a and c at the same time. Equality of place and simultaneity are frame-dependent properties.

Additionally, in spacetime diagrams, it is customary to use the same unit for both spatial distance and time. In fact, henceforth we choose to carry on the same custom and place space and time on an equal footing. We are familiar with the fact that years are used to measure distance, namely light years. A light year is the distance that light travels in empty space in one year. For our purposes, this is an oversized unit, but, fortunately, in the same manner meter can be used as the unit of time. One meter of time is the time that light takes to proceed one meter in a cavity. In conventional units, it is $1/c = 1/(2.998 \times 10^8)\,\mathrm{s} \approx 3.336 \times 10^{-9}\,\mathrm{s}$. The speed of light (in a vacuum) in meters per second, denoted by c, is only a conversion factor between meters and seconds with no deeper physical significance. In fact, c is only a historical accident in humankind's choice of units. Note that the speed of light in meters per meter of time is a unit-free constant and has a unit value. [77]

3.1.3 Worldline

A spacetime diagram can give a global picture of all significant events and, especially, if all of them come from the history of one particular particle, we can see the particle's history of travel. The thread connecting the events of the particle's history is the *worldline* of the particle. A worldline has its own unique existence in spacetime, independent of any frame by which we may choose to describe it. The same applies also to events. In a spacetime diagram, a worldline can be represented by a line, but, strictly speaking, it is not the worldline itself but just an image of the worldline in the spacetime diagram.

Figure 3.2 shows two worldlines in a spacetime diagram. The frame where the diagram is plotted is chosen so that two worldlines cross at the event of zero coordinates. The worldline that passes through event a represents a particle traveling at constant speed relative to the frame. The particle is sent from the reference event o and passes through several events shown in the figure. The line drawn through the events illustrates the worldline of the particle.

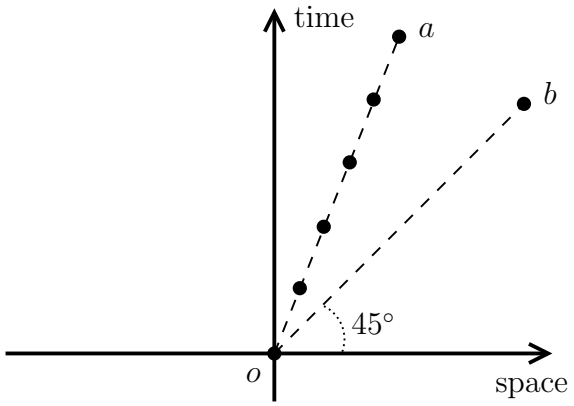


Figure 3.2: Worldlines of a particle (through event a) and light (b). The particle passes through several events, which form a worldline. A light-flash's worldline inclines at the unit slope, i.e., at an angle of 45 degrees to the horizontal axis.

The second worldline in figure 3.2 represents a light flash sent from the reference event o. Because light travels one meter in one meter of time, its worldline has been raised 45 degrees from the horizontal axis. As no object can travel faster than light, the inclined angle to the vertical axis of every worldline is less than 45 degrees. A slowly moving object draws an almost vertical worldline, whereas worldlines of faster objects have a more inclined

angle to the vertical but yet a slope greater that one, i.e., an angle less than 45 degrees to the vertical axis. [77]

If the particle changes its speed (relative to the frame), it draws a worldline that changes its inclination; i.e., the worldline is curved (see figure 3.3). A curved worldline is shown in figure 3.3, where the particle initially slowly

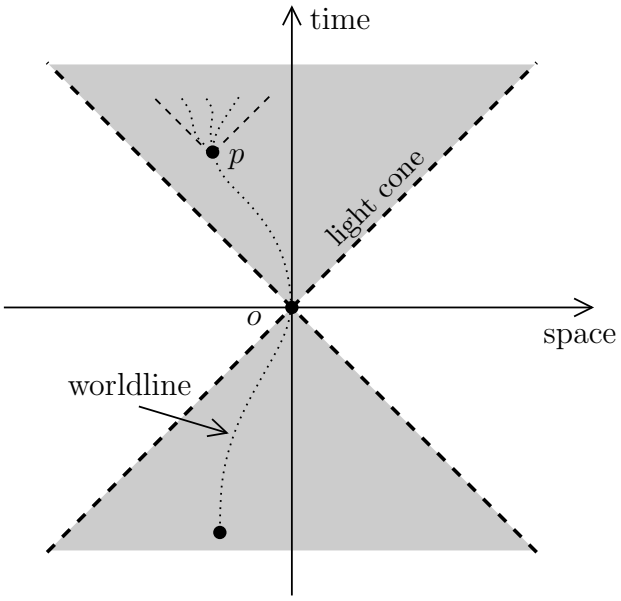


Figure 3.3: Curved worldline of a particle changing its speed. A lightcone limits the worldline's slope at every point. Shown are some acceptable worldlines for the particle after event p.

accelerates to the right but later slows down and comes to rest at event o. Subsequently, the particle accelerates to the left. At every point, the particle's worldline angle to the vertical is limited below 45 degrees. The lines at $+45°$ and $-45°$ angles to the vertical axis are called *lightcones*, because light travels along these lines. In three- and four-dimensional cases, these lines expand into two coaxial cones with their sharp ends, the apexes, facing each other—a shape which is known as the *dual cone*. Lightcones are shown at events p and o in figure 3.3 with some available worldlines after event p.

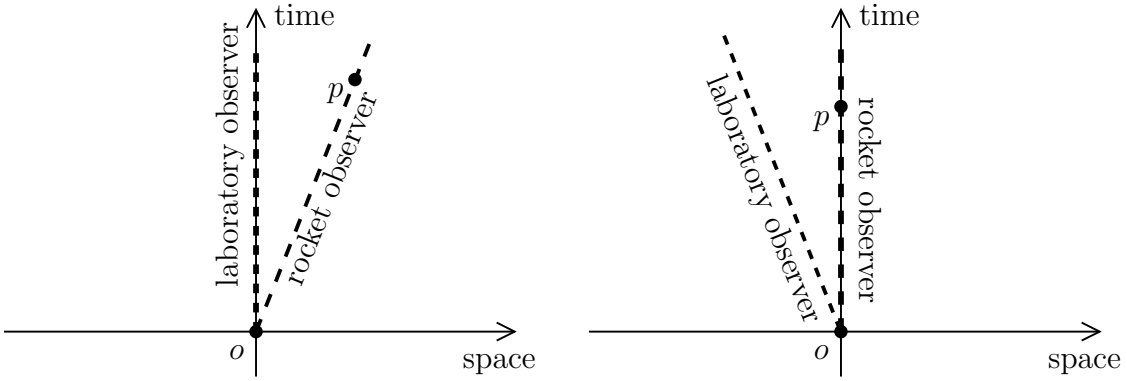
3.1.4 Spacetime interval

So far we have not found anything bizarre from the spacetime structure. Each structure has been substantially quite standard for any four-dimensional space, except some difference in the terminology. However, there are differences. The notion of distance in Euclidian geometry is incompatible with

spacetime, as we will see, and the concept of the spacetime interval is introduced to replace it.

Let us begin by considering two cartographers, who are making maps of (mainly) the same area. They both have the same unit for south-north and west-east directions but they disagree about the direction of north. We ease the situation and suppose that they both have the same physical position corresponding to the zero coordinates. Clearly, they disagree about the coordinates of most locations, but fortunately they agree the distances between points; i.e., they agree about the square sum of coordinate separations. This finding is one of the foundations of Euclidian geometry. [77]

Similarly, let us study two observers making spacetime maps of their neighborhoods; the first observer stays in the laboratory and the second travels in a rocket with constant velocity relative to the laboratory. In figure 3.4 a) the worldlines of both these observers are represented in the laboratory frame. Because the laboratory observer do not move with respect to the



a) Laboratory frame b) Rocket frame

Figure 3.4: Worldlines of laboratory and rocket observers in a) the laboratory frame and in b) the rocket frame. Two spark plug firing events (o and p) are shown in both frames. In the rocket frame the events happens in the same location, but in laboratory frame they have a nonzero space separation. Additionally, the period of time between the events o and p is shorter in the rocket frame compared to the laboratory frame.

laboratory, its worldline is vertical. In contrast, the rocket moves to the right at constant velocity, and thus its worldline tilts an angle relative to vertical. In figure 3.4 b) the same worldlines are drawn relative to the rocket frame and now rocket observer is immovable. For simplicity, let us assume that both frames have the same reference event. The conclusion is that the pointing directions of observers' time axes differ from each other; i.e., they

do not agree on the direction of the time axis. Later we find out that they disagree also about the direction of the spatial axis. Situation is somewhat identical to the cartographer's case.

The similarity with the cartographer's case extends also to coordinates: in different states of motion—in different frames—two observers disagree about the space and time coordinates of events. Because they also disagree about the squared sum of coordinates, the geometry cannot be Euclidian. To illustrate this phenomenon, let us study two spark plug firing events that the laboratory and rocket observer detect. The first plug firing happens when the rocket passes the laboratory—in the reference event o of both frames (see figure 3.4). The laboratory observer finds that the second spark plug fires ten meters from the first firing, just when the rocket passes that point—at event p in figure 3.4 a). Consequently, the rocket observer perceives that the two plug firings happen in the same location (figure 3.4 b)). Thus the space separation of the events—i.e., their spatial distance—is ten meters for the laboratory observer and zero for the rocket observer. In addition, they do not agree on the time (period) between these events.

Series of experiments and modelization have revealed [77] that observers agree on the *spacetime interval* (or *Lorentz interval*), which is defined as[4]

$$(\text{interval})^2 = (\text{time separation})^2 - (\text{space separation})^2, \tag{3.1}$$

where the space separation is an always positive distance of events, to be precise, $(\text{ space separation })^2 = (\text{ north separation })^2 + (\text{ east separation })^2 + (\text{up separation})^2$.

The *(Lorentz) invariance of the spacetime interval* says that the interval is independent of the state of motion of the observer.[5] Since space is different for different observers as is also time, but since their combined quantity, interval, is invariant, space and time cannot be separated and are fated to be part of a single entity, namely *spacetime*. Accordingly, spacetime is the same for all observers. The inventor of spacetime, Hermann Minkowski, described this union in his famous lecture on 21st September 1908 in Cologne [50] as follows: "*Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.*" Space and time are inseparable but not identical, because the minus sign in the equation of the interval distinguishes them.

[4]The sign of the squared spacetime interval varies from one source to another. This definition is referred to as the timelike interval; the other alternative is the spacelike interval. In the next chapter, we take the spacelike interval as our starting point.

[5]Later, we will find that the invariance of the spacetime interval is a consequence of Einstein's principle of relativity.

There is no way to get rid of the sign.[6]

3.1.5 Invariance of the spacetime interval

How is the invariance of an interval shown in a spacetime diagram? We demonstrate this with an example, where we plot the spacetime diagram of the same events in several frames. At first, we have three observers, i.e., three frames: laboratory, rocket, and super-rocket frames. The super-rocket is one that flies at constant speed in the same direction as the normal rocket but considerably faster. A flash of light is emitted at event o; i.e., at an event when and where the reference clocks of the frames coincide. A portion of the flash is reflected from a mirror traveling with the rocket and returns to the reference clock. In figure 3.5, the path in space of the flash of light is plotted in all three frames. In the rocket frame (figure 3.5 b)) light just goes back and forth to the mirror. The rocket travels so fast that the simple up-down track in its frame appears as a tent profile in the laboratory frame (figure 3.5 a)). In the rocket frame, the emission event o and reception event p take place in the same location, whereas in the laboratory frame they occur far away from each other. It may surprise us that the reception, too, does not occur at the same time in both frames. In fact, if we take for granted that the speed of light (in a vacuum) has the same value for both observers—which, in fact, is one major principle of relativity—this is quite obvious. Because light travels farther when recorded in the laboratory frame than in the rocket frame, the time between events o and p is greater in the former.

The super-rocket flies to the right even faster than the normal rocket, and, accordingly, the reception occurs to the left of the emission point (figure 3.5 c)) [77]. The distance and time separation between emission and reception are even greater than in the rocket and laboratory frames. However, the spacetime interval between the events is the same in all three frames.[7]

Figure 3.6 shows the emission and reception events along with the spacetime translation connecting them in spacetime diagrams given in all three frames. As we can see, the space and time components of the translation

[6] At least without using complex numbers (or other suchlike), whose interpretation is not so clear.

[7] A sketch of proof can be given. Let d be the distance between the reference clock and the mirror when they are at their closest. Equally, let l be the distance between events o and p. d has the same value in every frame, but the value of l varies. The time separation between emission and reflection—by Pythagoras' theorem and the unit speed of light—is $\sqrt{d^2 + (l/2)^2}$, and naturally o and p have twice the time separation, i.e., $2\sqrt{d^2 + (l/2)^2}$. In conclusion, the interval between o and p is (time separation)^2 − (space separation)^2 = $4(d^2 + l^2/4) - l^2 = 4d^2$, which is independent of the observer.

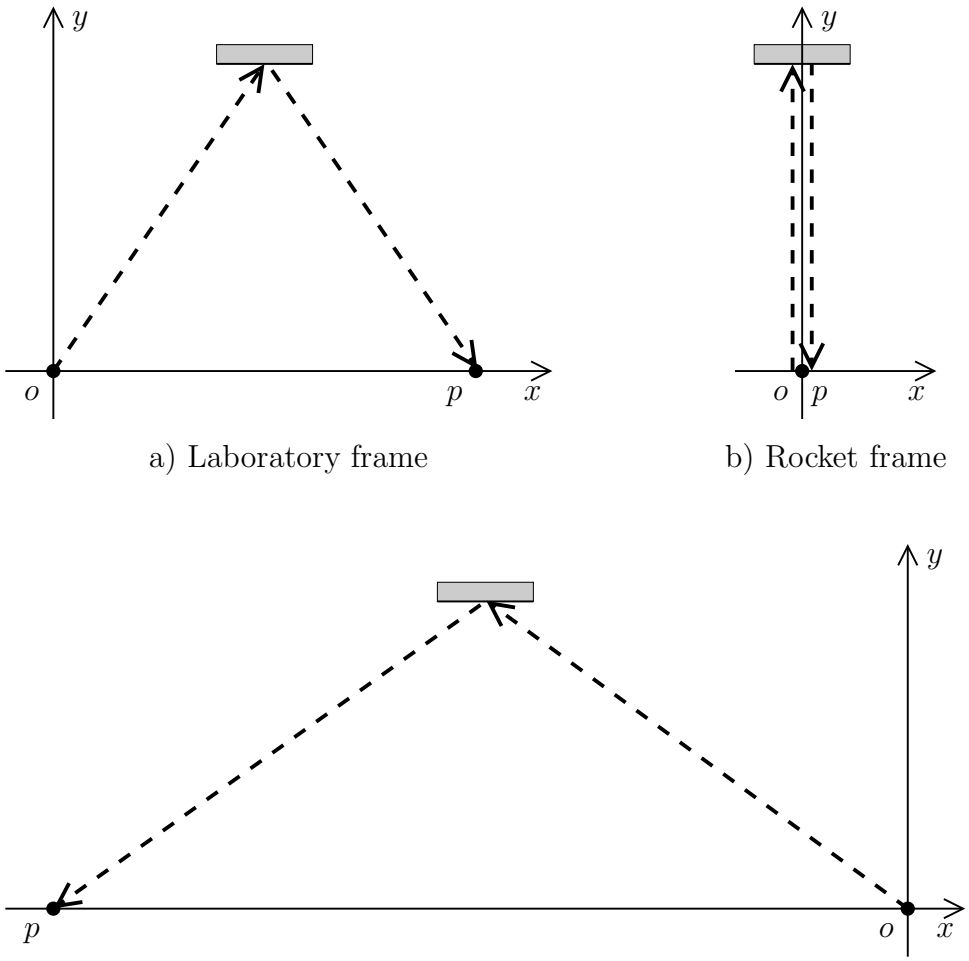
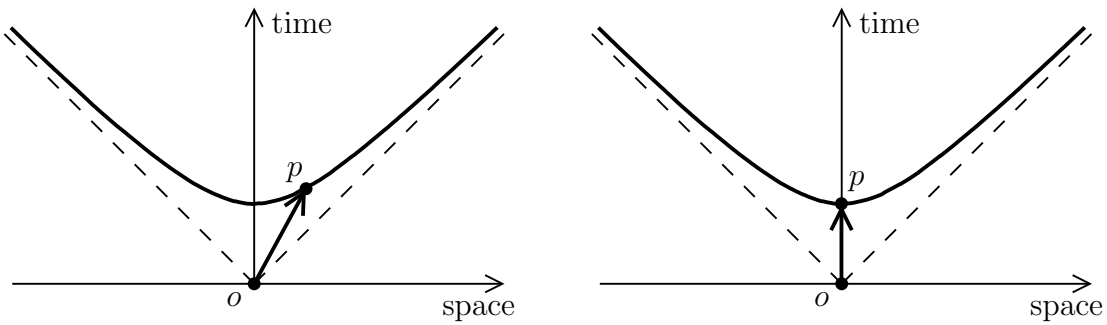


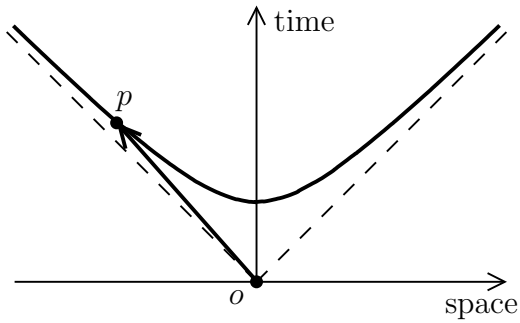
Figure 3.5: Path in space of a flash of light reflected from a mirror, represented in a) the laboratory frame b) the rocket frame c) the super-rocket frame. The rocket and super-rocket fly in the x-direction at constant speed. o and p are the places where the emission and reception events occur, not the events themselves, because this is no spacetime diagram.

differ in different frames—as well as its Euclidian length. Hence the invariance of the spacetime interval means something else than the invariance of distance in Euclidian geometry. The geometry of spacetime is not Euclidian but of another kind, called *Lorentz geometry*.

The arrow of a spacetime translation draws a hyperbola, when we change from one frame to another (see figure 3.6). To be exact, a time and space



a) Laboratory frame b) Rocket frame



c) Super-rocket frame

Figure 3.6: Events of the flash reflection example as shown in the spacetime diagrams of the three frames. The space-direction is the x-direction in figure 3.5. In each frame, the emission and reception events are drawn together with the spacetime translation connecting them. The reception event lies in every frame on a hyperbola, called the invariant hyperbola.

separation between two events satisfies the equation

$$(\text{time separation})^2 - (\text{space separation})^2 = (\text{interval})^2 = \text{constant}, \quad (3.2)$$

because the spacetime interval has the same value in every (inertial) frame. Identity (3.2) is an equation of a hyperbola—called the *invariant hyperbola*. In the Lorentz geometry, the end point of a translation with a constant interval draws a hyperbola, whereas in Euclidian geometry that with a constant length draws a circle.

Each above plot of spacetime translation that connects two fixed events represents the same translation; they are only different projections or perspectives of the same spacetime translation. Every event and interval has its own reality independent of any observer or frame.

We have many everyday interpretations of Euclidian distance, such as the

ruler or odometer, but how can we picture a spacetime interval? If we travel from one event to another at constant velocity carrying a wristwatch, we can measure the interval between events directly with our watch. But it is not enough to go through the points; we must pass through the actual events. From the point of view of the traveler, both events happen in the same place and therefore

$$\begin{aligned}(\text{interval})^2 &= (\text{time separation})^2 - (\text{space separation})^2 \\ &= (\text{time separation})^2 - (0)^2 = (\text{time separation})^2. \quad (3.3)\end{aligned}$$

Consequently, the time read on a wristwatch is the interval between the events. This value of the interval is independent of any frame or observer. This *wristwatch time* is also called the *proper time* or *eigenzeit* in Einstein's German. The clock measuring the proper time is also called the *proper clock*. Proper time can also be interpreted as the *aging* of the observer going through events at constant velocity.

The above interpretation of the spacetime interval as wristwatch time works only if it is possible to travel from one event to another—the interval is then said to be *timelike*. It is possible for events to occur so far separated in space and so close in time that we cannot travel from event to event—because we cannot travel faster than light. Then the interval is called the *spacelike interval*. In case of a spacelike interval, the squared interval (3.1) becomes negative, which is somewhat cumbersome, and thus we introduce a new definition

$$(\text{spacelike interval})^2 = (\text{space separation})^2 - (\text{time separation})^2. \quad (3.4)$$

The timelike interval (3.1) is measured directly using a wristwatch carried in a special frame in which the events occur in the same place, whereas the timelike interval is measured using a rod laid between the events in a frame in which they occur *at the same time*. Later, we shall more precisely describe the concepts spacelike and timelike.

3.2 Free-float frame

So far, we have been careless about the frames used in frame changes. The invariance of entities does not work in all frames but only in some special ones. For instance, spacetime intervals take the same value only in every inertial reference frame, not in every single frame. Now we repair this defect and therefore begin by defining the inertial frame.

Basically, an inertial reference frame is a frame in which the effects of gravity have been removed. We can eliminate these effects by observing

our surroundings in a free fall. If the cable of an elevator is cut and a ball is thrown inside the elevator cabin, the ball flies straight. What could be simpler? On the earth's surface, a ball does not fly a curved path because of the "force of gravity," but because of the force of the floor acting on us. In a free fall, or better put, in a free-float frame, an object remains at rest or moves at constant velocity, if no external forces are acting on it. Moreover, gravity is not an external force in this context, because we have eliminated it by using a free-float frame. In fact, from the point of view of relativity, gravity is never a force.

The experiences of an observer in free-float are similar to those in deep space, i.e., far away from any source of gravity [90].

However, free-fall is not enough to remove all the effects of gravity, because we can find evidence of gravity in its tide-producing action, if we examine a region large enough or for long time enough. Widely separated particles are affected differently by the gravitational interactions of the earth. For example, two particles released side by side are attracted toward the center of the earth and will, accordingly, move closer together while falling (see figure 3.7 a)). In contrast, vertically separated particles move farther apart as they fall (figure 3.7 b)). These forces are called *tidal*, because similar forces from the sun and the moon act on the oceans causing tides.

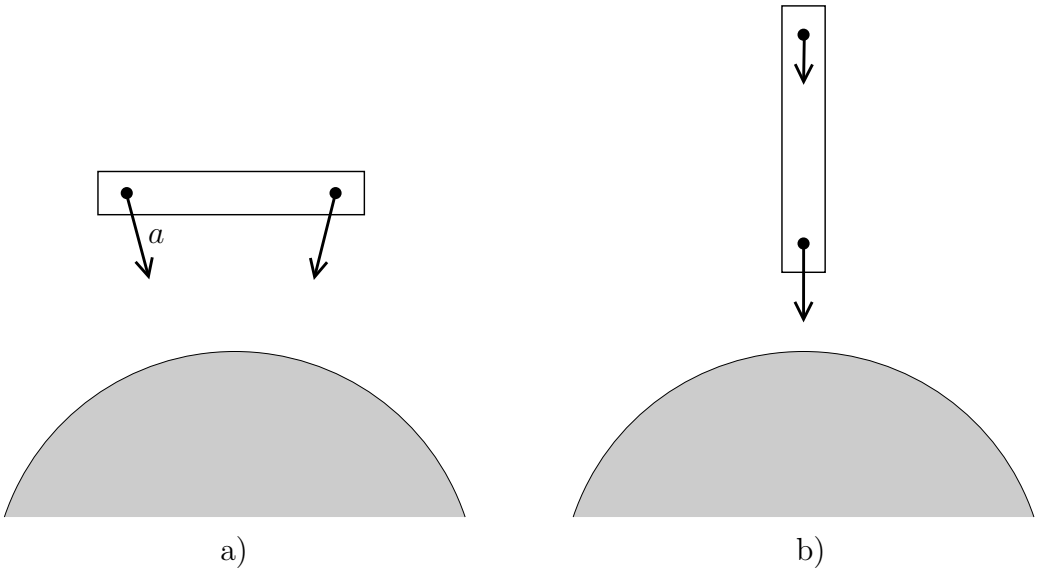


Figure 3.7: Tidal effects of widely separated particles near the earth. A horizontal (a)) and a vertical (b)) object are shown in a free fall, with particles inside having different accelerations.

The detection of tidal effects depends on the sensitivity of the measuring instruments. For a given accuracy, a room can always be made so small that these effects cannot be detected in a given time. This room and this period of time are called the free-float reference frame.

Definition 3.1. A reference frame is said to be *free-float* (or *inertial* or a *Lorentz*) *reference frame* in a certain region of space and time, when every free test particle remains at rest or retains its motion relative to the frame.

This test can be done locally, which makes it simple. Tidal forces make special relativity limited, but far from gravitational bodies, the free-float reference frame can constitute a large region of spacetime. Thus special relativity is a local theory, but the localness depends heavily on the location of the region under consideration.

3.3 Relativity

How can we know whether we are moving or at rest? This is one of the most elementary questions that relativity concerns. Because motion or being at rest are relative matters, we can speak about them only relative to something—relative to the same frame, for example.

Already in 1692, Galileo Galilei described his version of the *principle of relativity*, which stated that there was no absolute and well-defined state of rest as follows. A person in the cabin of a ship gliding at uniform speed cannot detect by an experiment of mechanics how fast she is moving [23]. Einstein generalized the Galilean relativity from mere mechanics to all laws of physics—including electromagnetics:

Postulate 3.1 (Einstein's principle of relativity)**.** *All the laws of physics are the same in every free-float (inertial) reference frame.*

The principle means that all physical laws and numerical values of physical constants are the same in every free-float frame—especially the speed of light. No test of these laws provides any way whatsoever of distinguishing one free-float frame from another. The principle of relativity rests on the emptiness of space; if space is empty, the surroundings are identical from the point of view of every free-float frame, and we cannot differentiate them from each other.

However, the principle of relativity does not say that everything is similar in every inertial frame. We have found that space and time separations differ from frame to frame, and therefore also (spacetime) direction. The velocity of an object relative to a frame is defined as the ratio of space

and time separations of the object's adjacent events; therefore, velocity is frame-dependent. Similarly, acceleration and—as a result—force are frame-dependent, along with electric and magnetic fields.

Einstein's principle of relativity has several important consequences. We have already found two major ones. First, the spacetime interval between two events is invariant. And second, in any free-float frame, no object moves at a speed greater than the speed of light in a vacuum. Other consequences—the relativity of simultaneity, the Lorentz contraction of length, and time dilation—are discussed below. We also introduce the Lorentz transformation, which is a helpful tool to systematize the study of special relativity and, additionally, a tool to illustrate it comprehensibly.

3.3.1 Relativity of simultaneity

One of the rather surprising consequences of the principle of relativity is the *relativity of simultaneity*. It says that two simultanous events as measured by one observer cannot be simultaneous in the perspective of a second observer, who is in (suitable) motion relative to the first observer.

Einstein demonstrated the relativity of simultaneity with his famous train paradox, shown in figure 3.8. Lightning strikes the front and back ends of a fast moving train, and the emitted flash spreads out in all directions from the ends. The train and its travellers move to the right while the flash propagates. The observer riding in the middle of the train concludes that the strikes are not simultaneous, because, first, the flash arrives first from the front of the train and, second, because she is equidistant from the front and back of the train. Furthermore, she knows that the speed of light is the same for every observer and for every direction. Therefore, the flash must have left the front before it did the rear.

A second observer standing by the tracks halfway between the strikes sees both flashes at the same time and concludes that the strikes were simultaneous—by reasoning identical to that of the observer on the train.

How can the strikes be both simultaneous and not simultaneous? This is because the *simultaneity* of events *is relative*, i.e., different in different frames. Simultaneity in one frame does not mean simultaneity in other frames. Only events in the same location or with a separation perpendicular to the observer's relative motion can be simultaneous in both frames. These findings can be summed up as follows [77].

Theorem 3.1 (Relativity of simultaneity)**.** *a) Two events that lie along the direction of relative motion between two frames cannot be simultaneous as measured in both frames. b) Two events with a separation purely transverse*

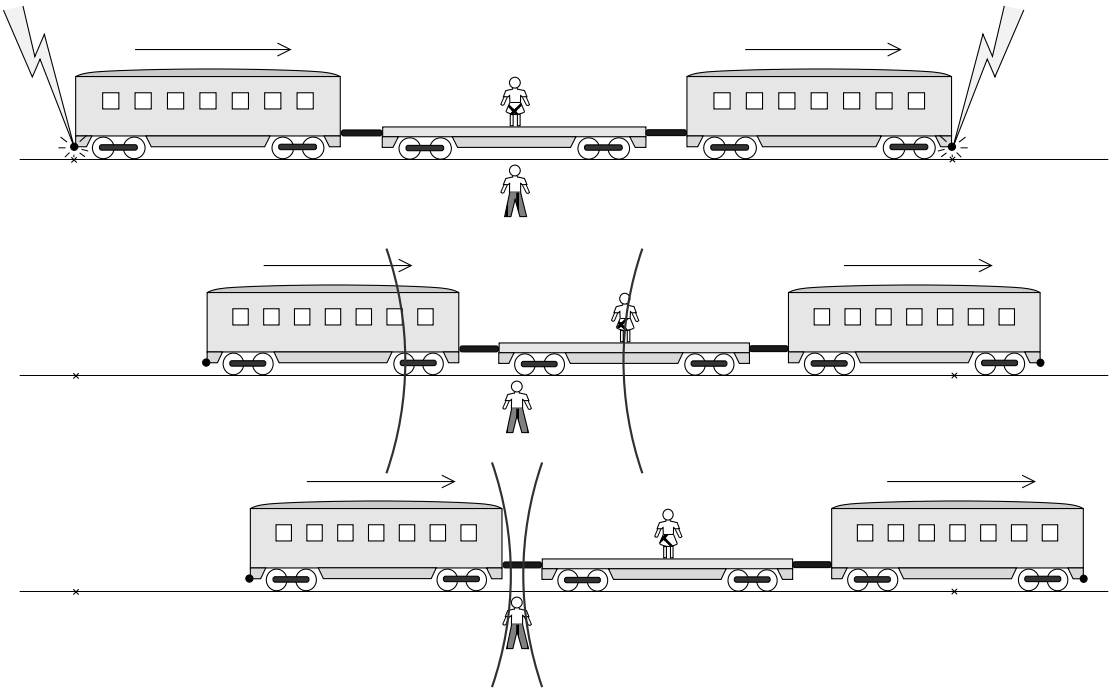


Figure 3.8: Einstein's train paradox. Lightning strikes both ends of a rapid train, and two observers see the flashes. The ground observer sees the flashes at the same time and concludes that the lightnings are simultaneous. The train observer moves towards the flash coming from the front and moves away from the rear flash. Thus she perceives the front flash first and disagrees about the simultaneity of the lightnings with the ground observer. [77]

to the direction of relative motion and simultaneous in either frame are simultaneous in both.

The relativity of simultaneity is the source of many paradoxes that are claimed to be faults of special relativity. In the end, they are all misunderstandings of the relativity of simultaneity—like Einstein's so-called train paradox.

3.3.2 Lorentz contraction of length

Yet another result of the principle of relativity is the Lorentz contraction of length. The length of an object is the space separation between its two ends measured simultaneously. Since observers do not agree about simultaneity, they do not agree about length either.

Let us return to the train paradox. For the observer on the ground, the lightnings happen at the same time, and he says that the length of the train is the distance between those lightning points. On the other hand, from the viewpoint of the observer on the train the measurement at the front happens long before that at the rear. Between these measurements, the train moves forward, and, accordingly, the measurement is too short—from the train observer's perspective. So her measurement is longer.

In fact, the observer at rest to the object measures the longest result, the *rest length* or the *proper length*. This effect is called the *Lorentz contraction of length*: an object in high-speed motion is measured to be shorter along its direction of motion than its proper length; i.e., the length measured in its rest frame. However, in the direction transverse to its motion, the dimensions of a moving object are measured to remain the same, whatever their speed relative to the observer. [77]

3.3.3 Lorentz transformation

Events, and the intervals between the events, define the layout of the physical world. We need only pairwise spacetime intervals between each pair of events to know all the locations of the events—with the aid of the laws of Lorentz geometry. We can analyze the physical world—do science—without a reference frame. From pairwise interval data, we can construct space and time locations of events as observed by an arbitrary observer. [77]

There is also another way to express the same information: the coordinates of a free-float frame. The events are now individual, and there is no apparent connection of events by intervals. The coordinate-based method has the advantage that doing spacetime physics can be mechanized. But it is useful only if we can translate coordinates from one frame to another overlapping frame. These mappings are called Lorentz transformations. In fact, they are not needed at all, if we deal only with intervals.

We omit the derivation of the Lorentz transformation. It is founded on the invariance of the spacetime interval and on the linearity of the transformation. We examine two observers with parallel coordinate axes, namely, a laboratory observer (with unprimed coordinates) and a rocket observer (with primed coordinates). The latter moves in the positive x direction at velocity v_{rel} relative to the laboratory. The two observers create two overlapping frames and the following *Lorentz transformation* can be used to transform

the coordinates of one frame to another:

$$t' = -v_{\text{rel}}\gamma\, x + \gamma t, \tag{3.5}$$

$$x' = \gamma x - v_{\text{rel}}\gamma\, t, \tag{3.6}$$

$$y' = y, \quad \text{and} \quad z' = z, \tag{3.7}$$

where $\gamma = \frac{1}{\sqrt{1-v_{\text{rel}}^2}}$ is the *Lorentz factor* or *stretch factor*.

We can demonstrate the relativity of simultaneity with the aid of the above equations and a spacetime diagram. The spacetime diagram represents the space and time coordinates x and t of events in a frame, e.g., the laboratory frame. Events in the horizontal line—e.g., a line with $t = 0$—are simultaneous in that frame. But which events in that diagram are simultaneous in the rocket frame? The equation for the line of events that are simultaneous with the reference event can be found from equation (3.5) by setting $t' = 0$. We get $t - v_{\text{rel}}x = 0$, and this line is shown in figure 3.9 a). The line also represents the space axis of the rocket frame drawn in the laboratory spacetime diagram. Similarly, the time axis is found by setting $x' = 0$ in (3.6). Both axes are shown in figure 3.9 b). The figure shows how the Lorentz transformation turns the axes of a spacetime diagram. Thus, this figure can be considered as a graphical presentation of the Lorentz transformation. Note that both axes of the moving frame tilt towards the lightcone, and that both have an equal tilting angle, a characteristic of all Lorentz transformations.

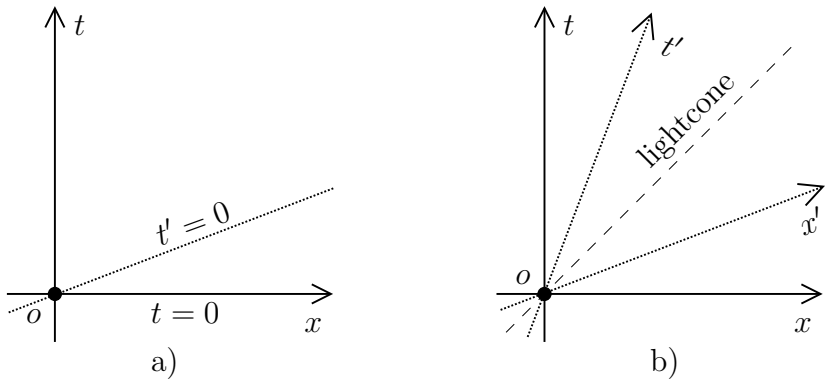


Figure 3.9: a) Spacetime diagram of the laboratory frame with the line of simultaneity of the rocket frame. b) Axes of the rocket coordinates drawn into the spacetime diagram of the laboratory frame. Both axes have the same angle with respect to the lightcone.

The Lorentz transformation also predicts how velocities are added or, rather, combined. With two frames in relative motion having a relative

velocity v_{rel}, the first observer measures the object's velocity to be v' relative to his frame. The *law of the combination of velocities* says that the second observer perceives the velocity

$$v = \frac{v' + v_{\text{rel}}}{1 + v'v_{\text{rel}}}. \tag{3.8}$$

Note that this is no simple summation, as we might expect, and, that, further, v cannot exceed the speed of light in a vacuum in any circumstances. Moreover, both observers detect the same speed of light, e.g., the unit speed.

The theory of hyperbolic geometry [58, 65] shows that in a two-dimensional case the place and time coordinates (x and t) of an event p in a frame can be expressed with a hyperbolic angle r called the *rapidity*. If s is the spacetime interval between p and the reference event o (at the origin of the coordinate system), the coordinates of p are $x = s \sinh r$ and $t = s \cosh r$. In addition, if o and p are nearby events in the object's worldline, $\tanh r = v_{\text{rel}}$, the object's velocity relative to this frame. Also if a second frame is moving along the worldline, the Lorentz factor can be represented as $\gamma = \cosh r$. The main benefit of the rapidity is that, unlike velocities, rapidities are additive [65].

In the language of Lorentz transformations, we have defined a scalar-valued entity to be (Lorentz) invariant, if the entity keeps it value in every Lorentz transformation. This property keeps a scalar quantity independent of any special choice of inertial reference frame. Is there a similar property, e.g., for a vector-valued quantity? It is irrational to say that a vector quantity should have the same value in every frame. Nevertheless, the components of a vector should follow some predetermined transformation. A vector and a vector-valued function in spacetime is said to be *Lorentz covariant*, if it transforms like a translation between events in an arbitrary Lorentz transformation. Lorentz covariant vectors have a special name, the *four-vector* (or *4-vector*). Thus, e.g., a translation between events is a four-vector.

3.3.4 Time dilation

In the context of the spacetime interval, we found that observers in relative motion measure a different frame time between the same events. Figure 3.6 of the flash reflection example showed that the coordinate time is the smallest in a frame where events happen in the same place; this time was called the proper time. Thus each observer always perceives that any other observer measures less time than herself, because the other measures the smallest possible time difference between events near her, i.e., the proper time. This phenomenon whereby an observer finds that another's clock is ticking at a

slower rate is called *time dilation*. A classical example of time dilation, the so-called twin paradox, is given in the next section.

If we set $x = 0$ in (3.5), t is the proper time and t' the frame time. Therefore, the ratio of the frame time and proper time equals the Lorentz factor γ. In everyday life, speeds are such small fractions of the speed of light that the Lorentz factor nearly equals unity, and thus time dilation or length contraction cannot be detected.

3.4 Proper time of worldline

We have discovered that the progression of time depends on the velocity at which we are proceeding. But a more practical question is, how much do we age in our curved worldline through spacetime, i.e., what is the proper time of our worldline? Measuring aging is analogous to measuring the length of a curved path in Euclidian geometry. Length can be measured, e.g., with an odometer or by dividing the path into short straight segments and summing up their lengths.

The proper time of a worldline can be found in the same way. The most straightforward manner is to travel through the worldline with a wristwatch. The time lapsed according to the wristwatch is the proper time.

We can also divide the worldline into small straight pieces and sum up the spacetime intervals of the pieces. Straight in this case means that velocity along that piece of the worldline remains constant. Segments must be so small that velocity is—within the accuracy—constant between its endpoints, and that the particle acts like a free-float particle along this piece of worldline (see figure 3.10 a)). Because all observers measure the same proper time for individual segments, they also agree on the total proper time. Thus the proper time of a curved worldline has the same value in every inertial reference frame.

The proper time—or cumulative interval—correctly predicts the aging of the traveler traversing the worldline. The shortest path between points in Euclidian space is a straight line, and the other paths are longer. In contrast, in spacetime with Lorentz geometry, the straight worldline between fixed events has the longest proper time and—unexpectedly—the curved worldline is shorter. This is caused by the minus sign in the equation of interval (3.1) (also see figure 3.10 a)). The *principle of maximum aging* says that, between two specified events, aging attains its maximum uniquely along the straight worldline, and, further, that a free particle follows this worldline of maximal aging. This is valid only in limited regions of spacetime described by special relativity, but as we will see, it can be modified to be sound also in the

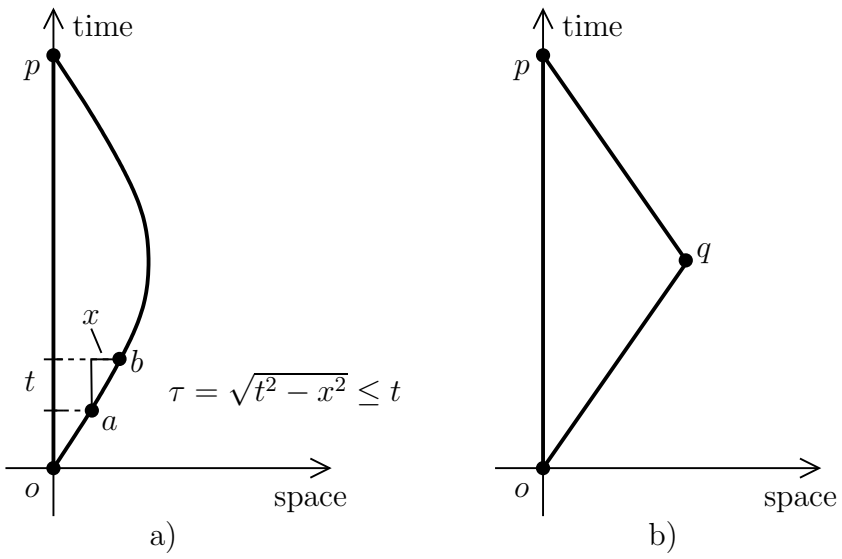


Figure 3.10: Proper time of worldlines. a) A worldline between nearby events a and b is approximately straight and behaves like a free-float frame. The proper time between events (τ) is less than the increase of time (t) between them. Consequently, a curved worldline from o to p takes less proper time than a direct one. b) A kinked worldline is a limiting case of greater and greater acceleration near the turning event. At event q, the worldline changes its direction sharply.

presence of gravity.

The worldline of light has zero proper time. The same worldline is unattainable for any material particle, but the speed of light (in a vacuum) can be approached arbitrarily closely. Let us demonstrate this with the back-and-forth travel shown in figure 3.10 a). As a limiting case when turnaround deacceleration and acceleration happen in an increasingly shorter and shorter period, the worldline turns into a kinked worldline, as in figure 3.10 b). If the traveling speed is sufficiently close to the speed of light, the lapse of proper time along back-and-forth journeys with a kinked worldline can be made as short as we want. This can be noticed from the Lorentz factor, which increases without limit when the speed of light is approached. The Lorentz factor expresses the ratio of the laboratory time and traveler's proper time.

The above characteristic of spacetime is often demonstrated with the so-called *twin paradox*. One twin is sent on a space expedition to a nearby star on a fast spaceship while the other twin remains on the earth. The spaceship's speed is almost that of light, and its journey can be simplified to resemble the kinked worldline in figure 3.10 b). When the traveling twin

returns to the earth, he realizes that he has aged less than his stationary brother. Hence, the fast traveling twin ages less. This violates no law but is only a result of Lorentzian geometry, where the traveler's elapsed proper time is his private matter and depends on the spacetime path taken.

The actual paradox has to do with the fact that the findings of earth and spaceship observers differ. It seems possible to reverse the situation and think that the spaceship remains stationary and the earth is moving—from the standpoint of the spaceship observer. Then the spaceship's worldline would be straight and the earth's kinked. In this case, how can we tell which worldline takes more proper time? We can distinguish the wordlines due to the sharp turnaround in the spaceship's worldline; this fold prevents us from regarding the spaceship's frame as a single free-float frame. The ship feels inertial forces when turning while the earth has no such turnaround.

The twin paradox is not a paradox but a misunderstanding of the relativity of simultaneity. Because the state of motion on the outward and inward trips is different, we must consider them in separate frames. On the outward trip, the earth observer detects that the time progression of the spaceship observer is slower than his, whereas the spaceship traveler observes the opposite. We end up having a similar outcome also on the inward trip. Now where is the catch when each observer thinks that the other measures less time than he? It is in that the outward and inward observer have a different line of simultaneity, as seen in figure 3.11. There is a vast difference between the earth time of the earth bound events a and b, which are simultaneous with the turnaround event q in the outward and inward frames, respectively. During the U-turn, the place of simultaneity sweeps a large segment of the life line of the stationary twin. That jump in the simultaneity of the traveler explains the supposed paradox. For a full explanation of frame change we need general relativity or a series of frames at different speeds [77].

On a high-speed roundtrip, we can travel to the earth's future, but we cannot go back in time. Thus time travel like this is only a one-way trip.

The twin paradox—or rather the *twin effect*—can be demonstrated with high-speed radioactive particles. Measured with the laboratory clock, the lifetime of a high-speed particle is significantly longer than the lifetime of a particle at rest.

3.5 Causality

What would happen if a signal could travel faster than light in a vacuum? A great confusion, because that would destroy the normal order of cause and effect. If an event b is (or could be) the direct consequence of an event a, we

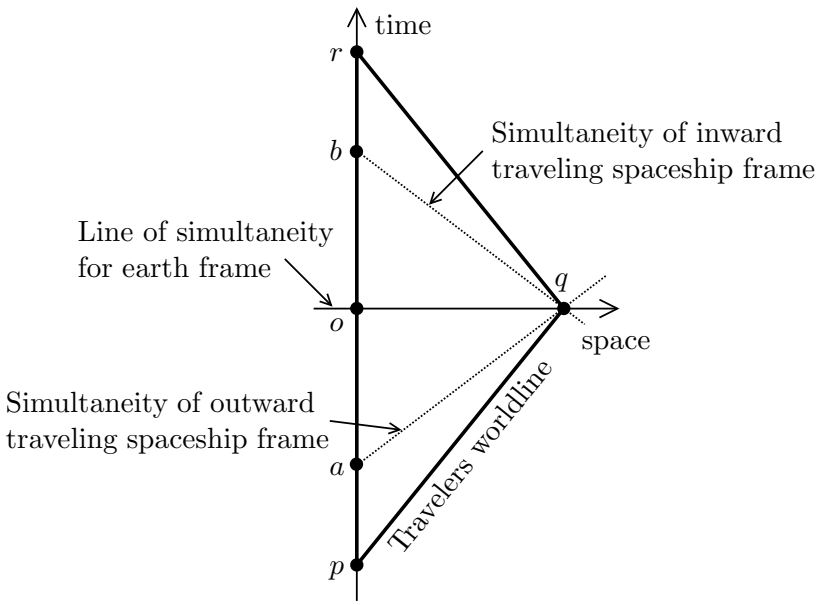


Figure 3.11: Worldline of the spaceship twin travelling to a nearby star and back. The line of simultaneity of three frames: the earth frame, the outward traveling spaceship frame, and the inward traveling spaceship frame.

say that a is the *cause* and b the *effect*. If a signal's traveling speed could exceed the speed of light in a vacuum, observers would disagree as to which of the two events is the cause and which the effect. This would destroy the foundations of present-day science.

Thus the speed of light (in a vacuum) sets the limit on causality. An event cannot cause another when their spatial separation is greater than light can travel in the time between them. The spacetime interval can be used to quantify this limit on causality.

The interval between far-away events (far in space and time)—unlike Euclidian distance—can be zero. In fact, the squared spacetime interval can be positive, zero, or negative, and the sign can be used to characterize the interval. Because the spacetime interval has the same value in every overlapping free-float frame, the sign is independent of the choice of inertial frame. The causal relation between events is arranged in three classes according to the sign of the interval:

timelike	if the time part of the interval dominates the space part
spacelike	if the space part of the interval dominates the time part
lightlike	if the time part of the interval equals the space part

More specific explanations are given below.

3.5.1 Timelike interval

When the time separation of two events is greater than their space separation, i.e., when the spacetime interval is positive, the events—and the translation between them—is said to be *timelike*. A material particle follows a *timelike worldline*; i.e., all the events of the worldline have a timelike relation to each other. A timelike interval τ has the well-known form

$$(\text{timelike interval})^2 = \tau^2 = (\text{time separation})^2 - (\text{space separation})^2. \quad (3.9)$$

As mentioned before, if we consider two fixed events with one placed at the common origin of several spacetime diagrams, in transition from one frame to another the second event draws an invariant hyperbola that opens up along the time axis, as in figure 3.6. The time separation is the smallest in the frame where the events occur in the same place. It is a frame-dependent issue whether the event is located to the left or the right, as seen in figure 3.6.

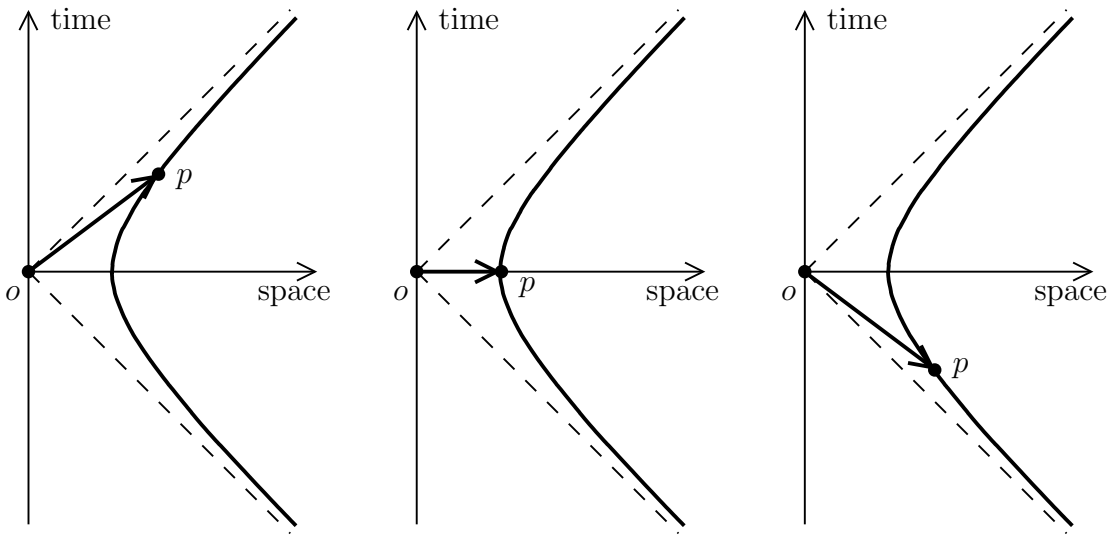
3.5.2 Spacelike interval

When the square of a timelike interval (3.9) becomes negative, i.e., when the space separation is greater than the time separation, events and the separation translation are *spacelike*. We define

$$(\text{spacelike interval})^2 = s^2 = (\text{space separation})^2 - (\text{time separation})^2, \quad (3.10)$$

and its value is invariant. If we draw the same pair of spacelike events in several spacetime diagrams like above, the second event creates a hyperbola opening in the space direction, as in figure 3.12. The space separation of events is the smallest in the frame where the events occur at the same time, in which case the separation is called the *proper distance*.[8] Before and after

[8]Do not confuse this with the proper length, which is the measurement of an object's length in the rest frame, i.e., the maximum space separation of two unfixed events. On the other hand, the proper distance is the miminum space separation of two fixed events. However, both of these extremals are attained when the events are simultaneous, and in both cases the value equals the value of the interval between the events under discussion.



a) Left moving rocket frame b) Laboratory frame c) Right moving rocket frame

Figure 3.12: Two spacelike events o and p in the spacetime diagram of three free-float frames, o being the reference event in each frame. Rockets are moving at equal speed in opposite directions relative to the laboratory. In the first frame, event a happens before b, but in the last frame their order is reversed. In the laboratory frame, events are simultaneous. Consequently, before and after are frame-dependent entities.

are frame-dependent entities, as demonstrated in figure 3.12. Spacelike events cannot be causally connected—i.e., they cannot get information from each other—because otherwise cause and effect would be scrambled. Accordingly, no material particle or signal can move between two events connected by a spacelike translation.

3.5.3 Lightlike interval

Two events are in a *lightlike* relation, when the interval between them is zero, i.e.,

$$(\text{space separation})^2 - (\text{time separation})^2 = 0. \tag{3.11}$$

Thus the magnitudes of the time and space separations of two lightlike events are equal. A lightlike (or *zero* or *null*) translation is plotted along ± 45 degree lines in the spacetime map. The influence of one event—spreading out at the speed of light—can directly affect a second event, if the events are in a lightlike relation. Only light—i.e., photons—and possibly gravitons can

move directly between events connected by a lightlike translation.

3.6 Lightcone

Events that can be reached by a pulse of light from o constitute the *future lightcone* of the event o. They satisfy

$$(\text{future time relative to } o) = +(\text{distance in space from } o). \tag{3.12}$$

A flash emitted from o expands as a sphere—or a circle in two-dimensional space—in space and traces out a cone opening into the future time direction in spacetime. The future lightcone is shown as an upward opening cone in a three-dimensional spacetime diagram in figure 3.13. Here the third space dimension is suppressed, i.e., two spatial dimensions and one temporal dimension are present.

Events that can send a light pulse to o make up the *past lightcone* of the event o. For them

$$(\text{past time relative to } o) = -(\text{distance in space from } o). \tag{3.13}$$

The cone opening downward in the diagram 3.13 traces the history of an incoming circular pulse perfectly focused to converge on the event o. Together, future and past lightcones constitute a *lightcone*. The intersection of a lightcone and a plane with constant time constitute a circle in the three-dimensional case (as in diagram 3.13) or a sphere in the general case.

The lightcone of o classifies every event, everywhere in spacetime, into one or another of five distinct categories according to the causal relation it has with the chosen event o (see figure 3.13) [77]:

1. A material particle emitted from o can affect what is going on to happen in an event inside the future lightcone of o—in the *future timecone* of o—(e.g., event a in figure 3.13)

2. On the future lightcone of o, a light ray emitted at o can affect—with no time to spare—what will happen (e.g., event b in figure 3.13)

3. A neutral or unreachable region: no effect whatsoever produced on o can affect what happens in the *spacecone* of o (e.g., event d in figure 3.13)

4. A material particle emitted inside the past lightcone of o—in the *past timecone* of o—can affect what is happening in o (e.g., event e in figure 3.13)

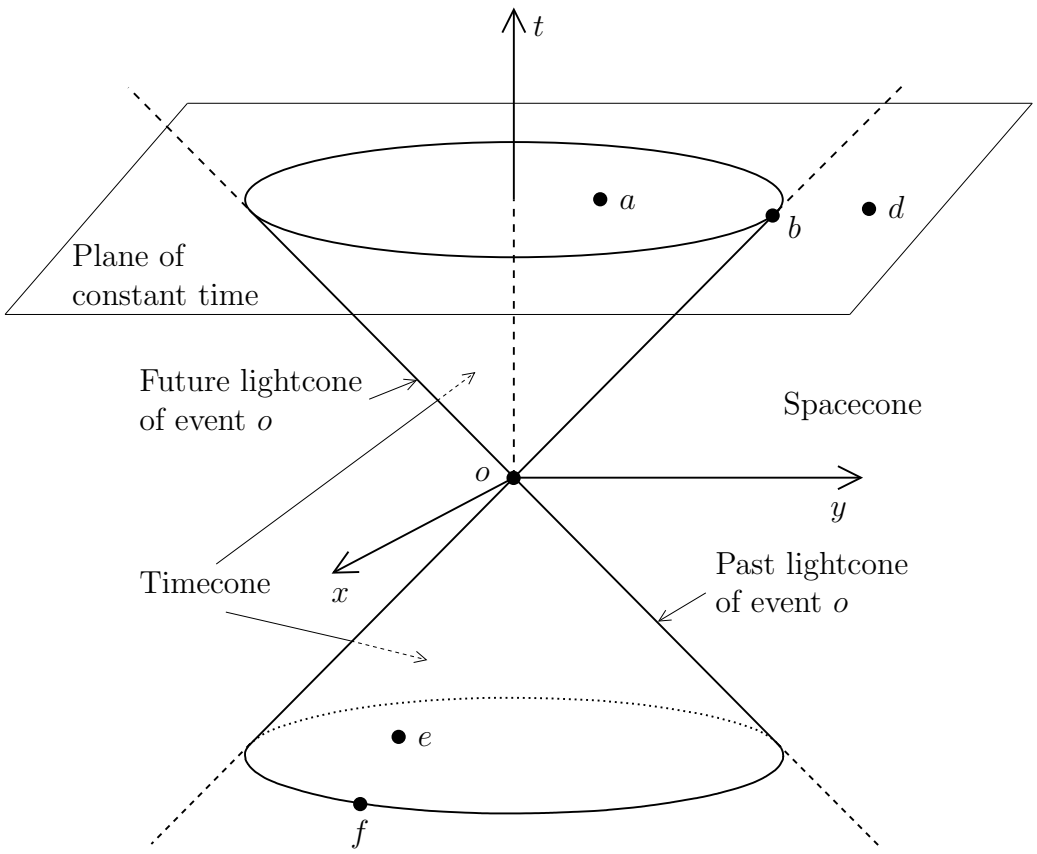


Figure 3.13: Lightcone shown in a three-dimensional spacetime map; the third space dimension is suppressed, and the two remaining space dimensions and a single time dimension are projected on paper in the familiar way of basic geometry. Shown are the five categories of spacetime events.

5. A light ray emitted from the past lightcone of o can affect—with no time to spare—what is happening in o (e.g., event f in figure 3.13)

Future and past timecones together form the *timecone* of o; i.e., it is composed of the events inside the lightcone. To conclude, Lorentz geometry retains by its nature the cause-and-effect structure of events. Lightcone and the above categories are identical in every overlapping free-float frame; i.e., they are independent of the state of motion of the inertial observer.

3.7 Momentum-energy

The energy and momentum of a system depend on the observer, but their combined quantity, called *momentum-energy*, is independent, i.e., a Lorentz covariant quantity. Momentum-energy is a four-vector—therefore, it is also called *four-momentum* or *4-momentum*—with momentum as its spatial part and total energy as its time component. Total energy subsumes all forms of energy, e.g., heat energy, chemical energy, potential energy, and electromagnetic energy.

In isolated mechanical and electromagnetic systems, both total energy and momentum are conserved, and thus also the momentum-energy 4-vector.

The magnitude of the 4-vector is calculated with the same signature as the spacetime interval, which, by the way, is the magnitude of the separation 4-translation (or 4-vector) from one event to another. The magnitude of momentum-energy is then

$$(\text{magnitude of momentum-energy vector})^2 = E^2 - p^2, \tag{3.14}$$

where E is the time component of momentum-energy—i.e., energy—and p is the magnitude of the vector of spatial components—i.e., the magnitude of momentum. As a magnitude of a four-vector, the magnitude of momentum-energy possess the same value in every inertial reference frame.

In particle mechanics, momentum-energy points along the particle's world-line; i.e., it is parallel to the particle's displacement vector. Thus it is always timelike. In fact, in this case, momentum-energy is the displacement vector of the particle divided by the proper time for that displacement and multiplied by the mass of the particle. Thus its magnitude is the mass (m) of the particle. Thus generally $E^2 - p^2 = m^2$.[9]

The above particle's rate of change of the displacement vector against the proper time is a four-vector called the *four-velocity* of the particle. Thus, familiarly, momentum-energy is the four-velocity times the particle's mass. Moreover, the rate of change of momentum-energy with respect to the proper time of the particle is called the *four-force* of the particle.

3.8 Gravity—curved spacetime

We have so far gained a good enough grasp of special relativity for our purposes. Now we can return to gravitational effects. How does gravity affect

[9] In the rest frame of particle $p = 0$, and thus $E = m$. In conventional units, $E_{\text{conv,rest}} = mc^2$, the most famous equation in physics. This manifests the equivalence of matter and energy.

the physics of spacetime? As mentioned before, locally in an inertial frame, physics is identical to the physics of gravitation-free space. But in a global perspective things are different. Gravity does not change the laws of physics, but it affects the spacetime structure itself, making it *curve*. From the point of view of mechanics, the theory of gravity can be summarized by quoting John Wheeler: "*Spacetime grips mass, telling it how to move; and mass grips spacetime, telling it how to curve*".

Einstein's gravity differs considerably from Newton's gravity. Newton described everything with respect to one rigid global frame, in which gravity was an external force from a distance. In contrast, Einstein tells us that locally physics is gravitation-free; i.e., locally, the worldline of a free particle is always straight relative to a free-float frame. But this works only in a limited region of spacetime, and thus we need a variety of local free-float frames. The concept of gravitational force acting on a body is replaced by spacetime geometry; the action of gravity is represented by the object's curvy—yet locally inertial—motion within a curved geometry of spacetime. [77]

Spacetime curvature is shown as a change in the separation of two originally parallel worldlines. For instance, consider two balls initially at rest relative to the earth then beginning to fall towards the earth's center. As they fall, their horizontal separation diminishes and their vertical separation widens due to gravitational interactions (as in figure 3.7), and their worldlines are no longer parallel. This *tide-producing action* results from spacetime curvature in the same way as the curvature of the earth causes the paths of two north-travelling mates to cross at the north pole, even if on the equator they move parallel towards the north.

No evidence of gravitation can be detected by following the motion of a single particle in a free-float frame. Gravity can locally be detected by its tide-producing action; i.e., it produces an acceleration of nearby particles relative to each other. By the very definition of the size of a free-float reference frame described in section 3.2, tidal effects cannot be detected in a single free-float reference frame. But by increasing the sensitivity of measuring instruments or the distance between the particles, tides can be detected. The differences in the directions of particles' worldlines, as observed in two nearby frames, are described in terms of the *curvature of spacetime*. These changes in local measures add up and yield the everyday appearance of long-range gravitation.

However, the situation is not quite as straightforward as described above, because besides mass also momentum-energy makes spacetime curve. For example, gravity waves demonstrate this phenomenon. A stricter version of Wheeler's quotation reads *"Spacetime acts on momentum-energy, telling it how to move; momentum-energy reacts back on spacetime, telling it how*

to curve." The curvature of spacetime is like the surface of a trampoline under our feet: it stretches under the feet, and the effect is transmitted in an ever dilute measure to remoter and remoter regions. The travel of a particle or light in spacetime is then analogous to the route of a ping-pong ball on the surface of a stretched trampoline. The extremal situation of spacetime curvature is the *black hole*, which causes a singularity in spacetime curvature—like an extremal sharp stretching of the trampoline, creating a subsidence deep enough for the ball not to escape.

Chapter 4

Mathematical theory

Our aim here is to study electromagnetic waves by numerical computation in spacetime. To achieve this goal, we must learn the basic concepts of mathematical modeling of both spacetime and electromagnetism.

Because we use semi-Riemannian manifolds to model spacetime, we start by discussing manifolds with an indefinite metric and the generalized concept of observer in this manifold. For electromagnetic modeling, we introduce cells, chains, and cochains. In addition, to apply constitutive laws to electromagnetic modeling, we need multivectors, multicovectors, and, last, differential forms. Finally, we need a Hodge operator for the constitutive laws and, additionally, a connection between cochains and differential forms.

We have to develop new mathematics for chains, cochain, and differential forms in semi-Riemannian (or Lorentzian) manifold, because these concepts are normally considered in the framework of standard Riemannian manifold. Even if no fundamentally new mathematical concepts are established, these results are vital for applied mathematics in spacetime computation.

4.1 Manifolds

To study electromagnetism in spacetime, we need a geometrical model for it. In general relativity, the concept of Lorentzian manifold has been established as the standard model of spacetime geometry.

But why do we need manifolds? The main philosophical reason is that using manifolds makes physical laws independent of coordinates, because there are no canonical (global) coordinates in a manifold. The notion of manifold unites the whole structure of a variety of frames, introduced in the last chapter, under a single concept. However, the main technical reason is undoubtedly that the manifold structure allows a curvature of the domain, for

without manifolds we would be constrained to flat affine spaces. For instance, we can model the surface of a sphere intrinsically as a two-dimensional manifold without embedding it in a three-dimensional space. Moreover, there is a strong link between our observations and intuition about the geometry and the manifold structure; the manifolds are locally Euclidian or Minkowski spaces. Another reason is that the mathematical theory of manifolds is so firmly established so as to afford a good basis for our study of spacetime electromagnetism.

We proceed hierarchically and construct a proper manifold for modeling spacetime. At bottom, each manifold is a topological manifold. However, almost without exception also a differentiable structure is included in it. These structures, and also tangent space, are standard in our case. What is not standard is an indefinite metric introduced to the differentiable manifold, which creates a semi-Riemannian manifold. We are particularly interested in the causality, orientation, and time-orientation of the Lorentzian manifold, a special case of the semi-Riemannian manifold. Moreover, important to the theory of relativity are geodesics and so also to us for interpretation. In the end, we pull together a complete model of spacetime.

4.1.1 Topological manifold

The topological manifold is the most fundamental manifold type, because at bottom each manifold is a topological manifold with possible additional structures. For example, a differentiable manifold is a topological manifold with some additional requirements for charts. We return to this later.

A topological manifold is a *Hausdorff space* with some additional structures. Specifically, it is a space equipped with a topology that fulfils the Hausdorff property, i.e., two disjoint points have disjoint neighborhoods [93]. Topology allows us to speak about limits and continuity, and the Hausdorff property makes the limits unique. Thus a manifold has a topological structure even if no metric is defined for it. This is important to us, because eventually we must define an indefinite metric whose connection to topology is not so straightforward.

The main idea of a manifold is that it is locally Euclidian; i.e., every point has a neighborhood that is homeomorphic[1] to an open subset of Euclidian space, namely $\mathbb{R}^n$ equipped with a Euclidian topology.[2] These homeomor-

[1]Homeomorphism is a continuous bijection with a continuous inverse. We can speak here about continuity, because we have a topology in the manifold and in a Euclidian space.

[2]Henceforth, we assume that (when discussing manifolds) $\mathbb{R}^n$ is the vector space of real arrays furnished with a Euclidian topology.

phisms are called *charts*. Locally, a manifold resembles a Euclidian space. Thus we have all the mathematical structures of $\mathbb{R}^n$ locally in the manifold, such as Euclidian topology and linear structure.

But we must note that it is only locally Euclidian, and that as a whole a manifold can be strongly non-Euclidian. For instance, a manifold topology is always locally Euclidian, but in a global perspective, it can be non-Euclidian. Some of these local structures can be extended to global structures, if we make some additional assumptions about the charts, e.g., that a differentiable manifold adds some differentiability requirements to the charts.

Formally, a manifold can be defined as follows [38]:

Definition 4.1. A *topological n-manifold* is a pair (M, A), where

1. M is a second countable[3] topological space with the Hausdorff property
2. A collection of charts A—called *atlas*—satisfying the condition that
   - the charts are homeomorphic maps from open subsets $U \subset M$ to open subsets of $\mathbb{R}^n$ or $\mathbb{R}^n_+$
   - $\bigcup_{\chi \in A} \mathrm{dom}(\chi) = M$,

where $\mathbb{R}^n_+ = \{\boldsymbol{x} \in \mathbb{R}^n | x_1 \geq 0\}$, the *half space* of $\mathbb{R}^n$. We assume that n, the dimension of $\mathbb{R}^n$, is constant for all charts and we call this constant *dimension* of the manifold.

The charts of an atlas give local coordinates to the manifold; hence maps are called *coordinate charts*, and $\mathbb{R}^n$ (or $\mathbb{R}^n_+$) is called the *coordinate space*.

If some of the points have their neighborhood homeomorphic to the open subsets of $\mathbb{R}^n_+$, the manifold is called a *topological manifold with boundary*. The points that are mapped to the hyperplane $x_1 = 0$ under some (and hence every [38]) chart, are called *boundary points*. The set ∂M of boundary points is called a *boundary* and it is a $(n-1)$-dimensional manifold without a boundary.

The atlas that determines a topological manifold is clearly not unique. By taking the union of all possible atlases determining a unique manifold, we get a unique *maximal atlas for a topological manifold* [37]. Thus a topological manifold is uniquely determined by the pair (M, A_M) of a proper topological space M and a maximal atlas A_M. A (topological) manifold is *connected*, provided it cannot be expressed as the union of two disjoint nonempty open sets.

[3]Second countable means that there is a countable basis for the topology.

4.1.2 Differentiable manifold

At bottom, the most primitive electromagnetic theory deals with real valued functions ($f : M \to \mathbb{R}$), their derivatives, and the relations between these objects. Therefore, differentiability of functions is an essential part of the mathematical theory we need here.

Because a manifold is locally homeomorphic to $\mathbb{R}^n$ via a chart, say χ, we can locally speak about differentiability: f differentiable, if $f \circ \chi^{-1} : \mathbb{R}^n \to \mathbb{R}$ is differentiable [91].[4] Unfortunately, this definition depends on the local chart, and differentiation is possible only locally. Nonetheless, we can make differentiation independent of a chart and extend its meaning to the whole manifold, if we make additional assumptions about the charts, i.e., limit our examination to some subset of homeomorphic charts of the maximal atlas.

Let χ_1 and χ_2 be charts of a topological manifold M, whose domains intersect (see figure 4.1). The composite map $\chi_2 \circ \chi_1^{-1} : \chi_1(U_1 \cap U_2) \to$

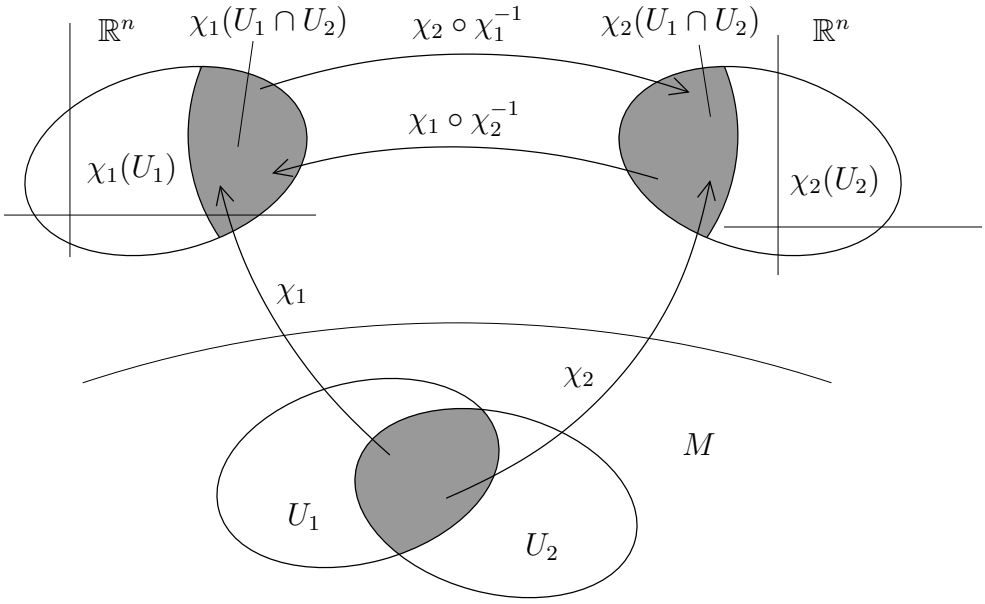


Figure 4.1: Transition maps χ_1 and χ_2 are two charts whose domain has a nonempty intersection $U_1 \cap U_2$. The composition of a chart and the inverse of other chart creates a transition map from one coordinate space to another; e.g., $\chi_2 \circ \chi_1^{-1}$ maps the χ_1-coordinates at $\chi_1(U_1 \cap U_2)$ to the χ_2-coordinates at $\chi_2(U_1 \cap U_2)$.

$\chi_2(U_1 \cap U_2)$ and its inverse $\chi_1 \circ \chi_2^{-1}$ are called *transition maps* or *changes*

[4]The differentiability of functions $\mathbb{R}^n \to \mathbb{R}$ is defined in an elementary way (the same for $\mathbb{R}^n \to \mathbb{R}^n$).

of charts (or *coordinates*), because they transit from one coordinate space to another.

Differentiability can be extended to the whole manifold, if all the transition maps are differentiable; the resulting manifold is called a *differentiable manifold* . Because the inverse of a transition map is also a transition map, both must be also diffeomorphisms. Furthermore, if the transition maps are smooth, we have a *smooth manifold*.

Choosing the atlas describes the differentiability we have in the manifold, but the same differentiable manifold can be described with several different atlases. That is why we are seeking a unique structure to describe a differentiable manifold. If two atlases can be combined into an atlas, whose transition maps are differentiable (to the same degree as the original atlases), the atlases are *compatible*. Compatibility of atlases is an equivalence relation; by combining all the atlases in the equivalence class of one atlas, a unique *maximal atlas* (of a differentiable manifold) can be constructed [38]. This system of equivalence classes or equivalently maximal atlas is called a *differentiable structure*. Consequently, a differentiable manifold can be uniquely described with the pair of a second countable Hausdorff space M and a differentiable structure on M [38].

Any topological manifold of a dimension smaller than four has only one possible differentiable structure.[5] However, if the dimension is greater than four, any *compact* topological manifold has a finite number of differentiable structures. The four-dimensional case is the most difficult, for there are noncompact and even compact four-dimensional manifolds with an infinite number of possible differentiable structures. [3]

The simplest example of a smooth manifold is $\mathbb{R}^n$ (with the Euclidian topology) encompassed with the differentiable structure containing the identity map. In other words, the identity map can be used as a chart to make $\mathbb{R}^n$ a manifold and we need only one chart. The resulting manifold having this differentiable structure is called the *Euclidian n-space*. However, $\mathbb{R}^4$ is a very bothersome example, because innumerable other differentiable structures can be defined for it [3, 76]. However, the above single differentiable structure of Euclidian n-space is enough for us, and henceforth we assume $\mathbb{R}^4$ to be equipped with it.

A general finite-dimensional vector space V (over real numbers) can be uniquely made into a differentiable manifold with coordinate functions in some basis of V. [53]

[5]The differentiable structures are count here modulo homeomorphism, as usual.

4.1.3 Tangent space

We have now defined differentiability, but we do not yet have any idea what kind of mapping the derivative is. In order to define a directional derivative, we must define the notion of direction; i.e., we must introduce an abstraction of vectors for manifolds. This abstraction is also essential in other ways, because it allows us to model displacements, velocities, and other quantities associated with direction. Defining a vector on a manifold is not a trivial task, because, unlike in an affine space, we have no translations from one point to another.

We now aim to define objects called tangent vectors and the linear space[6] of all tangent vectors at one point of a manifold, called the tangent space. Roughly speaking, tangent space models our intuition of linear space—such as plane or hyperplane—approximating a manifold in the neighborhood of a point x. This interpretation is not exact, because it requires some ambient space where the manifold is embedded and where the tangent space lies. Consequently, we must describe tangent space abstractly as an intrinsic structure of the manifold.

Tangent vectors can be defined in various equivalent ways [38], but we define them only via curves, mainly because it is traditionally done so in the field of electromagnetics. We start by defining a parametrized curve in a manifold—e.g., the trajectory of a particle in space as a function of time, or the worldline of a particle in spacetime parametrized by the particle's proper time—and use the travelling direction of the curve at one point as the basis for our definition.

A *finite curve* of a differentiable manifold M is a differentiable function

$$\gamma : (-1, 1) \subset \mathbb{R} \to M. \tag{4.1}$$

Remark 4.1. The differentiability of the finite curve γ is defined via charts, as always: γ is differentiable, if $\chi \circ \gamma : \mathbb{R} \to \mathbb{R}^n$ is differentiable, where χ is a chart in the neighborhood of x. And this is independent of the choice of chart, because the manifold is differentiable.

We are especially interested in *curves going through point* x, i.e., curves satisfying the condition $\gamma(0) = x$. Informally, we use these curves to represent a tangent vector at point x with direction defined by the curve's course and rate of progress. The direction and progression rate of the curve at point x is distinguishable by the vector that it defines in a chart χ: $(\chi(\gamma))'(0)$. But here we must be cautious, because several curves can have the same direction at

[6] In this text, linear space and vector space are synonymous. Notice that in Whitney [91] vector spaces are finite-dimensional linear spaces.

x, and none of them is preferable to the other. Therefore, we define tangent vector as the equivalence class of such curves.

Definition 4.2. Let γ_i and γ_j be two finite curves through $x \in M$, and let χ be a chart in the neighborhood of x. A *tangent vector* at point x is an equivalence class by the following equivalence relation of the curves:

$$\gamma_i \simeq \gamma_j \qquad \text{if} \qquad (\chi \circ \gamma_i)'(0) = (\chi \circ \gamma_j)'(0). \tag{4.2}$$

The equivalence class is uniquely determined by a real array, such as $(\chi \circ \gamma)'(0)$, in a coordinate space.

The definition is independent of the choice of chart, if the manifold is differentiable. By the definition of tangent vector, at each point $\gamma(s)$ of the curve, the curve represents a tangent vector. This vector is called the *velocity vector*, and we denote it $\gamma'(s)$ (notation explained later).

The above definition introduces a bijective map between tangent vectors and vectors in the coordinate space $\mathbb{R}^n$, which clearly depends on the chart used. Thus all information about a tangent vector is included in a triplet (point, chart, vector in $\mathbb{R}^n$). We can manage the dependence of triplets on a chart by describing an equivalence relation of the tangent vector triplets described with two different charts. Triplets $(x_1, \chi_1, \boldsymbol{v}_1)$ and $(x_2, \chi_2, \boldsymbol{v}_2)$ are equivalent if $x_1 = x_2$ and $\boldsymbol{v}_2 = (\chi_2 \circ \chi_1^{-1})'(\boldsymbol{x}_1)\, \boldsymbol{v}_1$, where $\boldsymbol{x}_1 = \chi_1(x_1)$.

The above bijection between tangent vectors and real arrays allows us to transfer the linear structure of $\mathbb{R}^n$ to the linear structure of the space of tangent vectors. In fact, this linear structure is independent of the choice of the chart. We call this linear space of tangent vectors at point x *tangent space* T^xM. If we extend the structure to the tangent spaces of all points in a manifold, we obtain a *tangent bundle*, which is a disjoint union[7] of the tangent spaces of a manifold: $TM = \coprod_{x \in M} T^xM$. An element of TM is the pair (v, x), where $v \in T^xM$ and $x \in M$, and where there is a natural projection $\pi : TM \to M : (v, x) \mapsto x$. As shown in [86], the tangent bundle becomes equipped with a topology and a differentiable structure in a canonical way from those of M, making it into a $2n$-dimensional manifold in its own right.

Basically, a *vector field* v on a manifold is a function that assigns each point $x \in M$ a tangent vector $v_x \in T^xM$. Strictly speaking, a vector field is a section of TM, i.e., a function $v : M \to TM$ such that $\pi \circ v = id$, the identity map. If M is a smooth manifold, and thus TM, we can also speak

[7]The disjoint union $\coprod_{i \in I} A_i$ is a union of pairs (y, i), where $y \in A_i$, $i \in I$, and where $\{A_i : i \in I\}$ is a family of sets indexed by I. The disjoint union gives a separate zero vector for every tangent space; therefore, each $v \in TM$ is contained in a unique T^xM.

about smoothness of vector field, i.e., v is smooth if its composition with the charts is smooth. By the definition of tangent bundle, the value of a vector field v at point x is a pair $(v_x, x) \in T_xM \times M$. We denote v_x the first member of the pair at x.

For subsequent use, we introduce a basis for tangent space. Let χ be a coordinate system of M in the neighborhood U of x. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ be a basis of $\mathbb{R}^n$ such that $[\boldsymbol{e}_i]_j = 0$, if $i \neq j$ and $[\boldsymbol{e}_i]_i = 1$. Each array $\boldsymbol{e}_i$ determines a unique tangent vector $e_i \in T^xM$ and as a whole we have the standard basis $e_1, e_2, \ldots, e_n$ for T^xM associated with chart χ.

In a finite-dimensional vector space V, there is a linear isomorphism between the tangent space T^xV and the vector space itself. For an arbitrary $v, x \in V$ the curve $\alpha(t) = x + tv$ represents the tangent vector at x isomorphic to v.

We are now ready to define the directional derivative of a real valued function and the derivative of a map between differentiable manifolds.

Definition 4.3. Let f be a differentiable real valued function on a manifold, $f : M \to \mathbb{R}$, and let γ be a curve that represents a tangent vector $v \in T^xM$. The *directional derivative* of f at x in the direction of v is the map

$$\mathrm{d}f(x) : T^xM \to \mathbb{R} : \quad v \mapsto (f \circ \gamma)'(0). \tag{4.3}$$

As shown in [38], the value of $(f \circ \gamma)'(0)$ is independent of the choice of γ.

Alternatively, we can define the tangent vector as a directional derivative [53].[8] It is shown in [38] that this definition is equivalent to our definition.

Definition 4.4. Let M and N be differentiable manifolds and $\phi : M \to N$ a differentiable map; i.e., for each chart χ and η of M and N the coordinate expression $\eta \circ \phi \circ \chi^{-1}$ is differentiable. The *derivative* or *differential map* of ϕ is the linear map $\mathrm{d}\phi(x, \cdot) : T^xM \to T^{\phi(x)}N$, satisfying

$$\mathrm{d}\phi(x, v) = (\phi \circ \gamma)'(0), \tag{4.4}$$

where γ is a curve representing a tangent vector $v \in T^xM$.

The differential map, obviously, induces a map $\phi_* : TM \to TN$ between the tangent bundles TM and TN, which is defined by (4.4) in a particular tangent space T^xM. The map ϕ_* is also called the differential or *pushforward*,

[8] The idea is same as in $\mathbb{R}^n$: we can characterize the vector $\boldsymbol{v} \in \mathbb{R}^n$ by the directional derivative operator $\frac{\partial}{\partial \boldsymbol{v}}$. Componentwise, if $\boldsymbol{v} = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 + v_3\boldsymbol{e}_3$, then $\frac{\partial}{\partial \boldsymbol{v}} = v_1\partial_1 + v_2\partial_2 + v_3\partial_3$ and hence there is an isomorphism between these objects.

and it maps the tangent vectors of the manifold M to the tangent vectors of N.

The notation of the velocity vector needs some explanation. Let $\gamma(s)$ be a curve. The derivative of the map $\gamma(s)$ maps the elements of $T^t\mathbb{R}$ to the tangent vectors of $T^{\gamma(t)}M$. Because the elements of the tangent vector space $T^t\mathbb{R}$ are isomorphic to $\mathbb{R}$, as mentioned before, we can identify $s \in T^t\mathbb{R}$ with an element of $\mathbb{R}$ and use the same symbol for both. Thus the derivative $\mathrm{d}\gamma(s) = \gamma'(s)$ maps an $s \in \mathbb{R}$ to the velocity vector of the curve at point $\gamma(s)$—hence the notation. Intuitively, the velocity vector represents the vector rate of change on γ at $s \in (-1, 1)$.

4.1.4 Submanifolds

Roughly speaking, a submanifold P of a differentiable manifold M is a subset of M, which itself has the structure of a differentiable manifold, and the *inclusion map* $j : P \to M : x \mapsto x$ satisfies some properties. Particularly, we require that subset V of P is open if and only if there is an open set V' of M such that $V' \cap P = V$, i.e., P is a *topological subspace* of M.

Definition 4.5. A manifold P is a (embedded) *submanifold* of a differentiable manifold M provided that [53]

1. P is a topological subspace of M.

2. the inclusion map $j : P \to M$ is differentiable, and its derivative $\mathrm{d}j$ is injective in each point $p \in P$.

A typical example of a submanifold is the plane $z = 0$ in $\mathbb{R}^n$. In fact, there is always a coordinate system of M where the $n-p$ coordinates are zero in the submanifold P [38, 53]. Here n is the dimension of M and p is the dimension of P. For example the boundary ∂M is a submanifold and by definition 4.1, the x_1-coordinate is zero on ∂M.

By definition, $\mathrm{d}j : T^xP \to T^xM$ is one-to-one onto its image; thus it is customary to identify T^xP with a vector subspace of T^xM, i.e., to ignore the negligible mapping $\mathrm{d}j$.

4.1.5 Semi-Riemannian manifold

We aim to look for a model for spacetime. All spacetime models derive from the theory of relativity. Relativity gives to spacetime a peculiar geometry, because, among others, both motion and simultaneity are not absolute but relative to an observer, as seen in chapter 3. Later, we will define observer

comprehensively and diversely and call the observer of previous chapter as the *physical observer*. It follows that every physical observer has his own coordinate system, i.e., in the language of manifolds, his own collection of charts.[9] This suggests that the manifold is a proper mathematical structure to model spacetime. However, these charts or coordinate systems are not unique to a physical observer, but the physical observer is associated with a (equivalence) class of charts (or coordinate systems). Later, observer refers to a set of physical observers such that their charts envelope the whole spacetime (or the whole domain).

In general, because gravity causes a curvature in spacetime, the latter cannot be modeled with a flat affine space. This is yet another reason why spacetime should be modeled with a manifold.

Scalar product and metric

All the mathematical structures introduced so far are standard. But the geometry of spacetime differs from the familiar Euclidian or Riemannian geometry in terms of the metric used in the manifold. As seen in the last chapter, we cannot use the standard Riemannian metric to measure the distance between events, because spatial and temporal separations are drastically different in separate observers' coordinates (in different charts) and also their Euclidian square sums differ.

In chapter 3, we also found that the measure of a spacetime interval is independent of the coordinates or charts used; thus we can use it as a basis for the spacetime metric. The spacetime interval is very dissimilar to the Euclidian distance. First, it can reach negative values and give zero values for nonzero separations. Second, it does not fulfil all the requirements for a (Euclidian) norm, e.g., the triangle equation. But even if the structure is quite nonstandard, it indubitably reflects spacetime geometry, because the concept of the lightcone—and hence the fact that the speed of light is equal to every physical observer—is intrinsically included in the metric.

In addition, our spacetime model contains not only a generalization of the metric but also a generalization of the inner product called the scalar product. The scalar product introduces the concept of the *spacetime angle*, which

[9]Physicist prefer the term *frame field* instead of charts. The frame field (or *tetrad*) is a (orthonormal) set of vector fields. Usually, one of them is timelike and three spacelike, and they correspond to the local (coordinate) basis of a physical observer, the timelike vector field always pointing along observer's worldline. In fact, frame field is a more general concept than coordinate basis, because collection of charts can be built to agree with a smooth frame field only if the frame field is *holonomic* [51]. In addition, physicist prefer the frame field, because it is more specifically a geometric objects, i.e., an object independent of the charts used.

in the theory of relativity is called *rapidity*, as mentioned in chapter 3. Physically, the spacetime angle between two worldlines of two physical observers models the relative speed of the observers. The Lorentz transformation from the coordinates of one inertial physical observer to those of another can be described with a spacetime angle.

To model spacetime, we must widen the concept of the metric—and the concept of the inner product as well—in a nonstandard direction; hence we begin by defining the scalar product in a vector space.

Definition 4.6. Let V be a vector space. A bilinear function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is a *scalar product*[10] if it is

1. *symmetric*, i.e., $\langle v, w \rangle = \langle w, v \rangle \quad \forall v, w \in V$
2. *nondegenerate*, i.e., $\langle v, w \rangle = 0 \quad \forall w \in V \;\Rightarrow\; v = 0$.

A vector space with a scalar product is a *scalar product space*. An inner product is a positive definite scalar product, i.e., $\langle v, v \rangle > 0 \; \forall v \in V, v \neq 0$. Accordingly, because a scalar product does not have to be positive definite, it is in general indefinite.

Definition 4.7. The index of a scalar product space V (or the index of the scalar product) is the dimension of the largest subspace $W \subset V$ on which the restriction $\langle \cdot, \cdot \rangle|_{W \times W}$ is negative definite, i.e., $\langle w, w \rangle < 0 \; \forall w \in W, w \neq 0$.

A subspace W of V is said to be *nondegenerate*, if $\langle \cdot, \cdot \rangle|_{W \times W}$ is nondegenerate. Because all subspaces of inner product spaces are nondegenerate, they are inner product spaces in their own right, but the indefinite scalar product space is a more complicated case. The latter always contains degenerate subspaces, e.g., a one-dimensional space spanned by a vector v satisfying $\langle v, v \rangle = 0$, the so-called null vector. Accordingly, a subspace of a scalar product space does not have to be a scalar product space—a characteristic becoming troublesome later.

The tangent space $T^x M$ is a vector space, and, accordingly, a scalar product can be constructed on $T^x M$ for each $x \in M$.

Definition 4.8. A *metric*[11] $\langle \cdot, \cdot \rangle$ on a smooth manifold M is for each $x \in M$ the scalar product $\langle \cdot, \cdot \rangle_x$ of a constant index on $T^x M$, which varies smoothly

[10] The term scalar product is here used to distinguish it from the positive definite inner product, but this terminology is not standard; scalar product is commonly used as a synonym for inner product.

[11] This definition is not the same as the classic definition of the metric in a set, which produces the distances of points in the set. By metric we mean the definition used in differential geometry, often called the *metric tensor*.

from point to point on M. This smooth dependence on x means that $\langle v, w\rangle_\cdot : M \to \mathbb{R} : x \mapsto \langle v_x, w_x\rangle_x$ is a smooth function for arbitrary smooth vector fields $v, w : M \to TM$ [38, 53].

A smooth manifold equipped with a metric is a *semi-Riemannian manifold*[12] M. The constant index of a scalar product ν—i.e., the value of the index that is equal in every tangent space—is the index of M. If $\nu = 0$, M is a *Riemannian manifold*, and then $\langle\cdot,\cdot\rangle_x$ is a (positive definite) inner product on T^xM. If $\nu = 1$ and the dimension of M $n \geq 2$, M is a *Lorentzian manifold* and the metric is a *Lorentzian metric*.

It turns out that the Lorentzian manifold with some extra assumptions is the right choice to model spacetime. Therefore, henceforth, we concentrate on Lorentzian manifolds but speak about semi-Riemannian manifolds, if the definition or theory is also valid for the general semi-Riemannian case. Additionally, for this reason, a point in a Lorentzian manifold is also called an *event*.

The simplest Riemannian manifold is the Euclidian n-space $\mathbb{R}^n$. Because there is a canonical isomorphism between $\mathbb{R}^n$ and $T^x\mathbb{R}^n$, the inner product of $\mathbb{R}^n$ also gives an inner product to $T^x\mathbb{R}^n$, i.e., a metric on $\mathbb{R}^n$.

The simplest example of a semi-Riemannian manifold with a nonzero index is the *semi-Euclidian space* $\mathbb{R}^n_\nu$, which is constructed from $\mathbb{R}^n$ by changing the first ν signs in the scalar product; $\langle v, w\rangle = -\sum_{i=1}^{\nu} v^i w^i + \sum_{i=\nu+1}^{n} v^i w^i$, where v^i and w^i are the components of vectors $v, w \in \mathbb{R}^n$. For $n \geq 2$, $\mathbb{R}^n_1$ is called the *Minkowski n-space*, the simplest example of relativistic spacetime. The Minkowski n-space was the one that we used to model the spacetime of special relativity in the previous chapter.

Remark 4.2. Note that every Lorentzian or Semi-Riemannian manifold is a smooth manifold, and that this requirement is passed on to the spacetime manifold. Smoothness assures that we can speak about derivatives of every degree and, additionally, makes some definitions easier. On most occasions, we could require a weaker condition of differentiability.

Is the metric in a tangent space enough to measure distances—or spacetime intervals—between the points in the manifold? Later we find this to be enough, because we can *integrate* distance by measuring the length of a special "straight" curve, called the geodesic, going between these points. The metric allows us to speak about infinitesimal distances in the neighborhood of one point and, besides, makes it possible to integrate the length of a curve segment. In every local neighborhood of spacetime, the metric has the same

[12]Strictly speaking, a semi-Riemannian manifold is a pair $(M, \langle\cdot,\cdot\rangle)$. Semi-Riemannian manifolds are often called pseudo-Riemannian manifolds.

form as in the Minkowski space, but in a global perspective the metric can differ due to the curvature of the spacetime. Moreover, the local interpretation of the metric also tells us that a Lorentzian manifold resembles locally a Minkowski n-space, which is no surprise after chapter 3.

Alternatively, we can reach the above conclusion on local geometry also as follows. Because the index of the scalar product has the value one in every tangent space T^xM of a Lorentz manifold, the geometry of the tangent space equates with Minkowskian geometry. With the aid of geodesics, we can construct a local diffeomorphism between T^xM and M, called the *exponential map*, in the neighborhood of an arbitrary point $x \in M$ [53]. Via the exponential map we can bring Minkowskian geometry locally to a Lorentzian manifold.

Let χ be a coordinate chart in the neighborhood U of x and $e_1, e_2, \ldots, e_n$ the standard basis (cf. section 4.1.3) on T^xM associated with that chart. With these coordinates, $v \in T^xM$ is given $v = \sum_{i=1}^{n} v_i e_i$ and, similarly, $w \in T^xM$. In these coordinates

$$\langle v, w \rangle = \sum_{i,j=1}^{n} \langle e_i, e_j \rangle v_i w_j, \tag{4.5}$$

where $g_{ij} = \langle e_i, e_j \rangle$ are components of the *metric tensor* $g(v, w) = \langle v, w \rangle$. These components obviously depend on the chart used. The components of the metric tensor form a matrix g_{ij}, which must be symmetric and nonsingular, because the scalar product is symmetric and nondegenerate. Altogether, the metric tensor is a smooth tensor field on M. In a Riemannian manifold, the metric tensor (and the matrix representing it) must be positive definite, but in a semi-Euclidian case it can be indefinite.

Geometry of the semi-Riemannian manifold

Many properties of inner product spaces can be carried over to scalar product spaces and the corresponding properties of geometry on semi-Riemannian manifolds are familiar from Riemannian manifolds. However, some distinctive new phenomena arise when the metric is indefinite.

A vector $v \in T^xM$ is *null*, if $\langle v, v \rangle = 0$ and $v \neq 0$. Clearly, null vectors exist only if the metric is indefinite. Vectors $v, w \in T^xM$ are *orthogonal* (denoted by $v \perp w$), if $\langle v, w \rangle = 0$. In indefinite metric spaces, the concept of orthogonality differs considerably from Euclidian orthogonality. A null vector, e.g., is always orthogonal to itself.

Example 4.1. In a two-dimensional Minkowski space $\mathbb{R}^2_1$, the scalar product

is $\langle v, w\rangle = -v_0w_0 + v_1w_1$.[13] Null vectors span lines at 45° with the coordinate axes (see figure 4.2, left). The set of all null vectors is called the *nullcone* or *lightcone*—as in the last chapter—because null vectors describe the path of light rays departing from or arriving at the origin. Curves with $\langle v, v\rangle = \pm a$, where $a > 0$, are hyperbolas asymptotic to the null lines. Figure 4.2 (right) shows pairs of orthogonal vectors, which are not orthogonal in Euclidian geometry.

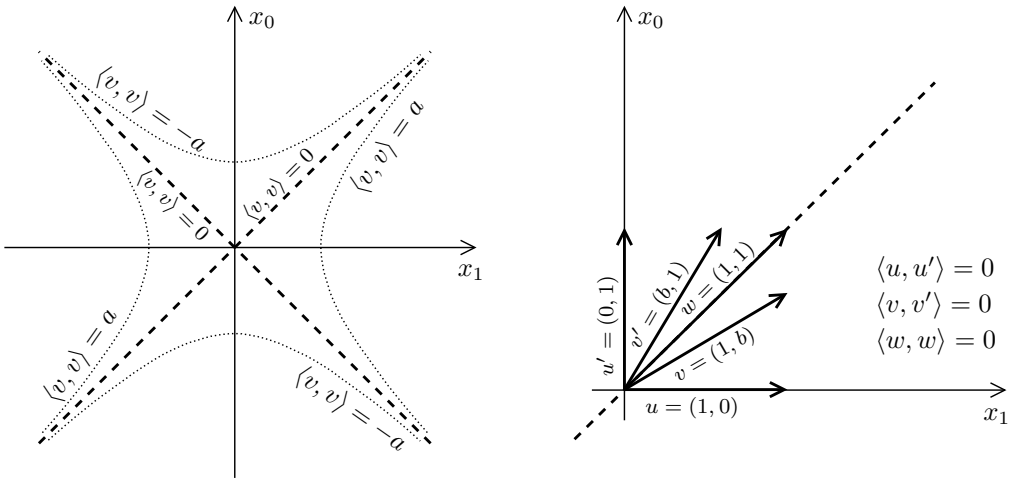


Figure 4.2: Geometry on a Minkowski 2-space $\mathbb{R}^2_1$. Let e_0 and e_1 be unit vectors of $\mathbb{R}^2_1$ of x_0- and x_1-directions, respectively. Arbitrary vectors can be written in the form $v = v_0e_0 + v_1e_1$ and expressed conventionally in drawings, which, in fact, are spacetime diagrams. Left: Nullcones and curves of constant norm, i.e., curves with $\|v\|^2 = |\langle v, v\rangle| = a$, where a is an arbitrary positive constant. Right: u and u' are orthogonal in both Euclidian and Minkowskian geometry, whereas v and v' are Minkowski-orthogonal but not Euclidian-orthogonal. The nullvector w is orthogonal to itself.

The *norm* of a vector $v \in T^xM$ is $\|v\| = |\langle v, v\rangle|^{\frac{1}{2}}$. Notice that it is always non-negative, even if the scalar product $\langle v, v\rangle$ can be negative. In the language of special relativity discussed above, the norm of a vector equals the absolute value of the vector's spacetime interval, i.e., proper time or proper distance between the vector's end events.

[13]In a Minkowski space and thus locally in a Lorentzian manifold, the scalar product can be expressed as a sum with one minus sign and $(n-1)$ plus signs. Components with a minus sign represent time components and are customarily denoted by a zero index. We use this custom henceforth.

Naturally, a *unit vector* is a vector with the norm 1, i.e., $\langle v, v\rangle = \pm 1$. The concept allows us to construct an *orthonormal basis* for all finite-dimensional scalar product spaces $V \neq 0$ [53].

The matrix representing the semi-Riemannian metric relative to an orthonormal basis $e_1, e_2, \ldots, e_n$ is diagonal; in fact, $g_{ij} = \langle e_i, e_j\rangle = \delta_{ij}\epsilon_j$, where δ_{ij} is the Kronecker delta and $\epsilon_j = \langle e_j, e_j\rangle = \pm 1$. By convention, in the orthonormal basis, vectors are ordered so that the negative signs of the diagonal are placed first, and the diagonal ϵ_i is called the *signature* of the scalar product. The number of negative signs in the signature equals the index of the manifold. In addition, the above shows us that the index is the number of negative eigenvalues of the scalar product, or the number of negative eigenvalues of the matrix representing the metric tensor in some basis—independent of the choice of basis. As a corollary to the previous definition, an arbitrary $v \in T^x M$ has the unique expression

$$v = \sum_{i=1}^{n} \epsilon_i \langle v, e_i\rangle e_i. \tag{4.6}$$

Let us generalize the concept of causality described in section 3.5 into a curved Lorentzian manifold, or even to a generic semi-Riemannian manifold. The speed of light describes the events in spacetime that can communicate, i.e., whether information can be carried from one event to another. This divides spacetime into different parts, depending on whether a cause-effect link exists between two events or not.

Every physical observer agrees on the speed of light locally in his own coordinates but may perceive it differently near a distant physical observer because of the curvature of the spacetime (see above similarity of local metrics). Nevertheless, local directions from a point—tangent vectors—are uniquely divided by the causal character determined by the speed of light.

This division by the causal character of tangent vectors applies also in all semi-Riemannian manifolds with an indefinite metric, yielding three possible *causal characters* to a tangent vector:

spacelike	if	$\langle v, v\rangle > 0$	or	$v = 0$
null	if	$\langle v, v\rangle = 0$	and	$v \neq 0$
timelike	if	$\langle v, v\rangle < 0$		

The set of all null vectors of $T^x M$ is the *nullcone* at x, and all timelike vectors form a dual cone called the *timecone*. In a Lorentzian manifold, a null vector is also called the *lightlike* vector and a set of all null vectors is a *lightcone*; a nonspacelike vector is said to be *causal*, and as a whole they form a *causal cone*. As can be seen, the terminology of Lorentzian (and partially

also semi-Riemannian) manifolds derives from the theory of relativity. For instance, if an object travels at the speed of light, it travels like a light ray, and the tangent vector of its worldline is said to be lightlike.

Definition 4.9. An *isometry* between the semi-Riemannian manifolds M and N is a diffeomorphism $\phi : M \to N$ that preserves the metric

$$\langle \mathrm{d}\phi(v), \mathrm{d}\phi(w) \rangle = \langle v, w \rangle \quad \forall v, w \in T^x M, \ \forall x \in M. \tag{4.7}$$

If there is an isometry between M and N, they are said to be *isometric*. Roughly speaking, it means that the manifolds are geometrically identical. For example, a scalar product space V with dimension n and index ν is, as a semi-Riemannian manifold, isometric to a semi-Euclidian space $\mathbb{R}^n_\nu$. In fact, the coordinate isomorphism of the orthonormal basis of V is such an isometry. Moreover, each tangent space of a semi-Riemannian manifold is linearly isometric to a semi-Riemannian space of the same dimension and an index like the manifold itself. Especially, the tangent space of a Lorentzian manifold is isometric to a Minkowski space. [53]

Causal character of the submanifold

Let P be a submanifold of a semi-Riemannian manifold M. Since each tangent space $T^x P$ is regarded as a subspace of $T^x M$, we can restrict the metric of M to $T^x P$ at each $x \in M$; i.e., we define the restriction $\langle \cdot, \cdot \rangle^P_x = \langle \cdot, \cdot \rangle_x |_{T^x P \times T^x P}$. However, if the metric of M is indefinite, $\langle \cdot, \cdot \rangle^P$ need not be a metric on P. It is symmetric but not necessarily nondegenerate. The situation is identical with the case where a subspace of an indefinite scalar product space does not need to be nondegenerate. In fact, P is a *semi-Riemannian submanifold*, i.e., a submanifold with a semi-Riemannian metric, if and only if $\langle \cdot, \cdot \rangle^P_x$ is nondegenerate on each $T^x P$, and if the index is the same for each $x \in P$. With a Lorentzian manifold M, if $\langle \cdot, \cdot \rangle^P_x$ is positive definite for each $x \in P$, P is a *spacelike* submanifold; if $\langle \cdot, \cdot \rangle^P_x$ is nondegenerate with the index one, P is *timelike*. Consequently, an arbitrary semi-Riemannian submanifold of a Lorentzian manifold has a *causal character*: it is either spacelike or timelike. [53]

Let us consider a third possible causal character. A submanifold P is said to be *lightlike* or *null*, if $\langle \cdot, \cdot \rangle^P_x$ is degenerate at each $x \in P$. Hence, for each x in the lightlike submanifold P, there exists a nonzero $v \in T^x P$ such that $\langle v, w \rangle^P_x = 0 \ \forall w \in T^x P$. In other words, there is a vector field $V : P \to TP$ so that $\langle V_x, w \rangle^P_x = 0 \ \forall w \in T^x P$ and $\forall x \in M$. In the Lorentzian manifold. the null vector field V is unique up to a scale factor. For example, (future and past) lightcones in a semi-Riemannian n-space $\mathbb{R}^n_1$ are lightlike hypersurfaces,

i.e., lightlike submanifolds of dimension $n-1$. In this case, V_x at $x \in P$ is a null tangent vector of the lightcone P and thus clearly unique up to the scaling. [18, 25, 28]

The properties of lightlike submanifolds are quite cumbersome, e.g., $T^xP \cap (T^xP)^\perp \neq \{0\}$, and thus the definition of the normal direction of a lightlike hypersurface is more demanding than that of (nondegenerate) semi-Riemannian hypersurfaces. For instance, if P is a lightcone in $\mathbb{R}^n_1$, $(T^xP)^\perp \subset T^xP$ for each $x \in P$ and thus $(T^xP)^\perp$ does not define a proper normal direction. [18, 43]

An arbitrary submanifold of a Lorentz manifold do not necessarily have a causal character, i.e., to be spacelike, timelike or lightlike. However, at each $x \in M$, each subspace of the tangent space T^xM has a causal character.

4.1.6 Orientation and time-orientation

A Lorentzian manifold has two separate concepts for sense of direction, called orientation and time-orientation. The former pays attention equally to all dimensions; the latter concerns only the time-dimension. Because orientation needs no metric structure, it is a property of a smooth manifold, whereas time-orientation is a special characteristic of a Lorentzian manifold.

Orientation

Orientation gives a unified abstraction to concepts such as direction of a line, sense of rotation on a surface, and left-handness or right-handness in a three-dimensional space. These are important, e.g., in integration, where the elementary integral of a real variable $\int_a^b f(x)dx$ changes its sign, if we change the direction of the integration from b to a. The sign thus changes when the direction—i.e., orientation—of the line segment $[a, b]$ changes. As we will see later, the same can be generalizes into integration over surfaces, volumes, and higher-dimensional regions.

Orienting a vector space V means fixing the order of the vectors in the basis. Two bases of V, $e_1, \ldots, e_n$ and $e'_1, \ldots, e'_n$, have the *same orientation* provided the change-of-basis linear transformation A—$e'_i = \sum A^j_i e_j$—satisfies $\det A > 0$. If $\det A < 0$, they have the *opposite orientation*. Having the same orientation brings an equivalence relation to the bases of V and divides the bases into two equivalence classes, called *orientations* of V. The bases of V belonging to the chosen orientation are called *positively oriented*; the rest are *negatively oriented*.

The *orientation* of a smooth manifold M assigns to each point $x \in M$ the orientation of the tangent space T^xM. We can always do this in a locally

coherent way in the domain of one chart as follows [38, 69]. Let χ be one chart and U its domain. We then have a bijective relation between the tangent vectors in T^xM and the vectors of $\mathbb{R}^n$ via $\mathrm{d}\chi$ for every $x \in U$. Let $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ be positively oriented bases of T^xM and T^yM respectively, where $x, y \in U$. T^xM and T^yM have the same orientation, if $(\mathrm{d}\chi(x)x_1, \ldots, \mathrm{d}\chi(x)x_n)$ and $(\mathrm{d}\chi(y)y_1, \ldots, \mathrm{d}\chi(y)y_n)$ have the same orientation in $\mathbb{R}^n$. If the orientation is the same for every tangent space T^xM, $x \in U$, it is locally coherent.

Similarly, we can try to extend this local orientation consistently to the whole manifold. If two domains of the charts overlap, we can compare their orientations with transition mappings and usually make the orientations compatible by possibly changing the orientation of one domain. But it is not always possible to finish this procedure, because there are manifolds whose orientation cannot be made consistent in every chart of their atlas. But if this is possible, we say that the manifold is orientable.

Definition 4.10. Let χ_i and χ_j be two charts of the differentiable structure of a differentiable manifold, and let U_i, U_j be their domains, respectively. If, for arbitrary charts χ_i and χ_j, $\det((\chi_i \circ \chi_j^{-1})') > 0$ at every point $x \in U_i \cap U_j$, the manifold is *orientable*.

Remark 4.3. If the manifold is orientable, we can divide the charts into two equivalence classes with the equivalence relation $\chi_i \sim \chi_j$ if $\det((\chi_i \circ \chi_j^{-1})') > 0$. If we choose one of the classes, we choose an orientation consistently for each tangent space. Thus an equivalence class O_r describes the orientation of the orientable manifold.

For example, $\mathbb{R}^n$ is orientable, but a Möbius strip is not. A connected orientable smooth manifold can be oriented with the above procedure by choosing a positively oriented frame in one tangent space. With a Möbius strip, the procedure runs into trouble because orientation cannot be consistent in the whole atlas.

Time-orientation

There are two timecones in each tangent space T^xM of a Lorentzian manifold M, and we have no intrinsic way of telling which one of them points to the future and which to the past. We must choose one of them as the future-pointing cone, i.e., to *time-orient* T^xM.

Definition 4.11. If it is possible to smoothly time-orient each tangent space, the manifold is *time-orientable*. Smoothness means that for some neighborhood U of each $x \in M$, there is a smooth vector field v such that v_y lies in a future timecone for $\forall y \in U$.

If the manifold is time-orientable, choosing a specific time-orientation on M means that we *time-orient* M. A Lorentzian manifold is time-orientable if and only if there is a smooth timelike vector field [53]. For example, a Minkowski space $\mathbb{R}^n_1$ is time-orientable with a constant time-directional vector field v such that $v_x = e_0$. Generally, every simply connected Lorentzian manifold is time-orientable [4, 53].

It is worth noticing that for a Lorentzian manifold there is no relation between orientability and time-orientability. For example, we can assign a time-orientable Lorentz metric to a Möbius band or a non-time-orientable metric to some orientable manifold.[14]

4.1.7 Curves and geodesics

Curves

We need some abstraction to model several kinds of curves: worldlines, geodesics and so on. To define tangent vectors, we used curves of a special kind, but the above definition of finite curves is not enough to model infinite worldlines. Hence we need a more general definition.

Definition 4.12. A *curve* is a continuous mapping $\alpha : I \to M$, where I is an open and possibly infinite interval of $\mathbb{R}$.

We obtain a *smooth curve*, if we require that α be a smooth mapping. The domain of a curve can also be a closed interval, and then the curve is called a *curve segment*. A (smooth) curve segment $\alpha : [a, b] \to M$ is a function that has a continuous (or smooth) extension to an open interval containing $[a, b]$. Thus a tangent vector $\alpha'(s)$ is also defined at the endpoints a and b. A curve is *piecewise smooth*, if there is a partition such that each part of the curve is a smooth curve segment. A curve, defined in an open interval I, is piecewise smooth provided the restriction to every closed interval is a piecewise smooth curve segment.

Sufficiently near points of the manifold can be connected by a smooth curve segment. It follows that a manifold is connected, if and only if any two of its points can be joined with a piecewise smooth curve segment. A manifold is *simply connected*, if it is connected, and if every closed curve can be continuously contracted to a point.

[14] Orientation is a metric-independent property, but time-orientation needs a Lorentz metric, because we cannot separate time-direction from other directions without a Lorentz metric. In fact, time-orientation depends only on the causal structure of the manifold, i.e., an equivalence class of metrics with an equal causal structure.

For simplicity, let a piecewise smooth curve $\alpha : I \to M$ be defined on $I = [0, b)$, where $b \leq \infty$. α is *extendible*, if it has a continuous extension $\tilde{\alpha} : [0, b] \to M$; otherwise it is *inextendible*. The case of a more general curve is given similarly.

A curve α is an *integral curve* of a vector field v, if $\alpha'(t) = v_{\alpha(t)}$ for all $t \in I$. Accordingly, at each point the curve α has v as the velocity vector. A vector field has a variety of integral curves. However, around a point x a (smooth) vector field v has a unique (smooth) integral curve α such that $\alpha(0) = x$. [53]

Causality relations

We mentioned earlier that a vector is timelike, null, or spacelike, i.e., has a unique causal character. At each point, a curve has a causal character, which is determined by the causal character of the velocity vector at that point. This causal character does not need to be the same for all points of the curve, but we are mostly interested in the curves with a constant causal character at every point. Therefore, we define the *causal character of a curve*: a curve is spacelike, null, or timelike, if $\alpha'(s)$ is spacelike, null, or timelike $\forall s \in I$. In a Lorentzian manifold, a *causal curve* is one whose velocity vectors are all causal.

In a time-oriented Lorentzian manifold, a (nowhere vanishing) causal vector field along a curve cannot continuously change from future-directed to past-directed. It follows that a (smooth) causal curve is always either future-pointing or past-pointing. [4]

Causality answers questions such as which events in a Lorentzian manifold can be joined by a causal curve, i.e., which events can influence or be influenced by each other. *Causality relations* on a connected time-oriented Lorentzian manifold M are defined as follows. If $x, y \in M$, then

1. $x \ll y$ means that there is a future-pointing *timelike* curve in M from x to y.

2. $x < y$ means that there is a future-pointing *causal* curve in M from x to y.

3. $x \leq y$ means that either $x < y$ or $x = y$.

The events x and y are *causally related*, if $x < y$ or $x > y$. The *chronological future*, of $x \in M$ is the subset $I^+(x) = \{y \in M | \, x \ll y\}$, and the *causal future* is $J^+(x) = \{y \in M | \, x \leq y\}$. The above definitions have their dual past versions, and *chronological* and *causal past* are denoted by $I^-(x)$ and $J^-(x)$, respectively.

For example, in a Minkowski space $\mathbb{R}^n_1$ the chronological future $I^+(x)$ is the future timecone of x, and $J^+(x)$ is the point x together with the future causal cone. Hence, $I^+(x) \subset J^+(x)$.

The causality relation $\leq$ is not necessarily a partial order on M, because it is possible that both $x \leq y$ and $x \geq y$ with two district points $x, y \in M$. This is possible, if there are *closed causal curves*. In spacetime, this would mean that a signal sent from event x could traverse some path and return to x—the place and time where it was sent—and thus violate causality. In contrast, such a manifold where $\leq$ is a partial order is said to be *causal* [4]. Furthermore, a manifold that does not even contain any timelike curves is said to be *chronological*. For our purposes, causal—or even chronological—manifolds are enough; thus we assume all connected Lorentzian spacetime manifolds to be causal.

In some cases, we need even a stricter notion of causality. A Lorentzian manifold is *strongly causal*, if every point x has arbitrarily small neighborhoods such that no causal curve that leaves one of these neighborhoods ever returns.

Geodesics

We are especially interested in a special class of curves having a causal character, called geodesics. They generalize the idea of straight lines for a curved manifold: they are locally straight in every local Lorentz frame. For instance, in a curved spacetime, the worldline of an object free of (nongravitational) forces is a geodesic. In a Riemannian manifold, a geodesic is defined as the shortest path between points. In a semi-Riemannian manifold with an indefinite metric, the definition is not so straightforward.[15] At least, we need a definition for the length of a curve.

Let $\alpha : [a, b] \to M$ be a piecewise smooth curve segment in a semi-Riemannian manifold M. The *arc length* of α is

$$L(\alpha) = \int_a^b \|\alpha'(s)\| ds, \tag{4.8}$$

where $\|\alpha'\| = |\langle \alpha', \alpha' \rangle|^{\frac{1}{2}}$ by the definition of norm. The length is independent

[15]For a precise definition, we should define *parallel transport* (or equivalently *connection* or *covariant derivative*), which is essentially a method for transporting tangent vectors along curves. A geodesic is a curve that parallel-transports the tangent vector of one point to the tangent vector of another point while keeping the magnitude of the vector. But this definition is beyond our scope. For us, the most important thing is that a metric determines a unique parallel transport.

of the parametrization of the curve, because it can be shown that the length remains unchanged by monotone reparametrization [53]. In the context of special relativity (section 3.4), we chopped a worldline into small straight segments and summed the intervals of the segments. The result was the proper time of the worldline. The above integral can be regarded as the limit of this summation process, when the segments are shrunk into smaller and smaller.

It is quite well known how curve length behaves in a Riemannian manifold, but indefinite metrics is somewhat different. For example, a null curve has the length zero, because $\|\alpha'(s)\| = 0$ for every $s \in [a, b]$ by the definition of null curve. The interpretation of the arc length of a timelike worldline in a curved spacetime is the same as in special relativity: the proper time or aging along the worldline. Thus we have a twin paradox (see section 3.4), which says that a kinked worldline is shorter than the straight one, and that the straight one is, in fact, the longest worldline between timelikely connected events. Note that this interpretation with proper time is valid only in a specific Lorentzian manifold; other semi-Riemannian manifolds are more complicated.

Note that arc length is always non-negative, even with a timelike curve in which case $\langle \alpha'(s), \alpha'(s) \rangle$ is negative at every point of the curve.

Definition 4.13. A curve with a causal character in a semi-Riemannian manifold is a *geodesic*, if it is null or, alternatively, if it is timelike or spacelike, and if it extremizes the arc length among curves between its endpoints.

Thus, first, all null curves are geodesics. A timelike geodesic locally *maximizes* the arc length or is a saddle point. Respectively, a spacelike geodesic locally minimizes the arc length or is a saddle point. To emphasize, this geodesic extremizes the arc length generally only locally, and thus there could be several geodesics between endpoints. The minimum is only local, e.g., with two antipodal points of a sphere acting as a Riemannian manifold. There are infinitely many geodesics with the same length between those points [89].

As mentioned before, a geodesic generalize the notion of straight line into a curved manifold. In the special case of a flat manifold—i.e., a manifold with zero curvature—the geodesics are indeed straight lines. For instance, in a semi-Euclidian space $\mathbb{R}^n_\nu$—the Minkowski n-space of special relativity, if $\nu = 1$—geodesics are straight lines; i.e., $\alpha(t) = x + tv$, where $x \in M$ is a point, and v is a vector describing the direction of the geodesics departing from x.

Distance function

If there is a geodesic joining any two points of a manifold, it is called a *geodesically connected*[16] (or *geodesic*) manifold. In a geodesically connected Riemannian manifold, geodesics allow us to speak about distances between points; the distance between two points in a Riemannian manifold is the length of the shortest geodesic joining these points. If the manifold is not geodesically connected, yet connected, *Riemannian distance* is the greatest lower bound of the arc lengths of the piecewise smooth curve segments joining the points. Moreover, in the Riemannian case, distance fulfils the axioms of metric [45].

The case of the Lorentzian manifold with an indefinite metric is more complex, because the geodesic is not a length-minimizing curve. If the manifold is geodesically connected, and if the geodesic joining two arbitrary events is unique, a (geodesic) distance can be defined analogously with the geodesically connected Riemannian case. A neighborhood $U(x)$ of a point x, where any two events can be joined by a geodesic lying entirely in $U(x)$, is called a *convex neighborhood*. For example, an entire Minkowski space is a convex neighborhood to each of its points. It is shown in [4] that there is a convex neighborhood around an arbitrary event in a Lorentzian manifold.

In a Lorentzian manifold satisfying the above conditions of a unique geodesic, *geodesic distance* is the length of the unique geodesic segment joining the events. Geodesic distance does not fulfil the axioms of metric, because, precisely, the distance is not positive definite, because the distance between two distinct events can be zero. Later we will see that triangle inequality is also violated.

Even if geodesic distance in a Lorentzian manifold is not as useful as Riemannian distance due to the above shortcomings, there are two occasionally more useful associates: Lorentzian distance and Synge's world function.

Let x and y be events in a Lorentzian manifold M. The *Lorentzian distance* function is defined by

$$d(x,y) = \begin{cases} \sup_{\gamma\in\Gamma} L(\gamma) & \text{if } x \leq y, \\ 0 & \text{otherwise,} \end{cases} \tag{4.9}$$

where Γ is the set of causal curves from x to y. Thus Lorentzian distance equals geodesic distance, if y is in the causal future of x, and if the points are

[16] Geodesically connectedness differs from (path) connectedness. The simplest example of a connected but not geodesically connected space is $\mathbb{R}^2 \setminus (0,0)$. There is no path between points $(-1,0)$ and $(1,0)$ which attains the minimum distance (i.e., value 2) between these points; i.e., there is no geodesic between these points. Thus there are non-compact manifolds where the conceps differ.

connected with a unique geodesic. Lorentzian distance has many properties similar to Riemannian distance though with some exceptions. For example, the distance function is not symmetric, because if $x < y$ then $d(x, y) \neq 0$ but $d(y, x) = 0$. Moreover, if $x \leq z \leq y$, Lorentzian geometry obeys the *reverse* (or *backward*) *triangle inequality* $d(x, y) \geq d(x, z) + d(z, y)$ in contrast to the opposite normal triangle inequality of Riemannian distance [4]. A reversal of inequality could be expected, because a causal geodesic in a Lorentzian manifold locally maximizes rather than minimizes the arc length.

The second variation of geodesic distance is *Synge's world function* [70], defined as the function of two points $x, y \in M$, satisfying

$$\Omega(x, y) = \frac{1}{2}\varepsilon L^2, \tag{4.10}$$

where L is the length of the unique geodesic joining the points, and ε is -1, if the geodesic is timelike, and $+1$, if the geodesic is spacelike. If the geodesic is lightlike, $L = 0$, and we can choose $\epsilon = \pm 1$, which is the most convenient. Hence, the absolute value of the world function is the squared geodesic distance between the points. The most striking benefit of the world function is that it is everywhere differentiable in terms of both its arguments, a property that even Riemannian distance lacks.

In the Riemannian manifold, the distance function is compatible with the original topology of the manifold; i.e., a topology can be generated from the open balls defined by the distance function. But with the general Lorentzian manifold, the connection between distance function and topology is not so straightforward. In the special case of the strongly causal Lorentzian manifold, however, the manifold topology agrees with the causality-related *Alexandrov topology* of M [4]. Alexandrov topology is one with the basis $\{I^+(x) \cap I^-(y) |\, x, y \in M\}$. Thus this basis consists of sets that are intersections of the chronological future and past of two points. For instance, in the Minkowski space, the bases are the intersections of the future and past timecones of separate points. Hence, in the strongly causal case, the topology of the Lorentzian manifold is uniquely determined by the causal structure of the manifold, but not the distance function itself, because M embodies several metrics with an identical causal structure.

4.1.8 Spacetime and general relativity

At this point, we have now enough mathematical structures to describe a definitive model for spacetime and introduce how the above concepts are utilized in general relativity. Because relativity is only a secondary matter

for us, this section—and relativity alike—is only an example of how the spacetime model could be exploited.

First, we introduce a model for spacetime, which is used both in general relativity and in our models for computational electromagnetics. Second, we exploit this model and briefly give the basics of general relativity.

Spacetime

Definition 4.14. The *spacetime* M is a connected time-orientable Lorentzian manifold of dimension ≥ 2.

Connectedness and time-orientability along with an indefinite Lorentzian metric provide causality relations for spacetime. The time-orientation of M is called *future* and its negative is *past*. As already mentioned, a tangent vector in a future causal cone is said to be *future-pointing* or *future-directed*, and, similarly, a causal curve is future-pointing, if all its velocity vectors are future-pointing.

Familiarly, the points of spacetime are called *events*. We give the name *particle* to the entity modeling the life history of an object that is negligibly small and whose path through spacetime is called a *worldline*. Particles are divided into two groups, material and lightlike particles, according to their worldline's causal character. First, the worldline of a *material particle* in M is a timelike future-pointing curve $\alpha : I \to M$ such that $\|\alpha'(\tau)\| = 1$ for all $\tau \in I$. The parameter τ is called the *proper time* of the particle. This definition is in accordance with the discussion in chapter 3, bacause the integral of arc length (4.8) with $\|\alpha'(\tau)\| = 1$ gives the total proper time along the object's path. Every particle carries a clock that measures its proper time. The velocity vector $\alpha'(\tau)$ is called the *four-velocity* of the particle. Secondly, the worldline of a *lightlike particle* is a future-pointing null geodesic $\gamma : I \to M$. Only massless objects can be identified as lightlike particles; contemporary physics recognizes only photons as lightlike and speculates about massless gravitons.[17] [53]

On physical grounds, spacetime is assumed to be chronological (or causal); i.e., it does not contain timelike (or causal) curves. If a closed timelike curve does exist, a person might traverse along this curve (as her worldline) and meet herself at an earlier age. Similarly, a closed causal curve would violate causality, a situation we wish to avoid.

A compact spacetime always contains a closed timelike curve. Consequently, spacetimes are usually considered noncompact in general relativity.

[17] As of 2010, contrary to earlier belief, neutrinos are believed to have a nonzero mass [20].

Luckily, the above result applies only to manifolds without boundary, whereas a spacetime with a boundary can be compact without causality problems. For our special purposes in computational electromagnetics, spacetimes always have a boundary. Another benefit is that any noncompact manifold admits a Lorentzian metric [4].

In the forthcoming chapters, we take a causal spacetime with a boundary as our background model for geometry in electromagnetic computation.

General relativity

As mentioned in section 3.8, momentum-energy causes spacetime to curve. The curving is determined by the *Einstein field equations*, which state that the *Einstein tensor* and the *stress-energy tensor* are linearly dependent. The Einstein tensor is related to the curvature of spacetime and is a function only of the metric and its derivatives. The components of the stress-energy tensor contain the flux of the components of momentum-energy across the surface of one constant coordinate. The solution of the Einstein field equations with a known stress-energy tensor is a metric of the spacetime manifold, which then gives possible forms of spacetime.[18] On the other hand, any (metric of) Lorentzian manifold is a solution of the Einstein field equations for some conceivable—but generally physically unreasonable—stress-energy tensor. [77, 67]

The simplest example of spacetime is the Minkowski spacetime, a spacetime where momentum-energy—and thereby also stress-energy—is zero. The *Minkowski spacetime* is a spacetime isometric to Minkowski 4-space $\mathbb{R}^4_1$. The Minkowski spacetime is simultaneously the geometry of special relativity and the geometry induced on each fixed tangent space of an arbitrary Lorentzian manifold. It plays the same role for a Lorentzian manifold that Euclidian geometry plays for a Riemannian manifold. Often the Minkowski spacetime is called *flat spacetime*, but strictly speaking, a spacetime is flat whenever its curvature is zero. [4]

Other well-known spacetimes related to feasible nonzero stress-energy include the *Schwarzschild*, the *Kerr*, and the *Robertson-Walker spacetimes*. The Schwarzschild spacetime is an empty spacetime outside a rotating and spherically symmetric body. It models the gravitational fields of slowly ro-

[18] In problems of this kind, general relativity uses a differentiable manifold as a starting point, i.e., as a background manifold. The solution of the problem is a Lorentz metric, and thus, a Lorentz manifold, when the metric is inserted in the background manifold. Our electromagnetic problems differ from this, because we take the Lorentz manifold readily as our starting point—as our background manifold. The solution to our problems is the electromagnetic field on that Lorentz manifold.

tating bodies such as stars, planets, and (dead) nonrotating black holes. In contrast, the Kerr spacetime represents stationary axisymmetric metrics outside rotating objects and corresponds to gravitational fields outside a rotating black hole. Last, the Robertson-Walker spacetimes are used in cosmological models of the big bang.

Curvature describes how objects move. The worldline of an object free of nongravitational forces is a timelike geodesic. This is strictly valid only for an object with a mass and energy small enough so that the object itself does not affect the curvature of the spacetime.

As we discovered earlier, in a small enough region, the geometry of the curved spacetime of general relativity is similar to that of the Minkowski space. Or, as we phrased it in chapter 3, in a small enough region of spacetime one cannot detect the effects of gravity. In these small regions the worldlines of free particles are (detected as) straight and thus their frames are inertial. We can do local Lorentz transformations from one frame to another, and all the transformations form a local Lorentz group. However, on a whole curved spacetime we do not have symmetry groups such as the Lorentz group; we have only local symmetries.

4.2 Observer

All our experience about any physical phenomenon depends on experimental devices; i.e., they rely on material objects in the vicinity of some observer. Thus all our experiences are relative to an observer. Those perceptions reflect not only the phenomenon but also the properties of the observer and its relation to the phenomenon. [44]

Our aim is to formulate the whole electromagnetic theory with no reference to any observer. Thus we try to separate all properties of electromagnetic phenomena from the impact of the observer. This is historically done by comparing the results of various observers and then extracting observer-independent information from these various observations.

Even if the theory is formulated independent of observer, it is sometimes more illustrative to consider phenomena from the point of view of one chosen observer. In our case, this connects our spacetime electromagnetic formulation to the more familiar three-dimensional one.

Our previous definition of physical observer is not enough for us, because one physical observer can make observations only from material points differentially close to its worldline. We want to widen this definition so that one observer can perceive the whole spacetime or a domain in it. Because our physical observer is defined by a worldline and the charts enveloping the

vicinity of this worldline, we need a collection of these physical observers such that their charts fill the whole spacetime. Because these worldlines continuously fill the spacetime, their velocity vectors define a continuous vector field u in M, called the *velocity field* of the observer. At every event, this field gives the velocity of the observer.

Every physical observer has its own local hyperplane of simultaneity. We can build a hypersurface such that its tangent hyperplane coincides with that of the local hyperplane of simultaneity of each physical observer. Additionally, we want this hypersurface to be a *Cauchy surface*; i.e., a hypersurface such that any inextensible causal curve (such as every worldline) pierces it exactly once. This hyperplane is called the *foliation* of M with respect to this observer. Each worldline is parametrized with its own proper time τ. Thus physical observers can be time-synchronized by choosing a single foliation corresponding to $\tau = 0$ on each worldline. We then parametrize the foliations by τ and denote them by M_τ. Because the submanifold M_τ consists of simultaneous events (i.e., at time τ) in the point of view of this (global) observer, we call M_τ a *purely spatial submanifold*.

Formally, an *observer* is the pair $(u, \tau \mapsto M_\tau)$ of a continuous causal vector field u and a foliation $\tau \mapsto M_\tau$ such that each tangent plane of M_τ is Lorentzian orthogonal to u. This requirement of orthogonality could be omitted, but we want to speak about simultaneity from the standpoint of the observer, and the simultaneity of M_τ is equivalent to u_x and $T^x M_\tau$ being orthogonal. Consequently, from the point of view of the chosen observer, an arbitrary event x in the spacetime M can be described by a pair of time τ and point x_τ in the purely spatial submanifold M_τ. For an observer given by the pair $(u, \tau \mapsto M_\tau)$, the integral curves of u are known as the worldlines of the observer.

Now we have an observer that we can use to perceive the whole manifold M. In the observer-based approach, it is customary to choose local charts so that in their standard basis $e_0, e_1, \ldots, e_{n-1}$ (cf. section 4.1.3) the vector e_0 agrees with the direction of u_x. The other basis vectors are purely spatial; i.e., they belong to $T^x M_\tau$ for some τ. Thus e_0 is the local basis vector in the time direction.

Even if the definitions of physical observer and the above observer are divergent, they model the same concept of real life. So if we claim that an entity is independent of any observer, it means the same thing from the point of view of both definitions.

4.3 Cells and chains

Electromagnetic quantities are inherently related to geometric objects. They are integral quantities, giving, e.g., circulations and fluxes on geometrical objects. For instance, electric field is modeled with a quantity that gives an electromotive force along an arbitrary path. Furthermore, we model electric current with a quantity that gives a current flux trough any properly oriented surface. Electromagnetic quantities are by their nature macroscopic quantities such as total electric charge of a capacitor plate, compared to local measures such as electric charge density. These quantities are mappings that take some geometric objects and return a number that represents the numerical value of the quantity on that object. Thus electromagnetic quantities belong to some dual space of some unknown space of geometries.

To start with, we should find the proper tools for macroscopic modeling of geometry so that the structure suits our purposes in electromagnetic computation, and we must define some mathematical expressions for our geometrical objects. Because, at bottom, we are seeking a linear map to model electromagnetic fields, we need a linear space structure also for the geometrical objects. That is, we must have mathematical structures that allow the addition and subtraction of geometrical objects. But why do we need a linear structure in the first place? Because it enables us to use linear algebra in computation, which is reason enough to demand a linear space structure for geometrical objects.

We start by defining the polyhedral cell, an elementary structure lying beneath all our geometric structures. Second, we define the polyhedral chain as a formal sum of polyhedral cells. For auxiliary structures, we need a boundary operator and such a topology for chains that the topology is complete (as a uniform space). In the end, we assemble all the geometrical structures into a compound with a chain complex entity (for advanced definitions, see [91, 73]).

4.3.1 Polyhedral cell

The polyhedral cell is a simple macroscopic model for geometric objects. The volumes, faces, edges, and vertices in the element mesh of the finite element method or geometric methods are examples of polyhedral cells. They are defined on a manifold, and like other structures on manifolds, they are defined with the aid of the corresponding structures in coordinate space. To define a polyhedral cell, we need a convex cell in coordinate space as an auxiliary structure.

Convex cell

For each subset D of a coordinate space $\mathbb{R}^n$, there exists a closed minimal convex set that contains D, called the *convex closure* of D. The convex closure of a finite set of points $P = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k\}$ in $\mathbb{R}^n$ is a *convex cell* $c(P)$, for short c. Its dimension is the dimension of the smallest affine subspace $A(c)$ that contains P; similarly, the orientation of a convex cell is the orientation of the affine space where the cell lies. In particular, when the points $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k$ are independent[19], $c(P)$ is called a k-dimensional *simplex*. [36]

Polyhedral cell

Informally, the *polyhedral cell* is the preimage of a convex cell by a chart. More precise, it is an equivalence class of pairs formed from a compatible chart and a convex cell. A chart is compatible with the convex cell, if the convex cell is in the range of the chart. Two pairs (χ_1, c_1) and (χ_2, c_2) are equivalent, if $\chi_2 \circ \chi_1^{-1}(c_1) = c_2$. The convex cells on both sides of the above equation must have an equal orientation; i.e., equality includes a comparison of orientation. The dimension and orientation of a polyhedral cell are determined by the corresponding convex cell. A p-dimensional polyhedral cell is written for short p-cell.

The polyhedral cell is also called an *inner oriented cell*. The orientation (now renamed as *inner orientation*) of a cell is an internal issue for the cell, or better phrased, a private character of the corresponding convex cell and the affine space where it lies. Let σ_1 and σ_2 be two polyhedral cells equivalent as sets, and let $(\chi_1, c_1), (\chi_2, c_2)$ be their representative pairs. The orientation of σ_1 and σ_2 can be compared by

$$\mathrm{ori}(\sigma_1, \sigma_2) = \mathrm{ori}(A(c_1), A(\chi_1 \circ \chi_2^{-1}(c_2))), \tag{4.11}$$

where the latter $\mathrm{ori}(\cdot, \cdot)$ is the orientation comparison of two affine subspaces. $A(c_1)$ is the smallest affine space containing c_1, and $A(\chi_1 \circ \chi_2^{-1}(c_2))$ is the smallest one that contains $\chi_1 \circ \chi_2^{-1}(c_2)$. The p-cells σ_1 and σ_2 *overlap*, if a p-cell σ' exists such that $\sigma' \subset \sigma_1$ and $\sigma' \subset \sigma_2$.

But for some purposes, we need a surface that is oriented by its crossing direction. For example, electric current is naturally described by its flow *trough* a surface, so the positive crossing direction and the sign of the total current gives the direction of the total net current through the surface.

[19] Choose one of the points and collect the $k-1$ translation vectors from this point into the other $k-1$ points. If the vectors are linearly independent, the points are said to be independent.

Generally, such an orientation is called an *outer orientation*. In the coordinate space $\mathbb{R}^n$, the outer orientation of a subspace S is the orientation of the complementary subspace of S. The inner orientation of S together with the orientation of $\mathbb{R}^n$ fixes the outer orientation of S. Similarly, we define the *outer oriented cell* to be an equivalence class of an inner oriented polyhedral cell and the global orientation of the manifold. Let us denote the manifold orientation by O_r; thus O_r assigns to each point x of a smooth manifold M an orientation of the tangent space T^xM. Later, we give a chain and a multivector field representative for global orientation. An outer oriented cell is an equivalence class of pairs (σ, O_r). The pairs (σ_1, O_r) and (σ_2, O'_r) are equal, if σ_1 and σ_2 are equal as sets and $\mathrm{ori}(\sigma_1, \sigma_2) = \mathrm{ori}(O_r, O'_r)$. The latter orientation comparison is the comparison of manifold orientations.

4.3.2 Polyhedral chain

All the information about the electric field is included in its integrals over oriented paths; i.e., the electric field is uniquely determined by its path-integral over all possible oriented paths. The electric field is then an object that yields a real number, when a path is inserted. In order to speak about the linearity of the electric field, we need a linear structure for the paths, and, more generally, a linear structure for the general dimensional group of polyhedral cells.

How do we add cells together? Could the union of cells be used to realize a sum of the cells? This seems valid, if we think, e.g., of the electric circulation of disjoint segments. Take two disjoint segments c_1 and c_2 and calculate the electric circulation over their union. We get

$$\int_{c_1 \cup c_2} e = \int_{c_1} e + \int_{c_2} e.$$

This seems encouraging, but, unfortunately, the equality fails if the segments intersect. So we cannot define $c_1 + c_2 = c_1 \cup c_2$, if we want the electric field to be a linear function. Thus we need a more elaborate structure to create a linear structure for cells—and we resort to the formal sum.

We build a linear space structure for polyhedral cells by defining more abstract objects called polyhedral chains, which are equivalence classes of formal sums of polyhedral cells. The *formal sum* (or *free vector space*) is a structure that, defined on some set A, defines an enlargement of A that has a linear structure [26]. The linear space of the formal sums of polyhedral p-cells is denoted by S_p.

The formal sums of p-cells are written in the form

$$\Sigma = \sum_{\sigma \text{ is a } p\text{-cell}} \alpha(\sigma)\,\sigma, \tag{4.12}$$

where α is a mapping from a group of polyhedral cells to real numbers so that $\alpha(\sigma) \neq 0$ only for a finite set of p-cells. An alternative expression is $\Sigma = \sum \alpha_i \sigma_i$.

The linear space structure of polyhedral chains is inherited from the characteristics of the formal sum. Formal sums have the following properties [26, 73]. A formal sum is zero, if and only if all the scalar coefficients $\alpha(\sigma)$ are zero. The scalar multiple of (4.12) by $\beta \in \mathbb{R}$ is a formal sum with scalar coefficients $\beta\alpha(\sigma)$. The sum of two formal sums is a formal sum with the scalar coefficients of each sum added together.

Because we know the region that a polyhedral cell covers, we can define the mapping $\alpha' : M \to \mathbb{R}$ so that $\alpha'(x) = \sum_{x \in \sigma} \alpha(\sigma)$. Defining the function α' is equivalent to defining a formal sum such as (4.12). The support of α' is the *subset the formal sum covers*.

However, our intention is to construct a structure that models some real geometrical objects, and, unfortunately, the formal sum structure is not congruent with our intention as such. The essential information about a geometrical object should still be the subset that it covers, and there are various formal sums covering the same subset. Consequently, we must identify all those formal sums that are identical in our intuition. By defining an equivalence relation for the formal sums of polyhedral cells, we make the different representations of a geometrical object identical. This equivalence class of formal sums of polyhedral cells is called the *polyhedral chain*. The cells in a formal sum must have the same dimension, and this dimension is also the *dimension* of the chain.

We define equivalence relation by defining the zero class Z_p of formal sums. The elements of the zero class, besides zero formal sums, can be represented by linear combinations of two basic types of elements. First, Z_p includes terms such as $\sigma_+ + \sigma_-$, where σ_+ and σ_- are identical as sets but oppositely oriented. Informally speaking, the minus sign of the scalar coefficient reverses the orientation of the cell.

Second, if a cell σ is subdivided into cells $\sigma_1, \ldots \sigma_m$, and if the orientation of σ_i agrees with that of σ, $\sigma - \sigma_1 - \cdots - \sigma_m$ belongs to the zero class. Informally, only the subset of the chain matters, not how the subset is divided into cells. This works only, if all the cell orientations match; otherwise the situation becomes more complicated.

By equivalence relation, two formal sums are equal, if their difference belongs to the zero class. The quotient space S_p/Z_p is the *space of polyhedral*

chains and it is denoted by $C_p(M)$. It is also customary to identify a cell with the chain that has a representative where the cell is the only member of the summation. In this way, polyhedral cells span the chain space $C_p(M)$.

The quotient space structure brings new aspects to the addition of chains besides the basic formal sum addition rules mentioned above. Given two polyhedral chains $A = \sum a_i\sigma_i$ and $B = \sum b_j\tau_j$, we may find a common subdivision of σ_i and τ_j, denoted by σ'_k. By the above equivalence relation, chains can be written in the form $A = \sum a'_k\sigma'_k$ and $B = \sum b'_k\sigma'_k$ and then $A + B = \sum(a'_k + b'_k)\sigma'_k$.

Even though a formal sum is always finite, the space of polyhedral chains is infinite dimensional, because the cardinality of a set of polyhedral cells is infinite. However, favourably to our purposes, any formal sum and similarly any element of $C_p(M)$ can be represented by a finite amount of information, by real coefficients of the representative formal sum.

By equivalence classification of polyhedral cells, every representative formal sum covers the same subset of M. We say that the chain *covers* this subset. The cells of the representative formal sum of a polyhedral chain can overlap, but we can choose an expression whereby the cells are non-overlapping [91]. Moreover, we can always choose a simplicial representative. With a non-overlapping representative, the subset that the cochain covers is the union of the regions of the representative's cells.

The *space of outer oriented chains* $\tilde{C}_p(M)$ is defined similarly to the above space of *inner oriented chains*. Its members are equivalence classes of formal sums of outer oriented cells; furthermore, it can be identified with the pair (Σ, O_r) of an inner oriented polyhedral chain Σ and a global manifold orientation O_r.

Remark 4.4. Global orientation can be described by an inner oriented n-chain. Whitney has proved that every smooth manifold has a (smooth) triangulation that covers the whole manifold by disjoint simplexes [91]. By orienting each simplex s_i according to the manifold orientation, the chain $\Sigma_M = \sum_i s_i$ represents the global orientation of the manifold. Thus an outer oriented cell can be denoted by (σ, Σ_M) and an outer oriented chain by (Σ, Σ_M).

4.3.3 Boundary operator

Electromagnetic laws connect quantities defined on a geometrical object to other quantities defined on the boundary of that object. Now, if the chain is our structure to model geometries, we need to know its boundary. We start from the boundary of a single cell and then extend the definition to

polyhedral chains.

Boundary of a cell

The boundary of a p-cell contains a combination of several $(p-1)$-cells. How could we handle this selection of cells as a single entity? Naturally, by combining them into a chain, a procedure which also gives a linear structure to the boundary abstraction. This is a major application of chains, an incentive to use chains in the first place.

The boundary operator is a mapping ∂ from the space of p-cells to the space of $(p-1)$-chains. The boundary of a polyhedral cell is defined via the boundary of the corresponding convex cell.

Let P be a set of points defining a convex p-cell $c(P)$. For a set S of a topological space, the boundary is the set of points in the closure of S, which does not belong to the interior of S. A convex cell has such a boundary as a subset of the affine space $A(c) \in \mathbb{R}^n$, with the topology induced from $\mathbb{R}^n$. We call it the *boundary of the convex cell* $c(P)$ and denote it by ∂c. If we take a subset $Q \subset P$ such that $\dim c(Q) = p-1$ and the $(p-1)$-cell $c(Q)$ is the intersection of the boundary of $c(P)$ and a $(p-1)$-dimensional affine subspace of $\mathbb{R}^n$, $c(Q)$ is called the *face* of $c(P)$. Second, the faces of a polyhedral cell are the preimages of the corresponding convex cell's faces by a chart. Similarly, the *boundary of a polyhedral cell* is defined via its representative's convex cell. Because the polyhedral cells are always closed, we have $\partial\sigma \subset \sigma$.

The boundary, as a polyhedral chain, is the formal sum of the faces of polyhedral cell. The scalar coefficient is $+1$ or -1, depending on whether the orientation of the face agrees with the cell's orientation or not. The boundary of a 0-cell is empty, and the boundary of a convex 1-cell from $\boldsymbol{x}$ to $\boldsymbol{y}$ (denoted $c = \boldsymbol{xy}$) is $\partial c = \boldsymbol{y} - \boldsymbol{x}$. Now let $p \geq 2$ and σ be a convex p-cell and c' its face. In addition, let $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$ orient c so that $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ lie in c'. Then the orientation of c' defined by $(\boldsymbol{v}_2, \ldots, \boldsymbol{v}_p)$ agrees with that of c, if $\boldsymbol{v}_1$ points out of c. By definition, the orientation of a polyhedral cell is inherited from the convex cell, and, accordingly, the orientation of a face agrees with the orientation of the polyhedral cell, if the orientations of the corresponding convex cells agree.

The $(p-1)$-faces of a p-cell have their own $(p-2)$-faces and so on up to 0-faces. The 0-faces are called *vertices* of the p-cell, the 1-faces are called *edges*, and the 2-faces are called *facets*.

Boundary of a chain

The definition of the boundary operator is extended to polyhedral chains by requiring the mapping to be linear. In other words, the boundary operator is a mapping $\partial : C_p(M) \to C_{p-1}(M)$ such that if a chain can be represented as $\Sigma = \sum \alpha_i \sigma_i$,

$$\partial\Sigma = \sum \alpha_i \partial\sigma_i. \tag{4.13}$$

The definition does not depend on the representative of the chain's equivalence class, because $\partial\Sigma = 0$ for every member of the zero class $\Sigma \in Z_p$.

4.3.4 Topology and completeness chain spaces

We have already found a linear space structure for chains, but can a chain space also be made into a topological space? Or can it even be equipped with a norm? In addition, not just any topology or norm will suit, because they must reflect the properties of the geometric object modeled by the chain. For example, we require that the boundary operator be continuous with this topology.

First, we try the standard way to produce a topology: we construct a norm for chain space. These norms are grounded on the measure (or volume) of individual cells and then generalized into general polyhedral chains. However, we find now that we cannot use any standard measures in our Lorentzian manifold, because the measure must be compatible with the metric. Therefore, the resulting chain norms have some unwanted properties, and we cannot use them to build a topology for chain spaces. The situation is similar to the semi-Riemannian manifold, where the connection between metric and topology is not as straightforward as in the Riemannian manifold.

We must now define topology independently of chain norms. Because we are modeling electromagnetic quantities with members of the dual space of chains (called cochains), we want the dual space to be complete. Because completeness is not a topological property[20] and because the norm we use is so cumbersome, we need a structure between topology and metric that possess this completeness property, called uniform structure.

[20]We cannot define Cauchy sequences for a general topological space and hence cannot define completeness. In a topological vector space (defined later), however, Cauchy sequences and completeness can be defined. Because a topological vector space always has a natural uniform structure, we treat topological vector spaces as uniform spaces.

Measure of cell

As discussed above, a chain models a real geometric object. Thus the measure of the chain must reflect some reasonable magnitude of the object. Several chain norms have been developed [91, 29], reflecting different intuitions; among them, the so-called flat norm should be the best for our purposes. For auxiliary structure and illustration, we also first define the *mass norm*.

The mass norm is a chain norm that stands for the volume that the chain covers. First, we define a norm for a polyhedral cell and then extend the definition additively to polyhedral chains with a non-overlapping representative. The mass of a polyhedral cell $|\sigma|_m$ stands for its volume; consequently, we must define the volume, or rather the measure, of a cell. We begin by defining a measure for the manifold and its submanifolds.

We can define a measure on a differentiable manifold in innumerable ways, but the measure becomes unique in a semi-Riemannian manifold, if we require that the measure be compatible with the metric. The situation is identical to the Lebesque measure of $\mathbb{R}^n$, which is required to give a unit value to a cube with unit length edges. The connection with the metric makes the measure also isotropic; i.e., loosely speaking, a set and its Lorentz transformation have the same measure. Consequently, the isotropy of the measure requires that every local physical observer detects the same measure for any cell; i.e., the measure is independent of the local chart used. In fact, this independence is compulsory for the model of a real object. We will later return to isotropy.

Let χ be a chart around x in a semi-Riemannian manifold M, and let $e_1, \ldots, e_n$ be the standard basis (cf. section 4.1.3) of $T^x M$; i.e., each e_i is an image of the basis vector of $\mathbb{R}^n$ by the derivative of the chart $\mathrm{d}\chi$, and the basis $\mathbb{R}^n$ is orthonormal, as defined in section 4.1.3. The above-mentioned unique *canonical measure on* M is defined as follows:

$$\mu_M \colon \mathcal{B} \to \mathbb{R} : \quad E \mapsto \int_{\chi(E)} \left| \det \begin{bmatrix} \langle e_1, e_1 \rangle & \langle e_2, e_1 \rangle & \ldots & \langle e_p, e_1 \rangle \\ \langle e_1, e_2 \rangle & \langle e_2, e_2 \rangle & \ldots & \langle e_p, e_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle e_1, e_p \rangle & \langle e_2, e_p \rangle & \ldots & \langle e_p, e_p \rangle \end{bmatrix} \right|^{\frac{1}{2}} d\mu_{\mathbb{R}^n}, \quad (4.14)$$

where $\mu_{\mathbb{R}^n}$ is the Lebesque measure on $\mathbb{R}^n$, and B is the set of the μ_M-measurable sets in M. The μ_M-measurable sets match the Lebesque measurable sets of $\mathbb{R}^n$. The integral in the definition is the Lebesque integral in $\mathbb{R}^n$. This definition is consistent, because it does not depend on the chart used. In fact, the determinant is the determinant of the metric tensor, and the canonical measure defines a unique *volume form*[21] ω_μ on an orientable

[21] Volume form is a machinery that takes in n tangent vectors and returns the (oriented)

manifold M. In Euclidian and semi-Euclidian spaces, the above determinant returns a unit value, and the canonical measure is reduced to the Lebesque measure. [24, 38, 53]

Also the above definition can be extended to a measure of a—possibly lower dimensional—submanifold P of M. Let p be the dimension of P. Then the Lebesque measure of $\mathbb{R}^n$ is replaced with that of $\mathbb{R}^p$, and $e_1, \ldots, e_p$ plays the part of the standard basis of $T^x P$ (cf. section 4.1.3). Further, the scalar product is replaced with the restriction of the scalar product on the submanifold, i.e., with $\langle \cdot, \cdot \rangle^P$. With these changes in relation (4.14), we define the canonical measure of P, denoted by μ_P. We can now reach the mass norm of cells and chains in every dimension.

The above definition is problematic only with lightlike submanifolds, whose measure now vanishes, because the restriction of the metric on P is degenerate. So the volume of a cell, which is nondegenerate but lightlike, is zero. This is not a good starting point for a chain norm and creates awkward properties for chains.

So why do we adhere to this measure? The simple answer is: it is isotropic. *Isotropy* means that the measure is invariant under a group of symmetry mappings called local isometries. A *local isometry* in $x \in M$ is a diffeomorphism $f : M \to M$ such that $\langle v, v \rangle_x = \langle f_* v, f_* v \rangle_x$ for each $v \in T^x M$. We also demand that $f(x) = x$. The mapping $f_* : T^x M \to T^x M$ is the pushforward of the isometry f. A group of local isometries is the (local) Lorentz group used in special relativity, i.e., a group that consist of all (local) Lorentz transformations. The analogy in the Riemannian manifold is the group of local rotations around a certain point x. The volume form ω_μ is invariant under the group of these local isometries; i.e., $\omega_\mu(f_* v_1, \ldots, f_* v_p) = \omega_\mu(v_1, \ldots, v_p)$ for each $v_i \in T^x M$ $(i = 1, \ldots, p)$ and each $x \in M$. Thus for the canonical measure, a set U and its image $f(U)$ under a local Lorentz transformation have identical measures; i.e., every local physical observer determines the same measure or volume for an arbitrary set. Equivalently, the Lebesque measure in a Riemannian manifold is (locally) rotation symmetric. Our canonical measure is the only measure that is both isotropic and compatible with the metric.

Finally, we can define the mass of a polyhedral cell. The mass of a polyhedral p-cell σ is the volume of the cell; i.e., $|\sigma|_m = \int_\sigma d\mu_P$, where P is the p-dimensional submanifold where σ lies, possibly the whole M.

---

volume of the infinitesimal room that the vectors span. It is a multilinear real valued function, and later we will name this object the n-dimensional differential form and define an integration theory for it. The measure of a measurable set U can be written $\mu(U) = |\int_U \omega_\mu|$.

Norms of chains

The definition of the mass norm can be extended into the space of polyhedral p-chains. Let us represent a p-chain Σ by a non-overlapping simplicial representative $\Sigma = \sum_{i=1}^{N} \alpha_i \sigma_i$, and let us define

$$|\Sigma|_m = \left| \sum_{i=1}^{N} \alpha_i \sigma_i \right|_m = \sum_{i=1}^{N} |\alpha_i| |\sigma_i|_m. \tag{4.15}$$

The above definition is independent of the choice of the (non-overlapping) representative of Σ. The mass of a polyhedral p-chain, or the p-mass for short, satisfies the axioms of a semi-norm but is not a real norm for the chains of the Lorentzian manifold. Moreover, it is somewhat unintuitive and thus rarely useful as such. First, the boundary operator is not continuous with the mass norm, a very unfavorable property. Second, if we, e.g., subtract a segment from a parallel disjoint segment, the mass norm of the result equals their combined volume regardless of their separation. For our purposes, the norm should converge to zero when the segments approach each other. Therefore, more elaborate (semi-)norms are normally used. [91, 29, 73]

With the aid of the boundary operator and mass norm, we can define the so-called *flat norm* for the chain space as follows [91]

$$|\Sigma|_\flat = \inf_{\Gamma \in C_{p+1}(M)} \{ |\Gamma|_m + |\Sigma - \partial\Gamma|_m \}. \tag{4.16}$$

Like the mass norm, the flat norm is also a semi-norm for the chains of the Lorentzian manifold. It cannot be a real norm, because it vanishes in the chains lying on a lightlike submanifold. Even though the definition looks cumbersome, it follows better our above-described intuition. In particular, the norm is constructed so that the boundary operator is continuous [91, 73].

The biggest problem with the above chain norms is that they give zero values for chains lying in a lightlike submanifold. Thus their continuous duals, cochains, must also vanish on these zero-normed chains. Because there is no reason for electromagnetic fields to vanish on lightlike manifolds, the cochains of this type cannot model electromagnetic quantities. More practically, these zero measured cells would cause division by zero problems in the constitutive relations of our numerical algorithms.

Not only lightlike submanifolds but also the cells near (by the norm) them cause problems. Because cochains are continuous, and because norms force cochains to vanish in any lightlike submanifold, this type cochains continuously approach zero in the neighborhood of a lightlike submanifold. This is again unphysical behaviour for electromagnetic quantities. Consequently, the

problem is not solved by forbidding cells from lying on any lightlike submanifold, which could technically be done with relative chain or cochain spaces. These problems disappear, if we consider only the chains that lie far enough from any lightlike submanifold.

Consequently, we must (at least partially) abandon these norms and look for a topology that does not directly originate from these norms. The situation is not unique, because we, in fact, did the same with semi-Riemannian manifolds. There the topology of the manifold is inherited from the structure of the topological manifold, and the Lorentzian metric is an additional structure that does not agree with the topology in a classical way. Thus in the Lorentzian manifold, we must maintain some separation between topology and metric structures.

Topology and completeness

For a moment, let us forget the norms of chain spaces introduced above. A chain space is a vector space, but we want it to be also a topological space. A structure that combines both of these aspects is the topological vector space.

Definition 4.15. *The topological vector space* is a vector space V over $\mathbb{R}$ and a Hausdorff topology $\mathcal{T}$ on V so that the vector addition ($V \times V \to V$) and the scalar multiplication ($\mathbb{R} \times V \to V$) are continuous functions. $\mathbb{R}$ is assumed to have its default topology, and, further, $V \times V$ and $\mathbb{R} \times V$ are equipped with a product topology. The topological vector space is denoted by $(V, \mathcal{T})$, or V for short, if there is no ambiguity about the topology. [26, 61]

In fact, with a fixed $v_0 \in V$ and a fixed $\lambda_0 \in \mathbb{R}$, the vector addition $v + v_0$ and scalar multiplication $\lambda_0 v$ are not only continuous but even homeomorphisms as a function of $v \in \mathbb{R}$. Consequently, we need only to build a base for the topology around the origin and then tranport the topology everywhere with these homeomorphisms.

Accordingly, we need to introduce a proper topology to the space of the polyhedral chains of each dimension. But what, for our purposes, would be the proper conditions for the topology?

First, we require that the boundary operator to be continuous with these topologies. Second, we want also the chain space to be complete. However, completeness is not a topological property but that of the metric—or that of some other additional structure. We can see this, e.g., in that a metrically complete space can be homeomorphic to an incomplete space. Fortunately, a topological vector space naturally retains a *uniform structure*, which allows us to speak about completeness. A space with a uniform structure is called

a *uniform space* (for a detailed definition of the uniform structure and how a topological vector space has a canonical uniform structure, see [26, 61]).

Loosely speaking, a uniform space is a space of structural complexity between a topological space and a metric space. For instance, every metric space is a uniform space. A topology allows us to speak about neighborhoods, and a metric gives us numerical values for the sizes of neighborhoods. In a uniform space, we have neighborhoods together with a certain ability to compare them with regard to size, but with no precise numerical measure of that size. [26]

With a uniform structure, we can complete a topological vector space [26]. The result of the completion of a polyhedral chain space is the space of *flat chains* $C_p^\flat(M)$ (or the space of outer oriented flat chains $\tilde{C}_p^\flat(M)$), and we can extend the boundary operator to that space and keep it continuous.

However, if we stay apart from lightlike submanifolds, the flat norm becomes a real norm, and the topology it introduces will be the right one for us. Thus we need only to choose a norm-independent topology for the chain space, if the chains are defined on lighlike manifolds or near them.

Open question 4.1. We leave open the question of fixing the chain topology and thus the final definition of flat forms. As mentioned above, we need these only if we want to consider chains on or near a lightlike submanifold. Thus, in fact, the norm topology is enough for our applications in this work.

On several occasions, we need chains that lie on a submanifold P of M. We define them as a subspace of polyhedral chains that covers a subset of P. However, the cells of a general representative can lie outside P; only the cells of a non-overlapping presentative are quaranteed to remain entirely on P. Typically, the dimension of P is lower than $\dim(M) = n$. Strictly speaking, the obvious choice for this space, $C_p^\flat(P)$, is not exactly what we need, because $C_p^\flat(P)$ is not a subspace of $C_p^\flat(M)$. Nevertheless, there is a natural *inclusion* $i : C_p^\flat(P) \to C_p^\flat(M)$ so that $i(C_p^\flat(P)) \subset C_p^\flat(M)$. This subspace $i(C_p^\flat(P))$ is our choice, but to simplify notation we denote it by $C_p^\flat(P)$. By this choice $C_p^\flat(P) \subset C_p^\flat(M)$. In addition, the outer orientation of its outer oriented counterpart $\tilde{C}_p^\flat(P)$ refers to the orientation of the whole M, not to the orientation of P.

4.3.5 Chain complex

Chains of different dimensions together with boundary operators constitute a structure called the chain complex. The chain complex includes all the information we need for a geometric model of electromagnetics. If our manifold

M is n-dimensional, the *chain complex* $C_\times(M)$ is the sequence

$$0 \to C_n(M) \xrightarrow{\partial_n} C_{n-1}(M) \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_1} C_0(M) \to 0, \tag{4.17}$$

where ∂_p is the boundary operator of p-chains, $\partial_p : C_p(M) \to C_{p-1}(M)$.

The $(p-1)$-chains that are images of the boundary operator ∂_p of p-chains are called *boundaries*. They form the space of $(p-1)$-boundaries,

$$\mathcal{B}_{p-1}(M) = \mathrm{im}(\partial_p) = \{\partial_p \Sigma \,|\, \Sigma \in C_p(M)\}. \tag{4.18}$$

The p-chains whose boundary is zero are called *p-cycles*. All p-cycles form the space of p-cycles,

$$\mathcal{Z}_p(M) = \ker(\partial_p) = \{\Sigma \in C_p(M) \,|\, \partial_p \Sigma = 0\}. \tag{4.19}$$

Because the boundary of a boundary is empty, i.e., $\partial_{p-1}\partial_p \Sigma = 0, \forall \Sigma \in C_p(M)$ and $\forall p > 0$, all boundaries are cycles, i.e., $\mathcal{B}_p(M) \subset \mathcal{Z}_p(M)$ [91]. In terms of chain complex (4.17), the image of ∂_p falls inside the kernel of ∂_{p-1}. The reverse is not generally true, because there are cycles that are boundaries of no chain, the so-called *nonbounding cycles*. However, for the sake of simplicity, we assume that there is no nonbounding p-cycles for $p > 0$.[22] Thus $\mathcal{B}_p(M) = \mathcal{Z}_p(M)$ $(p > 0)$ and sequence (4.17) is said to be *exact*. A chain complex can also be defined for flat chains, and the result is a *flat chain complex*; for outer oriented chains, we get an *outer oriented chain complex*.

4.3.6 Finite chain complex

For numerical computation, we want to have some finite counterpart of the chain complex. For numerical modeling, we cannot manage all the possible cells—and thus all the chains—of M, but we can use the finite subspaces of chain spaces. To restrict the chain spaces, we confine ourselves to considering only a certain subset of polyhedral cells. These cells of a particular dimension form the finite sets $S_0, \dots, S_n$. We require that the cells included in these sets fulfil the following conditions:

1. each S_p is composed of mutually non-overlapping polyhedral cells.
2. the union of the n-cells fills the whole manifold, i.e., $\bigcup_{c \in S_n} c = M$.

[22] Actually, with only minor extra work, everything in this thesis could be generalized into the case of topologically non-trivial domains. Thus, this restriction is quite artificial.

3. the faces of each cell of S_p are included in the set S_{p-1}.

4. the interior of any cell in S_p does not intersect lower dimensional cells in the sets S_q, $q < p$.

5. for $p = 0, \ldots, n$, for each $\sigma_1, \sigma_2 \in S_p$ so that $\sigma_1 \cap \sigma_2 \neq \emptyset$, we have $\sigma_1 \cap \sigma_2 \in S_{p-1}$.

The finite sets $S_0, \ldots, S_n$ are together called the *cellular mesh complex* K. The *finite chain complex* $C_\times(K)$ is the chain complex constructed from the formal sums of sets $S_0, \ldots, S_n$, and these finite chain spaces are denoted $C_p(K)$ for each $p = 0, \ldots, n$. Loosely speaking, the space $C_p(K)$ is a subspace of the flat chain space $C_p^\flat(M)$. Strictly speaking, the flat chain space consists of Cauchy sequences of polyhedral chains; thus we need an inclusion $i : C_p(K) \to C_p^\flat(M)$ so that $i(C_p(K)) \subset C_p^\flat(M)$. The inclusion is an isomorphism onto $i(C_p(K))$ and, additionally, a homeomorphism. However, for brevity, we drop the inclusion map from the notation. According to condition 3, the boundary operator can also be restricted to these finite chain spaces so that $\partial : C_p(K) \to C_{p-1}(K)$.

Because the chain spaces $C_p(K)$ are finite dimensional for every $p = 0, \ldots, n$, we can build a basis for each space $C_p(K)$. The p-cells in the finite set S_p of the cellular mesh complex K form a standard basis for the chain space $C_p(K)$. Then each $\Sigma \in C_p(K)$ can be written in the form

$$\Sigma = \sum_{\sigma \in S_p} \alpha(\sigma)\sigma, \tag{4.20}$$

where $\alpha(\sigma)$ is a real coefficient of the basis cell σ. The above representation of the finite chain is, in fact, unique and makes its numerical presentation elementary and the finite chain space truely usable in computation. If the cells are enumerated by I_p, i.e., $S_p = \{\sigma_i^p \mid i \in I_p\}$, we can write $\Sigma = \sum_{i \in I_p} \alpha_i \sigma_i^p$. If we choose a basis for each $C_p(K)$ with $p = 0, \ldots, n$, each $\Sigma \in C_p(K)$ can be represented by the real array $\boldsymbol{\Sigma} \in \mathbb{R}^{\kappa_p}$ with $[\boldsymbol{\Sigma}]_i = \alpha_i$, where κ_p is the cardinality of S_p.

The boundary operator of finite chain spaces $\partial : C_p(K) \to C_{p-1}(K)$ can now be represented with a finite linear map between these coefficient arrays, which represent the chains. Thus we have a matrix $\boldsymbol{\partial}_p \in \mathbb{R}^{\kappa_{p-1} \times \kappa_p}$ so that $\partial\Sigma$ is represented by the coefficient array $\boldsymbol{\partial}_p \boldsymbol{\Sigma}$. Hence $\partial\Sigma = \sum_{i \in I_{p-1}} [\boldsymbol{\partial}_p \boldsymbol{\Sigma}]_i \, \sigma_i^{p-1}$. If we use the standard basis of $C_p(K)$, i.e., the one composed of the p-cells of K, $\boldsymbol{\partial}_p$ contains only values -1, 0, and 1. $[\boldsymbol{\partial}_p]_{ij} = \pm 1$, if the i^{th} basis $(p-1)$-cell lies on the boundary of the j^{th} basis p-cell; it is positive, if the orientations match; otherwise it is negative. The matrix $\boldsymbol{\partial}_p$ is called the *incidence matrix*.

There is also a corresponding *outer oriented cellular mesh complex* $\tilde{K}$ build form a finite number of outer oriented cells and an *outer oriented finite chain complex* $\tilde{C}_\times(\tilde{K})$ built over the above mesh complex. Similarly, the *finite boundaries* $\mathcal{B}_p(K)$ are images of the boundary operator ∂ of $(p+1)$-chains in $C_{p+1}(K)$, and the *finite cycles* $\mathcal{Z}_p(K)$ are the boundaryless chains in $C_p(K)$.

4.3.7 Extrusion of chains

Our aim is to do electromagnetic modeling in the spacetime and use chains to model four-dimensional macroscopic geometry. However, because on some occasions we want to compare our model with the well-known three-dimensional one, we need a way to connect three-dimensional purely spatial chains with some four-dimensional ones. Consequently, we need a technique to stretch spatial chains into the temporal dimension and thereby obtain a four-dimensional spacetime chain. This procedure is called extrusion.

Let us begin with a p-dimensional convex cell c in $\mathbb{R}^n$. For an arbitrary vector $\boldsymbol{v} \in \mathbb{R}^n$ and a real parameter $t \in I = [0,1]$, we define the mapping

$$\mathcal{D}_{\boldsymbol{v}}(t \times \cdot) : \mathbb{R}^n \to \mathbb{R}^n : \boldsymbol{v} \mapsto \boldsymbol{x} + t\boldsymbol{v}. \tag{4.21}$$

For an arbitrary value of t, $\mathcal{D}_{\boldsymbol{v}}(t \times c)$ is the convex cell c moved in the direction of $\boldsymbol{v}$ according to the parameter t. Moreover, $\mathcal{D}_{\boldsymbol{v}}(I \times c)$ is a convex cell of dimension $p+1$ or less. We call $\mathcal{D}_{\boldsymbol{v}}(I \times c)$ the *extrusion* of c in the direction of $\boldsymbol{v}$. If $\boldsymbol{v}$ determines the orientation of the extrusion direction and $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$ orientates c, the orientation of $\mathcal{D}_{\boldsymbol{v}}(I \times c)$ is defined to be $(\boldsymbol{v}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$ [91]. Accordingly, its boundary is

$$\partial\mathcal{D}_{\boldsymbol{v}}(I \times c) = \mathcal{D}_{\boldsymbol{v}}(1 \times c) - \mathcal{D}_{\boldsymbol{v}}(0 \times c) - \mathcal{D}_{\boldsymbol{v}}(I \times \partial c). \tag{4.22}$$

The vector $\boldsymbol{v}$ does not have to be constant, because it can change from point to point in c; i.e., it can be a vector field in $c \subset \mathbb{R}^n$, i.e., $\boldsymbol{v} : c \to T\mathbb{R}^n$. However, we must make sure that $\mathcal{D}_{\boldsymbol{v}}(I \times c)$ remains a convex cell. If it does, everything else in the above case remains except that in definition (4.21) we have $\boldsymbol{v} \mapsto \boldsymbol{x} + t\boldsymbol{v}_{\boldsymbol{x}}$.

Via a chart χ, we can carry convex cells c and $\mathcal{D}_{\boldsymbol{v}}(I \times c)$ into a manifold as inner oriented polyhedral cells $\sigma = (\chi, c)$ and $(\chi, \mathcal{D}_{\boldsymbol{v}}(I \times c))$. The real translation vector $\boldsymbol{v}$ transforms into an inner oriented curve segment $\alpha : [0,1] \to M$. Thus with polyhedral cells, extrusion is made with respect to a vector field $v : M \to TM$—along the above described curve α we have $v_x = \alpha'(t)$ for each $x = \alpha(t)$.

If we want to define the extrusion of σ along an arbitrary vector field v, we must choose the chart χ so that there is a vector $\boldsymbol{v} : c \to T\mathbb{R}^n$, which via this chart maps to the integral curve of v near σ. With this chart, σ is represented by (χ, c). We denote by $\mathrm{extr}(\sigma, v, t)$ the equivalence class of polyhedral cells represented by $(\chi, \mathcal{D}_{\boldsymbol{v}}([0, t] \times c))$ and call it the extrusion of σ along v. The real parameter $t \geq 0$ tells us how far along v the cell is extruded. Yet again, the definition does not depend on the choice of the chart χ—satisfying the above requirements. The orientations of σ and $\mathrm{extr}(\sigma, v, t)$ are inherited from $\mathbb{R}^n$.

The only difficulty in the above definition is to find the proper chart. Does such a chart always even exist? Because the integral curve of a smooth vector field always creates a one-dimensional foliation, by the definition of foliation there does exist such a chart [39], hence extrusion exists at least for a smooth vector field. The smoothness of the vector field requires a smooth manifold. The assumption of manifold smoothness is vital here, because it makes the manifold sleeker in the local perspective.

Our main aim is to extrude in the time direction. Extrusion is thus taken along the velocity field u of the chosen observer so that u agrees with the positive time orientation. In a general case, u does not have to be smooth, but we assume that it is smooth near the cell under extrusion so that extrusion in the time direction is always possible.

Thus we can extrude polyhedral cells along vector fields (see figure 4.3). We can also extrude polyhedral chains by extruding cell-wise one representative of the chain. This extrusion is independent of representative. Similarly, the definition can also be extended to flat chains. The extrusion of a flat chain Σ along v is denoted $\mathrm{extr}(\Sigma, v, t)$.

Analogously to the above, we can also define the transportation of a chain along v. For a polyhedral cell σ, we denote $\Psi_t(\sigma) = (\chi, \mathcal{D}_{\boldsymbol{v}}(t \times c))$. This transports σ along the integral curve of v according to the argument $t \geq 0$. Yet again, the definition generalizes into the flat chains. According to equation (4.22), with the aid of Φ-transportation, the boundary of a chain can be written (compare to figure 4.3)

$$\partial \mathrm{extr}(\Sigma, v, t) = \Psi_t(\Sigma) - \Sigma - \mathrm{extr}(\partial\Sigma, v, t). \tag{4.23}$$

Note that with this notation, $\Sigma = \Psi_0(\Sigma)$. [44, 91]

Now we have the tools to extrude purely spatial chains in the time direction. But what do the terms "purely spatial" and "time direction" mean? First, both concepts are observer-dependent entities. And as we already mentioned, the time direction means that extrusion is made in the direction of the velocity field u of the chosen observer. This velocity field is everywhere

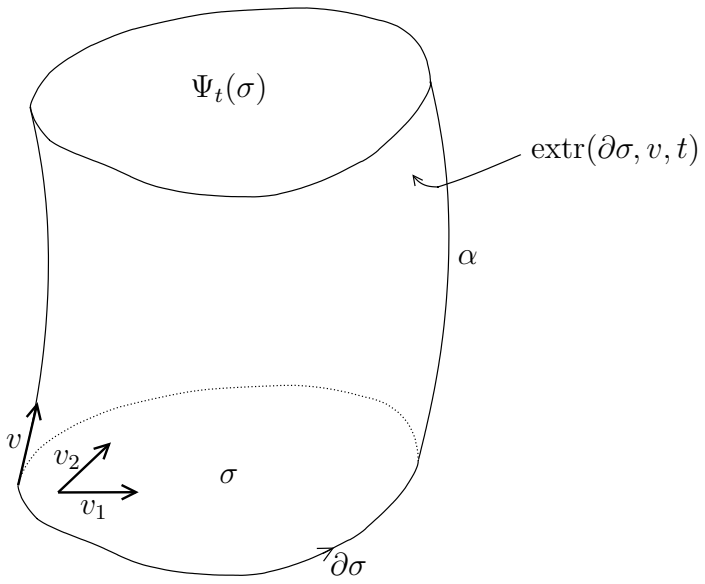


Figure 4.3: Extrusion of a polyhedral cell σ along a vector field v according to a real parameter t is denoted $\mathrm{extr}(\sigma, v, t)$. σ is here a 2-cell, and the extruded cell is a three-dimensional volume. α is an integral curve of the vector field v. The cell σ is oriented by (v_1, v_2) and $\mathrm{extr}(\sigma, v, t)$ by (v, v_1, v_2). The boundary of $\mathrm{extr}(\sigma, v, t)$ consist of σ at the bottom, $\Psi_t(\sigma)$ on the top, and $\mathrm{extr}(\partial\sigma, v, t)$ on the side surface. The first and the last come with a minus sign, as seen by comparing the orientations of the boundary and the extrusions [91].

tangent to the worldlines of the chosen observer. As discussed in the section 4.2, there are also $(n-1)$-dimensional submanifolds—called foliations—which consist of simultaneous events in the point of view of the observer under consideration. They are parametriced by the proper time of the observer τ and denoted by M_τ. The tangent space $T^x M_\tau$ is Lorentzian orthogonal to the velocity field u. The cells lying on this (purely spatial) submanifold M_τ are said to be *purely spatial*, and also chains construed from the formal sums of purely spatial cells are said to be purely spatial.

With almost identically to the above, also outer oriented chains can be extruded. Now the chain under the extrusion $\tilde{\Sigma}$ is outer oriented. If O_r describes the global manifold orientation, $\tilde{\Sigma}$ has the representative (Σ, O_r), where Σ is a proper inner oriented chain. Now, we define $\mathrm{extr}(\tilde{\Sigma}, v, t)$ to be the chain represented by $(\mathrm{extr}(\Sigma, v, t), O_r)$. Accordingly, the boundary of the extruded chain is similar to (4.23) with the exception that inner oriented chains are replaced by outer oriented ones.

The outer orientation of a p-chain $\tilde{\Sigma}$ in a submanifold M_τ can be defined relative to the global orientation of M or M_τ. In case of confusion, we use the notation $\tilde{\Sigma}(M)$ and $\tilde{\Sigma}(M_\tau)$, respectively. In the above definition of extrusion, all outer orientations are relative to the whole M. If M_τ is purely spatial and M is oriented and time-oriented, these orientations also define a canonical orientation for M_τ. If (e_0, e_1, e_2, e_3) is the local orientation of M, and if e_0 describes the local time orientation, M_τ is locally oriented by (e_1, e_2, e_3). Thus we also have a canonical one-to-one connection between $\tilde{\Sigma}(M)$ and $\tilde{\Sigma}(M_\tau)$.

Whitney [91] has shown that a submanifold can be approximated by a chain as accurately as possible.[23] In this way, we can also find a procedure to extrude submanifolds along a vector field.

4.4 Cochains

We must now decide what kind of mathematical objects we wish to use in our modeling of physical electromagnetic quantities. Because we aim at numerical computation, we choose models that are suitable for finite dimensional approximations.

Let us return to the question, what kind of quantitative electromagnetic phenomena can we detect in nature? What kind of electromagnetic quantities can we measure? Strictly speaking, we can never measure the pointwise values of classical electromagnetic quantities. For example, we cannot detect the pointwise value of an electric charge, but only the net charge in an observable small volume. In fact, in nature a charge is always quantized into the entities of a finite charge and volume; thus pointwise values are in a way unphysical. Other examples of difficult pointwise measurements are electric current and magnetic flux density, which are naturally measured by their fluxes through finite areas. However, we can also gain an understanding of pointwise measures by examining the net charge inside a series of ever decreasing volumes or the flux through a series of decreasing areas. The pointwise value of a quantity is the limiting value of the net charge or flux.

Consequently, are macroscopic quantities—such as net charge or fluxes—or pointwise quantities right for our purposes? The answer is that both are necessary for electromagnetic computation. Macroscopic quantities are the best for representing the Maxwell laws, whereas local quantities are suited for constitutive laws. We decide to model macroscopic quantities with cochains

[23] This proof applies here only, if we the have a flat norm for chains, not only a uniform structure. Our hypothesis is that this result holds also for flat chains with a uniform structure.

and local quantities with differential forms. Because we use both in essential parts of modeling electromagnetic phenomena, we need a connection between cochains and differential forms. In the end, we find some spaces of cochains and differential forms that are homeomorphic with each other.

Now we introduce a model for macroscopic quantities and return to local quantities in the following sections. We model macroscopic electromagnetic quantities with flat cochains and represent the Maxwell laws with the aid of the exterior operator of cochains. To model electromagnetic quantities that are associated with outer oriented chains, we need twisted cochains.

4.4.1 Flat cochains

The cochain is a linear and continuous mapping from the linear space of chains into real numbers; i.e., it belongs to the dual space of chains. This also suggests the background of the term cochain. A cochain is a mapping of the form

$$\Omega : C_p(M) \to \mathbb{R} : \Sigma \mapsto (\Omega \,|\, \Sigma), \tag{4.24}$$

also called a p-cochain.

Example 4.2. As we mentioned before, the integral of the electric field over a chain is a 1-cochain. The cochain is denoted by

$$\mathcal{E} : C_1(M) \to \mathbb{R} : c \mapsto \int_c \mathscr{e}, \tag{4.25}$$

where $\mathscr{e}$ is a—yet to be defined—differential form of the electric field. The integral over a 1-chain c is discussed later; here the above is only a notation representing the reader's intuition of an integral of the electric field. Similarly, the integral of magnetic flux density can be modeled as a 2-cochain.

As the dual space of a linear space, the linearity of a cochain is a straightforward property. Continuity is a property determined by the topology—or by the norm if it agrees with the topology—of the chain space and thus depends on the topology or norm used. For example, we have the *space of the polyhedral cochains* $C^p(M)$, with a mass norm, and the *space of flat cochains* $C^p_\flat(M)$, using the topology of flat chains in the corresponding chain space. Flat cochains are crucial for us, because they constitute the largest space of the cochains, which is homeomorphic with a subspace of the differential forms.

Moreover, we define a linear space structure for flat chains; consequently, $C^p_\flat(M)$ forms a linear space. If Ω and Λ are arbitrary cochains and $\alpha, \beta \in \mathbb{R}$,

their linear combination is defined as

$$(\alpha\Omega + \beta\Lambda) : \ C_p(M) \to \mathbb{R} : \ \Sigma \mapsto \alpha(\Omega \,|\, \Sigma) + \beta(\Lambda \,|\, \Sigma). \tag{4.26}$$

In addition, if the corresponding flat chain space possesses a norm, we can define a norm for flat chains, and thus $C^p_\flat(M)$ is a linear norm space. The norm of a flat cochain is defined as

$$|\Omega|_\flat = \inf\{\eta \in \mathbb{R} \mid \forall \Sigma \in C^\flat_p(M); \ |(\Omega \,|\, \Sigma)| \leq \eta |\Sigma|_\flat\}. \tag{4.27}$$

Observe that this agrees with the standard way of defining a norm for the metric dual space or for any other operator space.

If there is no real norm to give us topology, as in our case, topology must be defined separately. We assume that the chains space $C_p(M)$ is equipped with a topology. We can define several possible topologies for cochain spaces, e.g., the weak-$*$-topology, which is the coarsest topology for the dual pairing $C^p(M) \times C_p(M) \to \mathbb{R} : (\Omega, \Sigma) \mapsto (\Omega \,|\, \Sigma)$ to be continuous. The Cartesian product is assumed to be equipped with a product topology. Consequently, the cochain space turns into a topological vector space, and we can then complete this space. Thus we can assume that the space of the flat cochains $C^p_\flat(M)$ is complete. Furthermore, we require that the topology of the flat cochain space be such that the complete spaces $C^\flat_p(M)$ and $C^p_\flat(M)$ are dual to each other. This is always possible, at least with the chain and cochain topologies inherited from the Riemannian metric and from the corresponding measure and cochain flat norm.

Once again, we are free to use the above cochain flat norm, if we restrict our chain spaces to lie far enough away from lightlike submanifolds. For a submanifold P, the dual space of $C^\flat_p(P)$ is denoted by $C^p_\flat(P)$. We can define a *trace* operator $t_P : C^p_\flat(M) \to C^p_\flat(P) : \Omega \mapsto \Omega_P$ so that $\Omega(\Sigma) = \Omega_P(\Sigma)$ for each $\Sigma \in C^\flat_p(P)$; i.e., the trace operator restricts the cochain into a submanifold. Additionally, $t_P(C^p_\flat(M)) \subset C^p_\flat(P)$.

Open question 4.2. As with chains, we do not fix cochain topology in the lightlike case, because we can restrict ourselves to the normed case in our applications. We leave open the choice of a proper cochain topology in the general case.

4.4.2 Exterior operator

The *exterior operator* (or the *coboundary*) is an operator of cochains that correspond the derivatives of vector fields or differential forms. It is an entity that is similar to the boundary operator of chains and is, in fact, defined with the aid of the boundary operator.

The definition of the exterior operator stems from *Stokes' theorem* of the differential forms,

$$\int_{\Sigma} \mathsf{d}\omega = \int_{\partial\Sigma} \omega \qquad \forall \Sigma \in C_p^\flat(M) \;\text{ and }\; \forall\, p\text{-form } \omega. \tag{4.28}$$

The notation of the theorem is not yet defined. For now, we need only to know that the integral of the exterior derivative of the differential form $\mathsf{d}\omega$ over a chain Σ equals the integral of the original differential form over the boundary $\partial\Sigma$.

Based on the above, we define the exterior operator to be the linear mapping

$$\begin{aligned} \mathrm{d}: &\quad C_\flat^p(M) \to C_\flat^{p+1}(M) : \Omega \mapsto \mathrm{d}\Omega \\ \text{so that} \quad \mathrm{d}\Omega: &\quad C_{p+1}^\flat(M) \to \mathbb{R} : \Sigma \mapsto (\Omega \,|\, \partial\Sigma). \end{aligned} \tag{4.29}$$

Because $\partial\partial\Sigma = 0$ for each $\Sigma \in C_{p+1}(M)$, also $\mathrm{dd}\Omega = 0$ for all $\Omega \in C_\flat^p(M)$.

4.4.3 Twisted cochains

We have defined p-cochains as objects that send an inner oriented p-chain to a real number. But what if the chain is outer oriented? Then we must resort to a *twisted p-cochain* [47, 48].

A twisted p-cochain is an equivalence class of the pairs (Ω, O_r), where $\Omega \in C_\flat^p(M)$, and O_r is the global orientation of the manifold. (Ω, O_r) and (Ω', O_r') are equivalent if $\Omega' = \mathrm{ori}(O_r, O_r')\,\Omega$. The action of a twisted cochain (Ω, O_r) on an outer oriented chain (Σ, O_r') is defined as $((\Omega, O_r) \,|\, (\Sigma, O_r')) = \mathrm{ori}(O_r, O_r')(\Omega \,|\, \Sigma)$.

The linear structure, topology, and possible norm of twisted cochains are directly obtained by using straight versions for the straight representative cochain. We denote the space of flat twisted cochains as $\tilde{C}_\flat^p(M)$ (and in the P-submanifold case $\tilde{C}_\flat^p(P)$). The exterior operator of a twisted cochain is defined to be the equivalence class of the straight representative's exterior operator.

4.4.4 Cochain complex

Similarly to the chain complex, cochains of different dimensions together with boundary operators constitute a cochain complex. The *cochain complex* $C^\times(M)$ in a n-dimensional manifold M is the exact sequence

$$0 \xleftarrow{\mathrm{d}_n} C^n(M) \xleftarrow{\mathrm{d}_{n-1}} C^{n-1}(M) \xleftarrow{\mathrm{d}_{n-2}} \ldots \xleftarrow{\mathrm{d}_0} C^0(M) \leftarrow 0, \tag{4.30}$$

where d_p is the exterior operator of the p-chains, $\mathrm{d}_p : C^p(M) \to C^{p+1}(M)$.

The practice of naming cochain complexes is similar to that of naming chains, except for the co-prefix. *Coboundaries* $\mathcal{B}^{p-1}(M)$ are those $(p+1)$-cochains that are the images of the coboundary operator d_p of p-cochains, whereas those p-chains whose coboundary is zero are called *p-cocycles* and form the space $\mathcal{Z}^p(M)$. Because we chose $\mathcal{B}_p(M) = \mathcal{Z}_p(M)$ for $p > 0$, we also have $\mathcal{B}^p(M) = \mathcal{Z}^p(M)$ for $p > 0$.

Additionally, a cochain complex can be similarly defined for flat cochains, the result being a *flat cochain complex*; for twisted cochains, we gain an *twisted cochain complex* $\tilde{C}^\times(M)$.

4.4.5 Finite cochain complex

For numerical computation, we need a finite dimensional cochain complex. The obvious choice for a finite dimensional cochain space is the dual counterpart of the finite chain space $C_p(K)$. A *finite p-cochain* $\Omega \in C^p(K)$ takes a p-chain $\Sigma \in C_p(K)$ and returns a real number. Because the cochain is linear, we need to know its value only on each basis cell of the cell complex K. Thus $C^p(K)$ is spanned by the basis $\{\Omega_i^p\}$ which satisfies

$$(\Omega_i^p \mid \sigma_j^p) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \tag{4.31}$$

for each p-cell σ_j^p in the basis of $C_p(K)$, i.e., $C_p(K) = \mathrm{span}\{\sigma_i^p\}$. The exact sequence of finite cochains is called the *finite cochain complex* and denoted by $C^\times(K)$. Similarly we get the spaces of the *finite twisted p-cochains* $\tilde{C}_p(\tilde{K})$ and the *finite twisted cochain complex* $\tilde{C}^\times(\tilde{K})$.

Identically to the chain case, the choice of the basis for each $C^p(K)$ with $p = 0, \ldots, n$ allows a $\Omega \in C^p(K)$ to be represented by the real array $\mathbf{\Omega} \in \mathbb{R}^{\kappa_p}$. If we choose the above basis, we can write an arbitrary $\Omega \in C^p(K)$ in the form $\Omega = \sum_{i \in I_p} \beta_i \Omega_i^p$, and we have $[\mathbf{\Sigma}]_i = \beta_i$. The dimension κ_p is now the cardinality of $\{\Omega_i^p\}$. If now a finite p-chain Σ and a finite p-cochain Ω are represented by their real array counterparts $\mathbf{\Sigma}$ and $\mathbf{\Omega}$, we have $(\Omega|\Sigma) = (\sum_{i \in I_p} \beta_i \Omega_i^p | \sum_{i \in I_p} \alpha_i \sigma_i^p) = \sum_{i \in I_p} \alpha_i \beta_i = \mathbf{\Omega}^T \mathbf{\Sigma}$ by (4.31).

The exterior operator or coboundary operator resembles closely the chain boundary operator also in the finite dimensional case. If we write the definition of the coboundary operator (4.29) in the coefficient array form, the real array couterpart of d for p-cochains, denoted by $\mathbf{d}_p$, must satisfy $(\mathbf{d}_p \mathbf{\Omega})^T \mathbf{\Sigma} = \mathbf{\Omega}^T (\boldsymbol{\partial}_{p+1} \mathbf{\Sigma})$ for each p-cochain real counterpart $\mathbf{\Omega}$ and $(p+1)$-chain counterpart $\mathbf{\Sigma}$. Thus $\mathbf{d}_p = \boldsymbol{\partial}_{p+1}^T$.

We can define an *interpolator* from a finite cochain space into a flat cochain space, $\mathcal{W} : C^p(K) \to C_\flat^p(M)$ and the reverse operator $\mathcal{R} : C_\flat^p(M) \to$

$C^p(K)$. The range of $\mathcal{W}$ covers only a finite subspace of $C^p_\flat(M)$ denoted by $C^p_K(M)$. Thus $\mathcal{RW} : C^p(K) \to C^p(K)$ is an identity mapping, but $\mathcal{WR} : C^p_\flat(M) \to C^p_\flat(M)$ is an identity only on $C^p_K(M)$.

4.5 Exterior algebra

So far, we have introduced means to model macroscopic geometry by chains; in addition, we have introduced cochains to model electromagnetic quantities related to macroscopic geometric objects. Although these macroscopic objects are suited for modeling the Maxwell laws, we also need a model for pointwise electromagnetic quantities. The constitutive laws, i.e., the interaction between material and electromagnetic fields, are inherently local in nature. Because the interaction can be precisely defined only for pointwise values of electromagnetic quantities, we must formalize local geometry—also known as *virtual geometry*—and the field quantities on these geometric objects. The mathematical hierarchy of the structures is similar to the chain-cochain-structure. First, we define multivectors to model virtual geometry and then the space of multicovectors as their dual space. Multicovector-valued mappings on a manifold—called differential forms—are used as a model for pointwise field quantities. For the constitutive laws, we need a metric-dependent Hodge operator to connect the differential forms of distinct dimensions.[24]

4.5.1 Multivectors

By virtual geometry we mean objects that model infinitesimal lines, surfaces, volumes and so on. Tangent vectors model infinitesimal lines, but we need objects of the same kind also in the other dimensions. With vectors and their dual counterparts, we can only model displacement type objects and quantities associated with displacement, such as electric field intensity, but, e.g., for magnetic flux density and current density, we need a model for infinitesimal surfaces.

Our idea is grounded in that a pair of tangent vectors defines an infinitesimal surface and, more generally, the p-tuple of tangent vectors specifies a p-dimensional infinitesimal geometry. With some structure layers, these p-tuples are called multivectors.

With the aid of tensor algebra, we could define that a p-dimensional multivector is a completely antisymmetric contravariant tensor of rank p.

[24]It is questionable whether the Hodge operator and the inner product should be included in exterior algebra or not. We choose to include them.

Because we need only antisymmetric tensors, we define multivectors directly. We start our definition with an auxiliary entity called the simple multivector and add a linear structure to it with formal sums. With a proper equivalence classification, we obtain general multivectors. Moreover, we define the inner product along with the exterior product. Finally, we discuss the local and global orientations of a manifold and define the outer oriented counterpart of multivectors like we defined the outer oriented chains.

The abstractions of multivectors and their dual counterparts, multicovectors, are identical excluding the fact that tangent vectors are replaced by their dual associates. These dual structures are briefly introduced in the next section. The connection between multivectors and multicovectors with Riesz's representation theorem and the introduction of the Hodge operator for both sides are postponed to the section 4.5.3.

Simple Multivectors

The simple multivector formalizes the abstraction defined by a triplet: a subspace of T^xM, an orientation of the subspace, and a relative volume. *Relative volume* is an equivalence classification of regions; the comparison is made with the metric-free relative volume comparison described in [91].[25] Given a base of T^xM and $\epsilon > 0$, there exists $\delta > 0$ so that the regions R and R' belong to the same class, if the following is true. We cut T^xM into equal cubes of side $h \leq \delta$ along the basis vectors' directions; if N and N' are the number of cubes in R and R', then $|1 - N/N'| < \epsilon$.

Formally, a p-dimensional *simple multivector* is an equivalence class of p-tuples of tangent vectors. The p-tuples $(v_1, \ldots, v_p)$, $(w_1, \ldots, w_p) \in (T^xM)^p$ are equivalent if neither $v_1, \ldots, v_p$ nor $w_1, \ldots, w_p$ span a p-dimensional subspace of T^xM or, if there exists a coefficient matrix $K \in \mathbb{R}^{p \times p}$ so that $v_i = \sum_{j=1}^{p} [K]_{ij} w_j$ and $\det(K) = 1$.

Intuitively, a representative of a simple multivector is an infinitesimal parallelotope (see [16]) covering a specific volume. All parallelotopes on the same hypersurface with the same relative volume are identified. For example, a two-dimensional simple multivector is composed of all the equal-area parallelograms on a fixed plane, whereas a three-dimensional simple multivector is made up of all the equal-volume parallelepipeds on a three-dimensional subspace.

[25] This relative volume is in unison with the canonical measure of M as defined in the section 4.3.4. If the regions belong to the same relative volume class, also their measures agree, and vice versa.

The simple multivector has a causal character ; it is timelike, spacelike, or lightlike according to the causal character of its defining subspace.

General Multivectors

The space of simple multivectors adequately models infinitesimal geometries but does not possess a linear structure. Like for chain-cochain structure, linear structure is necessary for multivectors in order to model electromagnetic fields as linear mappings. As with the polyhedral cell, we add a linear structure to simple multivectors with formal sums and a proper equivalence classification.

Let S_p be the linear space of the formal sums of p-dimensional simple multivectors. We define an equivalence relation to these formal sums by defining a zero equivalence class. The zero class Z_p contains, besides zero formal sums, linear combinations of typical elements:

1. $1(v_1 + u_1, v_2, \ldots, v_p) - 1(v_1, v_2, \ldots, v_p) - 1(u_1, v_2, \ldots, v_p)$
2. $\alpha(v_1, v_2, \ldots, v_p) - 1(\alpha v_1, v_2, \ldots, v_p)$

Now the linear *space of p-dimensional multivectors* $T_p^x M$ is defined as the quotient space S_p/Z_p—its members are called *p-vectors* for short. Equivalence classification makes the space $T_p^x M$ finite dimensional. The linear structure is inherited from the linear structure of formal sums with the above two rules taken into account. By the definition of simple multivectors, the zero class Z_p contains also those tuples where vectors are linearly dependent. This reflects the fact that p dependent vectors do not span a p-dimensional subspace. Consequently, flipping two vectors on a tuple changes the orientation of the infinitesimal geometry that the p-vector stands for. [91, 73]

If $p = 0$, the tuple can only be empty, and the space of 0-vectors can be identified with real numbers. If $p = 1$, the zero class Z_1 is also trivial, and we can identify 1-vectors with tangent vectors; i.e., the spaces $T_1^x M$ and $T^x M$ are identical.

A multivector is *simple*, if it can be written as a tuple $(v_1, \ldots, v_p)$. In an n-dimensional manifold, 0-, 1-, $(n-1)$-, and n-dimensional multivectors are always simple [91]. Accordingly, in standard three-dimensional analysis, no general multivectors are required, but for our four-dimensional study they are a must. For example, let e_1, e_2, e_3, e_4 be a basis for a four-dimensional vector space V. Then $(e_1, e_2) + (e_3, e_4)$ is not simple. Therefore, we must take extra care of the non-simple multivectors. The problem is not insuperable, because simple multivectors constitute a base of multivector space; consequently, we can express a multivector as a linear combination of simple multivectors.

To define a p-vector-valued function, we need a bundle where the multivector spaces $T_p^x M$ for each $x \in M$ belong, such as the tangent bundle TM for 1-vectors. We follow the procedure of defining TM by replacing the spaces $T^x M$ with $T_p^x M$. We obtain a bundle $T_p M = \coprod_{x \in M} T_p^x M$. A p-vector field is a section of the *p-vector bundle* $T_p M$, i.e., a mapping $v : M \to T_p M : x \mapsto (v_x, x)$.

Let M and N be two differentiable manifolds and $\phi : M \to N$ a differentiable map, as in definition 4.4, and let $v \in T_p^x M$ be a simple multivector represented by $(v_1, \ldots, v_p)$. Then the derivative mapping $\mathrm{d}\phi(x, \cdot) : T_1^x M \to T_1^{\phi(x)} N$ generalizes into simple multivectors elementwisely for its representive tuple, i.e., $\mathrm{d}\phi(x, \cdot) : T_p^x M \to T_p^{\phi(x)} N : (v_1, \ldots, v_p) \mapsto (\mathrm{d}\phi(x, v_1), \ldots, \mathrm{d}\phi(x, v_p))$. The case of non-simple multivectors follows from the multilinearity of $\mathrm{d}\phi$. As before, the derivative mapping introduces a *pushforward* $\phi_* : T_p M \to T_p N$ between the bundles $T_p M$ and $T_p N$, which equals $\mathrm{d}\phi(x, \cdot)$ in a particular p-vector space $T_p^x M$.

Scalar Product

In section 4.1.5, we defined a scalar product for tangent vectors $\langle \cdot, \cdot \rangle : T_1^x M \times T_1^x M \to \mathbb{R}$ as a (possibly) indefinite generalization of the inner product. To impose the constitutive laws, we need a scalar product—and especially the concept of orthogonality—for the multivectors.

Because simple multivectors form a base of multivectors, we need only to define a scalar product for simple multivectors. The general case follows from the bilinearity of the scalar product. Let $(v_1, \ldots, v_p)$ and $(w_1, \ldots, w_p)$ be the representatives of two simple multivectors. Then their scalar product is defined to be the determinant

$$\langle (v_1, \ldots, v_p), (w_1, \ldots, w_p) \rangle = \begin{vmatrix} \langle v_1, w_1 \rangle & \langle v_2, w_1 \rangle & \ldots & \langle v_p, w_1 \rangle \\ \langle v_1, w_2 \rangle & \langle v_2, w_2 \rangle & \ldots & \langle v_p, w_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_1, w_p \rangle & \langle v_2, w_p \rangle & \ldots & \langle v_p, w_p \rangle \end{vmatrix}. \quad (4.32)$$

This definition is sound, because the scalar product of two multivectors is independent of the representative of the multivectors. In addition, the above scalar product is by construction symmetric and nondegenerate and thus satisfies the conditions of definition 4.6 [22]. The *norm* of the p-vector $v \in T_p^x M$ is $\|v\| = |\langle v, v \rangle|^{\frac{1}{2}}$.

Exterior Product

The Hodge operator is the fundamental tool for modeling the constitutive laws, and a mapping called the exterior product is needed to define it. The *exterior product* is a bilinear mapping of multivectors of the form

$$\wedge : T_p^x M \times T_q^x M \to T_{p+q}^x M. \tag{4.33}$$

By the bilinearity requirement, it is enough to define an exterior product for simple multivectors. If $(v_1, \ldots, v_p)$ and $(w_1, \ldots, w_p)$ are the representatives of two simple multivectors, their exterior product is defined by $(v_1, \ldots, v_p) \wedge (w_1, \ldots, w_p) = (v_1, \ldots, v_p, w_1, \ldots, w_p)$. Because we can validate that this definition does not depend on the representative of the multivector, it is properly defined for general multivectors.

In a Lorentzian manifold, if e_0 is the local basis vector in the time direction ($\langle e_0, e_0 \rangle = -1$), an arbitrary p-vector that can be written as $v \wedge e_0$ is said to be *in the time direction*, whereas a p-vector lying in M_τ is said to be *purely spatial*.

Global orientation and outer oriented multivectors

For some quantities, we need outer oriented infinitesimal geometries, i.e., outer oriented versions of multivectors. As with chains, the outer oriented version is defined as an equivalence class of pairs of a multivector and an (local) orientation. For twisted differential forms, we also need a characterization of the global orientation O_r.

Because the space on n-vectors, $T_n^x M$, is one-dimensional, it is spanned by one n-vector, σ. Because σ is n-dimensional, it is simple, and only its relative volume and orientations can be chosen freely. By choosing the value of the norm $\|\sigma\|$, we fix the relative volume. We choose $\|\sigma\| = 1$ and call σ a *unit n-vector*. There are only two possible unit n-vectors, namely σ and $-\sigma$; consequently, a unit n-vector can be used to fix the *local orientation* of the tangent space $T^x M$.

Can the (global) orientation O_r of an orientable manifold be described by unit n-vectors? Manifold orientation means describing the orientation of all tangent spaces in a consistent way. Consistency means that all change-of-chart mappings have a positive determinant, as explained in section 4.1.6. We can construct a smooth n-vector field

$$\sigma : M \to T_n M : x \mapsto (\sigma_x, x), \tag{4.34}$$

where σ_x agrees with the orientation of $T^x M$ for each $x \in M$. We call this mapping σ the *global orientation* of M. We need a global orientation to

define the Hodge operator and the twisted differential forms. We can easily see that this definition coincides with that of the manifold orientation O_r and its chain representative Σ_M.

As with chains and cochains, we need outer oriented counterparts of the multivectors and their duals to model certain pointwise electromagnetic field quantities. For one thing, if we change the inner orientation in the definition of simple multivectors to an outer orientation, we get *outer oriented multivectors*. Formally, they are equivalence classes of pairs (v, σ) of multivector v and local orientation σ. (v, σ_1) and (w, σ_2) are equivalent, if $w = v$ and $\sigma_1 = \sigma_2$, or if $w = -v$ and $\sigma_1 = -\sigma_2$. The space of outer oriented p-vectors is denoted by $\tilde{T}^x_p M$.

Scalar and exterior products of outer oriented multivectors are made by their representatives of equal local orientation.

4.5.2 Multicovectors and differential forms

In the above section, we introduced multivectors to model infinitesimal geometries. They were constructed from tuples of tangent vectors with a formal sum structure and an equivalence classification. Now we introduce the dual counterparts of tangent vectors and multivectors, called covectors and multicovectors. In fact, we must define only covectors in depth, because multicovectors are built from covectors analogously to the construction of multivectors from tangent vectors.

Multicovector-valued functions on a manifold, called differential forms, are used to model pointwise field quantities. However, for us, differential forms are only an auxiliary structure to model the constitutive laws. The rest examination is done with cochains; thus the most important aspect of differential forms is their connection with cochains.

First, we define covectors and then briefly revise the description of multicovectors. After that, we can proceed to defining differential forms.

Covectors

The space of *covectors* is the dual space of tangent vectors, denoted by $T^1_x M$. That is, a covector $\omega \in T^1_x M$ is a linear functional from the tangent vector space $T^x_1 M$ into real numbers,

$$\omega : T^x_1 M \to \mathbb{R} : v \mapsto (\omega \,|\, v). \tag{4.35}$$

Additivity and scalar multiplication are defined via the covectors' vector-wise values; hence $T^1_x M$ is a linear space. The linear space $T^1_x M$ is also called the *cotangent space*. [53]

Remark 4.5. We do not require continuity for the covector functional, because in a finite dimensional space every linear functional is continuous. And continuity is related to the unique norm topology of finite dimensional space. That is why the algebraic dual, without any continuity requirement, and the continuous dual are the same in the finite dimensional case, but not in infinite dimensional space. The connection of the norm topology of $T_1^x M$ with the (indefinite) scalar product is somewhat more complicated (see section 4.1.7).

Covectors at $x \in M$ can also be defined with equivalence classes of smooth real valued mappings—imitating the definition of tangent vectors with smooth curves. Mappings ψ_1 and ψ_2 defined in the domain U of a chart χ are equal, if $(\psi_1 \circ \chi^{-1})'(\mathbf{x}) = (\psi_2 \circ \chi^{-1})'(\mathbf{x})$, where $\mathbf{x} = \chi(x)$. In this way, a covector can be represented by the triplet of a point, a chart, and a real array. Let a mapping ψ represent a covector ω and a curve ϕ represent a tangent vector v. Then the action of the covector is defined as $(\omega \,|\, v) = \frac{\partial \psi \circ \phi(t)}{\partial t}|_{t=0}$. [73]

Again let M and N be two differentiable manifolds and $\phi : M \to N$ a differentiable map (see definition 4.4). The derivative $\mathrm{d}\phi(x, \cdot) : T_1^x M \to T_1^{\phi(x)} N$ defines an associated linear mapping $\phi^* : T^1_{\phi(x)} N \to T^1_x M$ called the *pullback* so that

$$(\phi^* \omega)(v) = \omega(\mathrm{d}\phi(x, v)). \tag{4.36}$$

Thus the pullback carries the covectors in the reverse direction to the original mapping and its pushforward. The pullback is an essential tool to shift structures from one manifold to another, especially when ϕ is the chart χ and $N = \mathbb{R}^n$ (or $\phi = \chi^{-1}$ and $M = \mathbb{R}^n$) accordingly. For instance, the integral of differential forms is defined with the aid of the pullback and integration in $\mathbb{R}^n$.

Multicovectors

Multicovectors are defined identically to multivectors, with the exception that tangent vectors are replaced with covectors. Simple multicovectors are equivalence classes of tuples of covectors with the same equivalence classification as with simple multivectors. Additionally, general multicovectors follow the same procedure of formal sums and equivalence classification as multivectors. The space of p-dimensional *multicovectors* is denoted by $T_x^p M$, and especially 1-covectors are identified with covectors. As for multivectors, 0-covectors are identified with real numbers. The bundle of p-covectors is defined to be $T^p M = \coprod_{x \in M} T_x^p M$.

The space of p-covectors is the dual space of p-vectors. The bilinear action of a p-covector ω on a p-vector v is defined as follows. Once again, thanks

to bilinearity, we can assume ω and v to be simple. Let ω be representable with $(\omega_1, \omega_2, \ldots, \omega_p)$ and v with $(v_1, v_2, \ldots, v_p)$. Then

$$(\omega \,|\, v) = \begin{vmatrix} (\omega_1 \,|\, v_1) & (\omega_2 \,|\, v_1) & \ldots & (\omega_q \,|\, v_1) \\ (\omega_1 \,|\, v_2) & (\omega_2 \,|\, v_2) & \ldots & (\omega_q \,|\, v_2) \\ \vdots & \vdots & \ddots & \vdots \\ (\omega_1 \,|\, v_q) & (\omega_2 \,|\, v_q) & \ldots & (\omega_q \,|\, v_q) \end{vmatrix}. \tag{4.37}$$

The dual space of outer oriented p-vectors is the space of *twisted p-covectors*, denoted by $\tilde{T}^p_x M$. Its members are pairs of a p-covector and a local orientation with an equivalence classification similar to that of outer oriented multivectors. The action of a twisted p-covector on an outer oriented p-vector is done by the representatives of identical local orientation.

We can extend the definition of the pullback also to multicovectors. For a simple p-covector $(\omega_1, \ldots, \omega_p)$, $\phi^*(\omega_1, \ldots, \omega_p) = (\phi^*(\omega_1), \ldots, \phi^*(\omega_p))$ and for a general multivector, the pullback follows from linearity. Thus the multivectors of N are pulled back to M with the associate of the operator $\phi : M \to N$. From the definition also follows that $\phi^*(\omega \wedge \eta) = \phi^*\omega \wedge \phi^*\eta$.

For later use, we also define a new linear map, a contraction [86].

Definition 4.16. Let v be a tangent vector in $T^x_1 M$. The *contraction* is a map

$$i_v : T^p_x M \to T^{p-1}_x M : \omega \mapsto i_v \omega, \tag{4.38}$$

where

$$i_v\omega : T^x_{p-1} M \to \mathbb{R} : w \mapsto (i_v\omega \,|\, w) = (\omega \,|\, v \wedge w). \tag{4.39}$$

Differential forms

Finally, we are ready to define the tools for modeling the constitutive laws, i.e., differential forms. With differential forms, we can model the electromagnetic field problem in its entirety, but we choose to use cochains instead, because they are better interrelated with our numerical computation model. However, we cannot discard differential forms, because they are the right tool for modeling the material-dependent parts of the electromagnetic field problem.

A *differential form* ω of dimension p—a p-form for short—is a mapping that assigns each point $x \in M$ an element ω_x of the p-covector space $T^p_x M$,

$$\omega : M \to T^p M : x \mapsto (\omega_x, x). \tag{4.40}$$

In other words, a p-form is a section of the p-covector bundle. The linear space structure of multicovectors follows pointwisely to differential forms,

and the linear *space of p-forms* is denoted by $F^p(M)$. For instance, a 1-form is a covector-valued mapping, and a 0-form is a real valued function on M. In tensor language, a differential p-form is an antisymmetric covariant tensor field of rank p. The action of a smooth p-form ω on a smooth p-vector field v is a smooth real valued function on M, i.e., $(\omega|v) : M \to \mathbb{R} : x \mapsto (\omega_x|v_x)$.

As a pointwise measure, the electric field is an object that takes a tangent vector and returns the virtual work done by the field when a unit charge is infinitesimally moved along a curve that represents the vector. Hence a pointwise electric field is a 1-form. Similarly, magnetic flux density is a 2-form. One more, the differential of a smooth real valued function defined in equation (4.3) is a 1-form.

Pointwisely—as a covector—a differential form could be written componentwisely according to any local chart. Like in section 4.2, we choose an observer and e_0 as the local basis vector in the time direction. Then the 1-form component associated with the time direction is denoted dt, i.e., $(dt|e_0) = 1$ and $(dt|e_i) = 0$ for $i = 1, 2, 3$.

A twisted differential form resembles a differential form, except that p-covectors are replaced by twisted ones. But the definition can be made directly with a differential form and a global orientation. Because the global orientation requires an orientable manifold, twisted differential forms are defined only on an orientable manifold. A *twisted differential form* is an equivalence class of pairs (ω, σ) formed by a differential form ω and a global orientation σ. (ω_1, σ_1) and (ω_2, σ_2) are equivalent, if $\omega_1 = \omega_2$ and $\sigma_1 = \sigma_2$, or if $\omega_1 = -\omega_2$ and $\sigma_1 = -\sigma_2$. The linear space structure is inherited from the space of differential forms, and the space of twisted p-forms is denoted by $\tilde{F}^p(M)$. As we can see, our definition of the twisted differential forms requires an orientable manifold in the background; however, there are also more general definitions (see [31]).

The pointwise versions of electric flux and electric current are typical examples of twisted 2-forms in electromagnetics. Additionally, electric charge can be represented by a twisted 3-form.

Let again $\phi : M \to N$ be a differentiable map between differentiable manifolds M and N. If ω is a p-form in $F^p(N)$, then for any $x \in M$, $\omega(\phi(x))$ is a p-covector and ω can be pulled back to M by the pullback of p-covectors ϕ_x^*. This describes a p-form on M by

$$(\phi^*\omega)(x) = \phi_x^*(\omega(y)), \quad \text{where } y = \phi(x), \tag{4.41}$$

and thus we have defined a *pullback for differential forms*, denoted by ϕ^*. For instance, the pullback by the chart χ gives a one-to-one mapping of p-forms in a manifold and p-forms in the coordinate space $\mathbb{R}^n$.

A p-form ω on a smooth manifold M is said to be *smooth*, if $\chi^*\omega$ is smooth for each chart χ. The p-form $\chi^*\omega$ is defined on $\mathbb{R}^n$, and it is smooth, if $(\chi^*\omega|v)$ is smooth for each p-vector v. Because the manifold is smooth, this definition does not depend on the choise of χ. Similarly, if $\chi^*\omega$ for each χ is continuous, ω is said to be continuous.

For a differential form, the contraction with a vector field is pointwisely the covector contraction of the form and the vector field. Informally speaking, the contraction drops the degree of the differential form by one by inserting pointwisely the vector field as a part of the argument multivector. For example, if $\boldsymbol{b}$ is the pointwise magnetic flux density and $\boldsymbol{v}$ the velocity of the test particle, $i_{\boldsymbol{v}}\boldsymbol{b}$ is the magnetic force acting on the particle scaled by the charge of the particle. In this case, $i_{\boldsymbol{v}}\boldsymbol{b}$ is defined only on the worldline of the particle, because $\boldsymbol{v}$ is defined only on the worldline.

4.5.3 Hodge operator

The most essential application of differential forms for us are the constitutive laws of electromagnetic field problems. In a typical three-dimensional formulation, we need, e.g., a connection between electric field intensity and electric flux density. Because the former is a 1-form and the latter is a twisted 2-form, we cannot get from one to the other with a scalar valued permittivity function, as in classical vector formulation. So with $n = 3$, we need a mapping between 1- and 2-forms, one straight and one twisted. More generally, we need a mapping between p-forms and $(n-p)$-forms, one straight and one twisted. The Hodge operator is the proper tool for this job, adequate also for modeling the constitutive laws in a four-dimensional case.

To define a Hodge operator, we need an equivalent of Riesz's representation theorem of Hilbert spaces. Because in our analysis the inner product is replaced with an indefinite scalar product, we cannot rest on the results of Hilbert spaces, especially on the Hilbert space version of Riesz's theorem. Riesz's representation theorem gives an isomorphism between a Hilbert space and its dual; we need an identical result for indefinite scalar product spaces. To be precise, we need an isomorphism between p-vectors and p-covectors.

Because in the computational electromagnetics indefinite scalar product spaces are used infrequently, also the theory and application of the Hodge operator is truly unique. Consequently, we will take our time on this subject now and use the given results throughout the rest of the thesis.

First, we generalize Riesz's representation theorem into multivectors of a semi-Riemannian manifold. Then we define a Hodge operator for the multivectors, and with this Hodge we construct a Hodge operator for the differential forms.

Riesz's representation theorem

With a moderate effort, a general version of Riesz's representation theorem for the infinite dimensional indefinite scalar product space was not found in the literature. However, we can restrict our examination to finite dimensional spaces, because the space of p-vectors is finite dimensional. An orthonormal basis always exists for a finite dimensional vector space with a scalar product, and thus the following associate of Riesz's representation theorem can be proven [22].

Theorem 4.1 (Riesz's representation theorem in the semi-Riemannian manifold)**.** *Let M be an n-dimensional semi-Riemannian manifold, $0 \leq p \leq n$, and let $x \in M$ be arbitrary. $\forall\, \omega \in T^p_x M \quad \exists_1\, v \in T^x_p M$ such that*

$$(\omega \,|\, w) = \langle v, w \rangle \quad \forall\, w \in T^x_p M. \tag{4.42}$$

Proof. Because $T^p_x M$ is a finite dimensional space, there is an orthonormal basis $w_1, w_2, \ldots, w_p$ on $T^p_x M$. If $(\omega \,|\, w_i) = \alpha_i$ for $i = 1, \ldots, p$, we choose

$$v = \sum_{j=1}^{p} \langle w_j, w_j \rangle \alpha_j w_j. \tag{4.43}$$

For this v

$$\langle v, w_i \rangle = \sum_{j=1}^{p} \langle w_j, w_j \rangle \alpha_j \langle w_j, w_i \rangle = \alpha_i = (\omega \,|\, w_i), \tag{4.44}$$

with an arbitrary w_i in the basis, since w_j's are orthonormal. Because (4.44) is valid for every w_i in the basis, it is also valid for an arbitrary $w \in T^x_p M$. The v is unique due to the nondegeneracy requirement of the scalar product in definition 4.6. □

Example 4.3. In a three-dimensional Riemannian manifold, electric field intensity can be represented with a 1-form; i.e., at a point $x \in M$ the electric field intensity $\boldsymbol{e}(x)$ is a 1-covector. With Riesz's representation theorem, we can find a vector representative $\mathcal{E}_x \in T^x_1 M$ for $\boldsymbol{e}(x) \in T^1_x M$ and represent electric field intensity by a metric-dependent vector field $\mathcal{E} : M \to TM : x \mapsto (\mathcal{E}_x, x)$.

For another example, the volume form defining a measure and the global orientation on $x \in M$ are related by Riesz's representation theorem.

Definition of the Hodge operator

Now we have the tools to define the Hodge operator of a p-vector. Second, a Hodge operator can be defined for the differential forms with the aid of this p-vector Hodge.

Hodge operator of multivectors

Take an arbitrary p-vector $v \in T_p^x M$ and consider the following mapping

$$v \wedge \cdot : T_{n-p}^x M \to T_n^x M : w \mapsto v \wedge w. \tag{4.45}$$

As mentioned before, the n-vector $v \wedge w$ can be represented as a multiple of a unit n-vector σ, called the local orientation of $T^x M$. Then for each $v \in T_p^x M$ there exists a unique $\mathcal{U}(w) \in \mathbb{R}$ so that $v \wedge w = \mathcal{U}(w)\,\sigma$. When we pull all this together, we get the mapping

$$\mathcal{U} : T_{n-p}^x M \to \mathbb{R} : w \mapsto \mathcal{U}(w). \tag{4.46}$$

$\mathcal{U}$ is a real valued linear mapping, and by Riesz's representation theorem (theorem 4.1) there exists a unique $v_\sigma \in T_{n-p}^x M$ such that

$$\mathcal{U}(w) = \langle v_\sigma, w \rangle \quad \forall w \in T_{n-p}^x M. \tag{4.47}$$

Notice that v_σ changes its sign when the sign of σ is changed; i.e., the sign of v_σ depends on the local orientation. Based on the above discussion, this object is called an outer oriented multivector. Thus the pair (v_σ, σ) is a representative of an outer oriented $(n-p)$-vector, denoted by $\star v$.

The *local Hodge operator* is a mapping that sends a p-vector v into the outer oriented $(n–p)$-vector $\star v = (v_\sigma, \sigma)$. Similarly, it sends an outer oriented p-vector represented by (v, σ) into the $(n-p)$-vector $v_\sigma = \star(v, \sigma)$. If we fix the local orientation to σ, the Hodge operator for $v \in T_p^x M$ satisfies

$$v \wedge w = \langle \star v, w \rangle\, \sigma \quad \forall w \in T_{n-p}^x M. \tag{4.48}$$

For any (inner or outer oriented) p-vector v in an n-dimensional manifold, $\star \star v = (-1)^{p(n-p)+\nu} v$, where ν is the index of the inner product [22].

We can also define a *Hodge operator for a smooth p-vector field*. For every smooth p-vector field $v : M \to T_p M$, and a fixed global orientation σ,

$$v \wedge w = \langle \star v, w \rangle\, \sigma \quad \forall w \in T_{n-p} M. \tag{4.49}$$

Thus for a vector field v, the Hodge operator returns an outer oriented vector field $(\star v, \sigma)$ and for an outer oriented vector field an inner oriented vector field.

The local Hodge operator includes all the same information as the scalar product of a tangent space, because the scalar product determines the Hodge uniquely and, moreover, the Hodge operator determines uniquely the scalar product, because $v \wedge \star w = w \wedge \star v = (-1)^\nu \langle v, w \rangle \sigma$. Here the global orientation σ could be defined independently of the scalar product by $\sigma = \star 1$, where 1 is a real valued function on M, returning the unit value at every point; i.e., 1 is a smooth 0-vector field.

Because the indefinite scalar product of a Lorentzian manifold differs considerably from the standard Riemannian inner product, the local Hodge operator in the Lorentzian manifold is not well known. Additionally, the Hodge operator in indefinite scalar product spaces is a major part of the truly unique content of this thesis. Thus we cease here momentarily and give several examples of Hodges in definite and indefinite scalar product spaces with a variety of dimensions. These results are used later in examples of numerical computation.

Example 4.4. Let us consider the Hodge operator of a Euclidian space $M = \mathbb{R}^2$. We remember that because the tangent space $T_1^x \mathbb{R}^2$ can be identified with $\mathbb{R}^2$ itself, we are dealing with the Hodge of the vectors of $\mathbb{R}^2$. There is an orthonormal basis e_0, e_1 in $\mathbb{R}^2$, i.e., $\langle e_i, e_j \rangle = \delta_{ij},\ i, j = 0, 1$. Let us fix the (local) orientation to $\sigma = e_0 \wedge e_1$ and look for representatives for the outcomes of the Hodge operator. We can seek $\star e_0$ by substituting $\star e_0 = \alpha e_0 + \beta e_1$ for $\star v$ in (4.48), using e_0 and e_1 as w alternately, and thus solve the coefficients $\alpha, \beta \in \mathbb{R}$. We get $\alpha = 0$ and $\beta = 1$, thus $\star e_0 = e_1$, and similarly $\star e_1 = -e_0$. For general $v = v_0 e_0 + v_1 e_1$, $\star v = -v_1 e_0 + v_0 e_1$.

Example 4.5. For comparison, let us also consider the Hodge operator in the Minkowski 2-space $\mathbb{R}^2_1$. We have an orthonormal basis e_0, e_1 in $\mathbb{R}^2_1$, now with $\langle e_0, e_0 \rangle = -1$ and $\langle e_1, e_1 \rangle = 1$. With this change compared to the above, we get $\star e_0 = e_1$ and $\star e_1 = e_0$, so the sign of $\star e_1$ is changed in comparison to $\mathbb{R}^2$. Now $\star v = v_1 e_0 + v_0 e_1$. Figure 4.4 shows graphically how the Hodge operator maps vectors in two-dimensional Euclidian and Minkowski spaces. Both operators map vectors into orthogonal vectors, but the concept of orthogonality differs in Minkowski space from that in Euclidian space (see figure 4.2). Orthogonality does not impose the sign of the $\star v$, but the sign follows from (the version of) the Hodge operator and from the orientation of the space, here fixed to $\sigma = e_0 \wedge e_1$.

Example 4.6. Let us continue with the Hodge operator of the 1-vectors in the Minkowski 3-space $\mathbb{R}^3_1$. Once again, let e_0, e_1, e_2 be an orthonormal basis with $\langle e_0, e_0 \rangle = -1$ and $\langle e_i, e_i \rangle = 1,\ i \neq 0$. Let us choose $\sigma = e_0 \wedge e_1 \wedge e_2$ for

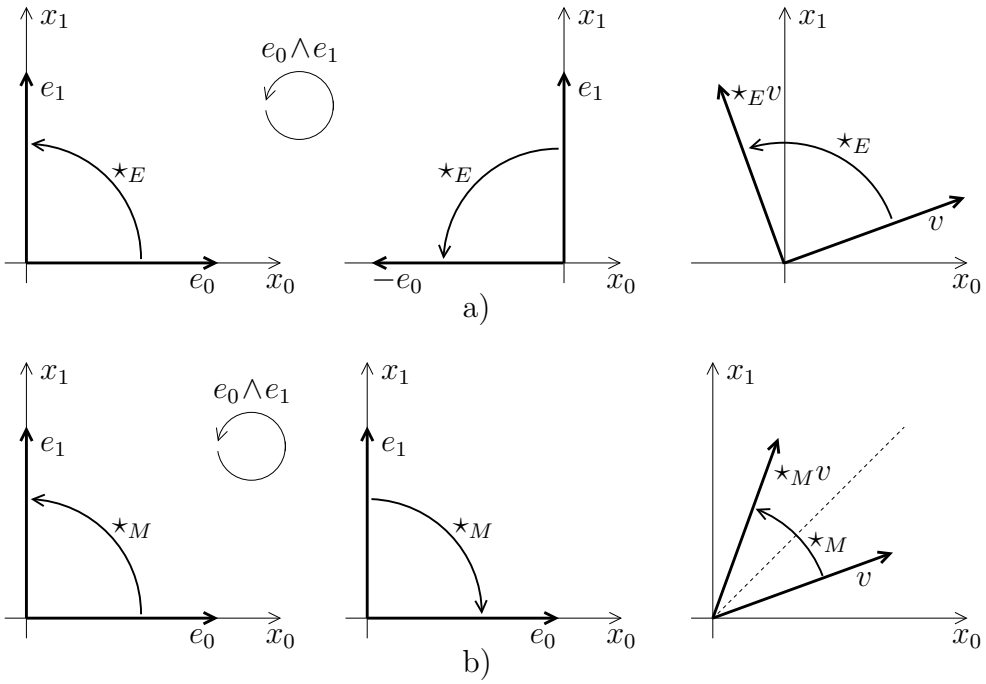


Figure 4.4: Hodge operators in two-dimensional Euclidian and Minkowski spaces. a) In Euclidian space, the Hodge operator turns all vectors counterclockwise 90° according to the orientation determined by $e_0 \wedge e_1$, b) whereas in the Minkowski 2-space, the Hodge operator turns the x_0-component 90° counterclockwise and the x_1-component clockwise. Generally, loosely speaking, the Hodge turns the vector around the closest branch of the lightcone, maintaining the same angle respect to the lightcone as in the original.

a fixed (local) orientation and $e_{01} = e_0 \wedge e_1$, $e_{12} = e_1 \wedge e_2$ and $e_{20} = e_2 \wedge e_0$ for the basis of the 2-vectors. It is an orthonormal basis with $\langle e_{01}, e_{01} \rangle = -1$, $\langle e_{12}, e_{12} \rangle = 1$ and $\langle e_{20}, e_{20} \rangle = -1$. Now, we look for $\star e_0 = \alpha e_{01} + \beta e_{12} + \gamma e_{20}$ by substituting it for equation (4.48). We get $\alpha = 0$, $\beta = 1$, and $\gamma = 0$ and thus $\star e_0 = e_{12}$. Similarly, $\star e_1 = -e_{20}$ and $\star e_2 = -e_{01}$, which means that the signs of $\star e_1$ and $\star e_2$ are changed in compared to the Euclidian case. With $p = 2$, $n = 3$, and $\nu = 1$, $\star \star v = -v$ for all 2-vectors v, and we can use the above result to find a Hodge for the 2-vectors—e.g., $\star e_{01} = - \star \star e_2 = e_2$, $\star e_{12} = -e_0$, and $\star e_{20} = e_1$. Figure 4.5 compares geometrically the effects of the Hodge operator in three-dimensional Euclidian and Minkowski spaces.

Example 4.7. In the Minkowski 4-space $\mathbb{R}^4_1$, we are mostly interested in the Hodge operator of the 2-vectors. Once more, e_0, e_1, e_2, e_3 is an orthonormal

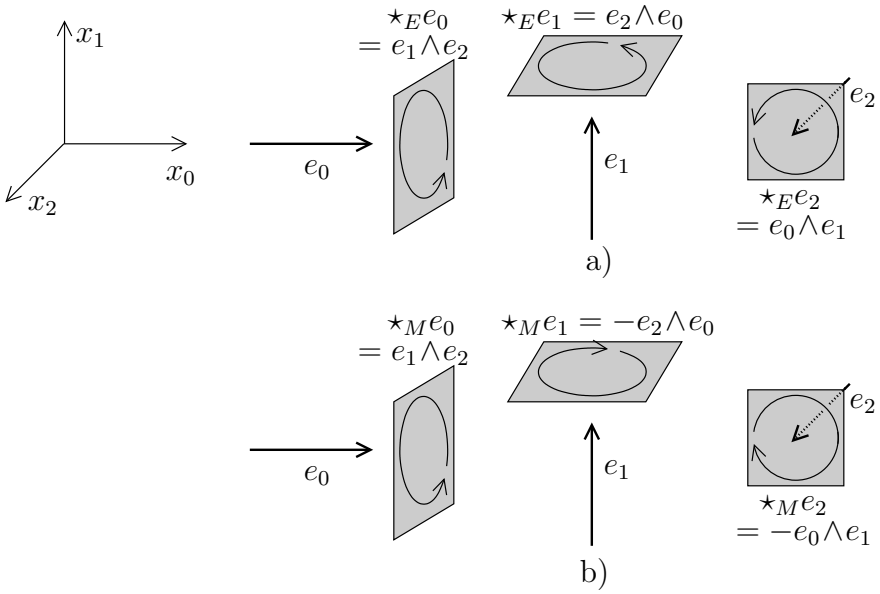


Figure 4.5: Hodge operators in three-dimensional Euclidian and Minkowski spaces. a) In Euclidian space, the Hodge operator maps a vector to a 2-vector on the orthogonal plane. Since the orientation is righthanded ($\sigma = e_0 \wedge e_1 \wedge e_2$), the orientation of the 2-plane is determined from the vector direction by the right hand rule. b) In contrast, in the Minkowski 3-space, the orientation of $\star e_1$ and $\star e_2$ is changed compared to Euclidian space. For an arbitrary vector v, the geometric interpretation of $\star_M v$ is complicated, but similar to that in the Minkowski 2-space.

basis with $\langle e_0, e_0 \rangle = -1$ and $\langle e_i, e_i \rangle = 1,\ i \neq 0$. An orthonormal basis for the 2-vectors is $e_{10} = e_1 \wedge e_0$, $e_{20} = e_2 \wedge e_0$, $e_{30} = e_3 \wedge e_0$, $e_{12} = e_1 \wedge e_2$, $e_{23} = e_2 \wedge e_3$, and $e_{31} = e_3 \wedge e_1$. If we choose $\sigma = e_0 \wedge e_1 \wedge e_2 \wedge e_3$ for a fixed (local) orientation, we get with the procedure used above $\star e_{10} = -e_{23}$, $\star e_{20} = -e_{31}$, $\star e_{30} = -e_{12}$, $\star e_{12} = e_{30}$, $\star e_{23} = e_{10}$, and $\star e_{31} = e_{20}$. If we compare these results with the Euclidian 3-space and denote by $\star_3$ the Euclidian three-dimensional Hodge operator in $\mathbb{R}^3$, we have $\star(v \wedge e_0) = -\star_3 v$ and $\star w = (\star_3 w) \wedge e_0$ for all $v \in \mathbb{R}^3$ and $w \in T_2^x \mathbb{R}^3$. Thus basically the Hodge operator of the Minkowski 4-space differs from the three-dimensional one in the wedge product with e_0 and in a change of sign with purely spatial 2-vectors.

Hodge operator of differential forms

The main issue in modeling the constitutive laws is the Hodge operator of differential forms. Again, we have two types of Hodge operators; the first one sends the differential forms to twisted ones and the second in the re-

verse direction. The definitions use pointwisely the local Hodge operators of multivectors.

The first one is of the form $\star : F^p(M) \to \tilde{F}^{n-p}(M) : \omega \mapsto \star\omega$, where $\star\omega$ is defined to be

$$\star\,\omega : M \to \tilde{T}^{n-p}M : x \mapsto \tilde{\omega} \tag{4.50}$$

where

$$\tilde{\omega} : \tilde{T}_{n-p}M \to (M \to \mathbb{R}) : (v, \sigma) \mapsto (\omega| \star (v, \sigma)). \tag{4.51}$$

Remember that $(\omega| \star (v, \sigma))$, for a smooth p-form ω and a smooth p-vector field v, is a smooth real valued function on M. In other words, for a p-form ω, $\star\omega$ is a twisted $(n-p)$-form that takes an outer oriented $(n-p)$-vector $\tilde{v}$ and gives pointwisely $\star\tilde{v}$ for ω and returns the result. Mapping (4.51) is independent of the representative of the outer oriented $(n-p)$-vector.

The second version, expressed with one representative of the outer oriented vector, is $\star : \tilde{F}^p(M) \to F^{n-p}(M) : (\omega, \sigma') \mapsto \star(\omega, \sigma')$, such that

$$\star\,(\omega, \sigma') : M \to T^{n-p}M : x \mapsto \omega' \tag{4.52}$$

and ω'_x is defined to be

$$\omega' : T_{n-p}M \to (M \to \mathbb{R}) : v \mapsto ((\omega, \sigma')| \star v). \tag{4.53}$$

Notice that the abstractions of outer oriented multivectors and twisted differential forms enable the Hodge operator to be independent of the orientation of the manifold. If we fix the manifold's orientation, we fix a straight representative for outer oriented multivectors and twisted differential forms—and, similarly, we fix a Hodge operator for these straight representatives. Thus the Hodge becomes orientation-dependent only if it is defined as a mapping between straight representatives.

Example 4.8. In a three-dimensional electromagnetic field problem, *electric field intensity* can be modeled—as a pointwise object—as a 1-form $\mathscr{e}$ and, correspondingly, *electric flux density* as a twisted 2-form $\mathscr{d}$. If the medium is isotropic and the *permittivity* is expressible with a real valued function $\varepsilon : M \to \mathbb{R} : x \mapsto \varepsilon(x)$, the constitutive law between $\mathscr{e}$ and $\mathscr{d}$ can be written $\mathscr{d} = \varepsilon \star \mathscr{e}$. Here the Hodge operator is produced by the Riemannian definite inner product. Similarly, if *electric current density* is modeled as a twisted 2-form $\mathscr{j}$, *magnetic flux density* as a 2-form $\mathscr{b}$, and *magnetic field intensity* as a twisted 1-form $\mathscr{h}$, the remaining constitutive laws can be written $\mathscr{j} = \sigma \star \mathscr{e}$ and $\mathscr{b} = \mu \star \mathscr{h}$, where $\sigma : M \to \mathbb{R}$ is the function of *electric conductivity*, and $\mu : M \to \mathbb{R}$ is that of *permeability*.

Example 4.9. In the four-dimensional formulation, electromagnetic phenomena are modeled pointwisely with two 2-forms: electromacnetic field $f = \mathscr{e} \wedge dt + \mathscr{b}$ and electromagnetic displacement $g = \mathscr{d} - \mathscr{h} \wedge dt$. These decompositions of f and g into three-dimensional quantities are observer-dependent, and dt is the 1-form associated with the time direction of this observer. In a vacuum—with units chosen so that $c = 1$—the constitutive law between f and g can be written $g = \sqrt{\varepsilon_0/\mu_0} \star f$. The Hodge operator here is produced by the scalar product of the Lorentzian manifold. These fields and this Hodge operator are discussed in detail below.

4.6 Relation between cochains and differential forms

We have now structures ready to model electromagnetics, but still lack a connection between cochains and differential forms. We are seeking a linear and bijective function between p-cochains and p-forms. This would allow us to model the Maxwell laws with cochains and the constitutive laws with differential forms.

Mapping from p-forms to p-cochains is done by integrating the p-forms over chains. Thus we first define the integration of the differential forms. Reverse mapping is done as a limiting process, because a p-cochain can be approximated with a p-form the more precisely the smaller the p-chain inserted in the cochain. Then, in the definition of the map from chains to forms, we are left with an ever diminishing sequence of p-chains. Last, we must make sure that these two mappings are really inverses of each other.

4.6.1 Integration

By integration we mean two separate but related mappings:

1. mapping from differential forms to cochains
2. mapping from chains to real numbers.

Let us define the second one first. We must define the integral of a given p-form over a p-chain. Intuitively, the integral of a 1-form over a curve c is done by chopping c into segments and then converting these segments into vectors and inserting these vectors in the 1-form. Summing the results over the curve gives the desired outcome of the integral. More generally, we need a mapping from p-simplexes to p-vectors, because we can express an arbitrary p-chain as a formal sum of p-simplexes.

Our intention is to define the mapping

$$\int_{\cdot} \omega : \ C_p(M) \to \mathbb{R} : \ \Sigma \mapsto \int_{\Sigma} \omega, \tag{4.54}$$

so that this mapping is linear. Thus we must define the mapping only for simplexes, because every p-chain Σ can be expressed as a linear combination of simplexes, i.e., $\Sigma = \sum_{i=1}^{N} \alpha_i s_i$, where s_i is a simplex for all $i = 1, \ldots, N$. The general case follows from linearity as

$$\int_{\Sigma} \omega = \sum_{i=1}^{N} \alpha_i \int_{s_i} \omega. \tag{4.55}$$

Let a simplex s be in the domain of a chart χ; by definition its image $\chi(s)$ on the coordinate space $\mathbb{R}^n$ is a simplex. By definition [91], the *integral of a continuous p-form ω* over the simplex s in a smooth manifold M is

$$\int_{s} \omega = \int_{\chi(s)} (\chi^{-1})^* \omega, \tag{4.56}$$

where $(\chi^{-1})^*$ is the pullback of χ^{-1}. This definition does not depend on the chart χ employed, as proved in [91]. Thus we must define only the integral of the p-form $\omega' = (\chi^{-1})^* \omega$ over a simplex in $\mathbb{R}^n$.

We must now somehow transform a convex p-simplex s into a p-vector in $\mathbb{R}^n$ in order to use it as an argument of the p-form ω'. But at what point of the simplex should the p-form be evaluated? The barycenter seems the natural choice. A convex p-simplex $s(P)$ is described by the set of its vertices $P = \{\boldsymbol{x}_i\}_{i=0}^{p}$. The p-form ω' is evaluated at the *barycenter* $\boldsymbol{c} = \frac{1}{p+1} \sum_{j=0}^{p} \boldsymbol{x}_j$ of the simplex s.

Because the value of ω' varies from point to point, this evaluation at one point can contribute the greater error, the larger the simplex. We work this out by *subdividing* s in $\mathbb{R}^n$ into smaller simplexes so that those of the subdivision $\mathfrak{S}_1 s$ form a complex K and the union of the cells in K equals s. Each simplex of $\mathfrak{S}_1 s$ can be subdivided in itself to form a finer subdivion $\mathfrak{S}_2 s$ for s. We can continue this process and get a sequence of simplex subdivisions $\mathfrak{S}_1 s, \mathfrak{S}_2 s, \ldots$, each element being constructed so that $s = \sum_{s_i \in \mathfrak{S}_k s} s_i$. The largest diameter of the simplices of $\mathfrak{S}_i s$ approaches zero as $i \to \infty$. The value of the integral is defined via this sequence of ever smaller and smaller simplexes, reflecting the limit aspect of the integral.

A p-simplex is determined by an affine subspace, an orientation and a convex subset of the subspace. The first two of this triplet are the same as in the

triplet specifying a simple p-vector, and the third one—the relative volume—can be concluded from the subset. If we denote $\boldsymbol{v}_i = \boldsymbol{x}_i - \boldsymbol{x}_0,\ i = 1, \ldots, p$, the p-vector corresponding to s is $\{s\} = \frac{1}{p!}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$, if the orientation of the frame $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ matches the orientation of s. We have $\|\{s\}\| = |s|_m$.

With the above definitions, the integral of a p-form ω' over a simplex s in $\mathbb{R}^n$ is defined as

$$\int_s \omega' = \lim_{k \to \infty} \omega' \circ \mathfrak{S}_k s, \tag{4.57}$$

where $\omega' \circ \mathfrak{S}_k s = \omega' \circ \sum_{s_i \in \mathfrak{S}_k s} s_i = \sum_{s_i \in \mathfrak{S}_k s} (\omega'(\boldsymbol{c}_i) | \{s_i\})$, where $\boldsymbol{c}_i$ is the barycenter of s_i. Whitney has shown that the limit of the above sequence is independent of subdivisions [91].

With (4.55), (4.56), and (4.57), we can define the integral of an arbitrary p-form over an arbitrary p-chain Σ in an n-manifold M. The use of a simplex representative of Σ is purely for convenience, because then the simplex subdivision $\mathfrak{S}_k s$ also starts from a simplex. Thus this definition applies also to an arbitrary p-cell or to the formal sum of arbitrary p-cells.

Now we can also give the second definition of the integral: a mapping from the space of differential forms to the space of cochains. The mapping from p-forms to p-cochains is defined as $\mathcal{C} : F^p(M) \to C^p(M) : \omega \mapsto \int_\cdot \omega$. However, the limit of (4.57) exists $\forall s$ only for a specific subspace of p-forms, called *flat p-forms*, which form the space $F^p_\flat(M)$. In fact, the image of $F^p_\flat(M)$ by the mapping $\mathcal{C}$ is not the whole $C^p(M)$ but the space of flat cochains $C^p_\flat(M)$. Additionally, the mapping $\mathcal{C} : F^p_\flat(M) \to C^p_\flat(M)$ is linear as we wanted, but is it also bijective? To work this out, we look for a mapping from the space of flat cochains to the space of flat differential forms.

4.6.2 Cochain → differential form

We aim for a mapping $\mathcal{F} : C^p_\flat(M) \to F^p_\flat(M)$, which is linear and the inverse of $\mathcal{C}$. Informally, it transforms a macroscopic quantity into a pointwise one by a limiting process. For an arbitrary flat p-cochain $\Omega \in C^p_\flat(M)$ it defines a flat p-form $\mathcal{F}\Omega \in F^p_\flat(M)$. To attain this, we need a way to transform an arbitrary p-vector v to a p-cell σ. We can restrict ourselves to simple p-vectors, because a p-form $\mathcal{F}\Omega$ is fully described by its outcome $(\mathcal{F}\Omega | v)$ for all simple p-vectors v.

Let v be an arbitrary simple p-vector and let us take one of its representatives $(v_1, \ldots, v_p)$. We define $(\mathcal{F}\Omega(x) | v)$ as follows. The tuple $(v_1, \ldots, v_p)$ has a counterpart in coordinate space with the real array tuple $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$. In $\mathbb{R}^n$, we can define a simplex s with $\boldsymbol{x} = \chi(x)$ at the barycenter and $\boldsymbol{v}_j$'s as edges originating from one vertex. The pair (χ, s) represents a polyhedral

simplex σ_v in M, corresponding to the p-vector v. Similarly, we can define a sequence of polyhedral simplexes $\sigma_1, \sigma_2, \ldots$ by replacing $\boldsymbol{v}_j$'s by $\frac{1}{i}\boldsymbol{v}_j$'s in the definition of σ_i. Now we can define

$$(\mathcal{F}\Omega(x)|v) = \lim_{i\to\infty} i^p \cdot (\Omega|\sigma_i). \tag{4.58}$$

Whitney has shown in [91] that limit (4.58) exists for flat cochains in the Riemannian manifold, if the simplexes are not too flat. To specify this, we define the *diameter* of a p-simplex $\operatorname{diam}(\sigma) = \sup_{x,y\in\sigma} d(x,y)$, where d is the definite metric defined in the manifold, and the *fullness* $\rho(\sigma) = \frac{|\sigma|_m}{(\operatorname{diam}(\sigma))^p}$. Now, if we have a lower bound $\eta > 0$ so that $\rho(\sigma_i) \geq \eta\ \forall i$, limit (4.58) exists. In our Lorentzian case, we cannot use this result (see the open question 4.3 below) but take a similar result as a hypothesis.

Hypothesis 4.1. *The limit of equation* (4.58) *exists, if the simplexes are regular enough.*

Though the mapping from simple p-vectors to p-cells depends on the representative of the vector, $\mathcal{F}$ does not and is thus well defined. The outcome of $\mathcal{F}$ is a flat differential form, but the result is ambiguous in that two outcomes can differ in a set of zero measure. But if we define an equivalence classification for the space $F^p_\flat(M)$ so that $\omega, \omega' \in F^p_\flat(M)$ are equal if $\int_\Sigma \omega = \int_\Sigma \omega'\ \forall \Sigma \in C^\flat_p(M)$, the result is unique. Thus, from now on, the elements of $F^p_\flat(M)$ are assumed to be equivalence classes by this classification.

Whitney also showed in [91] that $\mathcal{C}$ and $\mathcal{F}$ are inverses of each other in the Riemannian case—with the above equivalence classification of flat forms. We take the same result as a second hypothesis.

Hypothesis 4.2. *With the flat forms and cochains of the Lorentzian manifold, mappings $\mathcal{C}$ and $\mathcal{F}$ are inverses of each other.*

Open question 4.3. The proof of our hypotheses 4.1 and 4.2 by Whitney in [91] lean on Riemannian structures, which are not available for us in our Lorentzian spacetime manifold. The biggest flaw is the lack of a chain norm and a chain topology inherited by the norm. In spacetime, we have only the topology of a uniform space structure. Additionally, the proof exploits the Lebesque measure of the Riemannian manifold, which was replaced with a totally differently behaving canonical measure of equation (4.14) in order to keep the measure compatible with the metric. Furthermore, we cannot use the fullness concept to keep the simplexes regular, because fullness can approach infinity when the mass norm $\operatorname{diam}(\sigma)$ drops to zero on lightlike submanifolds.

Consequently, the exact form and proofs of the hypotheses are left open in this work. Because we can use Riemannian structures in regions far from the lightlike submanifold and thus use Whitney's results, we assume that a similar result is valid also for the whole spacetime manifold. However, the exact form of the result is not self-evident. We should specify how the regularity of the simplexes in hypothesis 4.1 is quantitated. In fact, even the exact definition of flat forms is somewhat open, because we have not fixed the topology of the chains.

We could speculate if the natural cochains of Jenny Harrison [29] would help formulate and prove these hypotheses.

Consequently, flat cochains and differential forms contain the same information, and we can move from one structure to another using the above mappings. However, the structures differ both mathematically and thus computationally. Because cochains are more compatible with our numerical computation methods, they are our primary modeling structures, but we are compelled to use differential forms as an auxiliary model for the constitutive laws.

Example 4.10. For illustration, we show how the exterior derivative of flat differential forms can be defined with the aid of the exterior operator of cochains and the mappings $\mathcal{C}$ and $\mathcal{F}$. The exterior derivative for smooth differential forms could be defined directly in coordinate space and then pulled back to M, but the next procedure extends the definition also to discontinuous flat forms. Now, if d is the exterior operator of cochains, the exterior derivative of flat differential forms d is defined as

$$\mathsf{d}\omega = \mathcal{F} \circ \mathrm{d} \circ \mathcal{C}\omega \qquad \forall \omega \in F^p_\flat(M). \tag{4.59}$$

This definition agrees with the direct definition of Lipschitz-continuous flat forms [91]. With the aid of definition (4.59) and the inversity of $\mathcal{C}$ and $\mathcal{F}$, we have $\mathcal{C}\mathsf{d}\omega = \mathrm{d}\mathcal{C}\omega \ \forall \omega \in F^p_\flat(M)$ and $\mathcal{F}\mathrm{d}\Omega = \mathsf{d}\mathcal{F}\Omega \ \forall \Omega \in C^p_\flat(M)$. Additionally, because $\mathrm{dd}\Omega = 0 \ \forall \Omega \in C^p_\flat(M)$, also $\mathsf{dd}\omega = 0 \ \forall \omega \in F^p_\flat(M)$.

Stokes' theorem for differential forms states that

$$\int_\Sigma \mathsf{d}\omega = \int_{\partial\Sigma} \omega \qquad \forall \Sigma \in C^\flat_p(M) \ \text{ and } \ \forall \omega \in F^p_\flat(M). \tag{4.60}$$

The theorem could be proved directly, but here it is, in fact, a trivial consequence of the definitions of the exterior operator for cochains and exterior derivative for differential forms.

Chapter 5

Electromagnetic theory formulated in spacetime

There is at least one exhaustive way to rationalize the theory of electromagnetism: to axiomatize it. However, these axioms must be chosen so that they agree with electromagnetic measurements. This approach is used by Hehl and Obukhov [31]. However, since this would be too extensive for us, we rather stick to the traditional way and begin with basic measurements and build the requisite mathematical concepts and laws based directly on basic sightings and measurements. Mercifully, either way, we arrive at the same theory.

Starting with basic electromagnetic measurements, we build a proper mathematical model to describe electromagnetic phenomena in a four-dimensional spacetime. First, we introduce the basic quantities and the Maxwell and constitutive laws to describe the behaviour of the quantities. However, because we have also material interfaces and boundaries, we need means to model them. After all, we want to solve the electromagnenic quantities, and potentials are a vital means for solving them. Finally, we consider cases where the model does not have to be four-dimensional, and which, with some simplifying assumptions, we can be modeled using two- or three-dimensional spacetimes.

In the following chapters, we use normal italic fonts, such as F, g, and v, to denote quantities defined in the whole four-dimensional spacetime, and calligraphic fonts, such as $\mathcal{E}$, $\mathcal{d}$, and $\mathcal{v}$, to denote quantities in a spatial submanifold that depends on the observer.

5.1 Basic electromagnetic quantities and laws

What can we measure in the electromagnetic phenomena? First, we assume that we have charged test particles, whose charge and mass are known or is measurable (we leave out the hard questions about their definition and measurability). Additionally, we assume that the location of the test particle can be detected at various time instants; i.e., we can find the worldline of the particle. By detecting the accelerations of the test particle, we find the forces affecting it and can thus discover the work done by this force.

We want to build a model to describe the measured forces and to predict them in more complicated cases. Some quantities of this model derive directly from the forces, whereas others model how materials react to these phenomena and how the reactions affect the forces. We also build a model to describe the state of the motion of charges, called the charge-current. Furthermore, based on the measurements, we formulate laws to describe the behaviour of these new quantities, i.e., the Maxwell laws.

5.1.1 From measurements to a pointwise electromagnetic field

Electromagnetism, like perhaps all fields of physics, is in one way or another based on some simple measurements. These basic measurements have a double function: they are used to define quantities and concepts to model a phenomenon, and, by this model, they predict the behaviour of the phenomenon in complex cases.

One of the basic measurements in electromagnetism is the force affecting a small test charge in various states of motion. If we know the forces on a single small charge, we can build a model of electromagnetic phenomena to predict the forces in more complex cases, e.g., for a particle with an arbitrary charge distribution.[1] The acceleration of a charged test particle reveals the electromagnetic forces acting on it. In electromagnetic phenomena, the momentum of a particle changes gradually as opposed to particle collisions, where the change is impulsive [51].

For the above reasons, we at first analyze the effect of an electromagnetic phenomenon on a particle. To describe this phenomenon, we first introduce a differential form model for electromagnetic field, even though we eventually move to a cochain model. We do this mainly for technical reasons, which arise

[1] Of course, this measurement is not enough for a comprehensive model of electromagnetic phenomena, because it reveals only the effect of the phenomena on a charged particle, not how this charge itself creates electromagnetic fields around it. Thus we also need other basic measurements and additional model structures.

when we erect these concepts from basic measurements. To familiarize ourselves with the concepts, we first examine the well-known three-dimensional case.

Example 5.1 (Basic electric and magnetic measurements in the language of differential forms)**.** In classical three-dimensional electromagnetism, we have two similar basic measurements, one for electric and one for magnetic phenomena. As we noticed earlier, electric and magnetic fields depend on the frame in which they are studied.

First, let us consider the electric field in a frame. Take a charged test particle at rest relative to a frame. Then a electromagnetic phenomenon accelerates the particle, and the acceleration gives us the *electric force* acting on the particle.

The same experiment can also be interpreted in terms of the virtual work principle. When the field accelerates the particle, it also does some work by displacing the particle. We can build a machine that takes a tangent vector and returns the virtual work made by the field, when a unit charge is infinitesimally moved along a curve that represents the vector. Furthermore, we realize that the machine is a linear map. If we know this map, we know the electric force acting on the particle. Thus in the language of differential forms this machine is a 1-form, which we can call the *pointwise electric field* and denote it by $\mathscr{e}$.

The magnetic field is more cumbersome, because now the test particle must be in motion. If we want to observe the magnetic phenomenon, we must make the test particle move relative to the frame. Compared to the previous case, we now detect some additional acceleration. If we subtract the effects of the electric field, the residual acceleration is due to *magnetic forces*.

The virtual work made by magnetic forces is now linear with respect to the velocity of the test particle (v) and, in addition, with respect to the vector associated with the displacement of the particle (w). Furthermore, this map is zero if both argument vectors v and w are parallel. Thus the map, in fact, is a 2-form and takes in a 2-vector (v, w) and returns the additional virtual work acting on the unit charge test particle. We call this 2-form the *pointwise magnetic flux*.

Let us denote the pointwise magnetic flux by $\mathscr{b}$. In the language of differential forms, the force $\frac{dp}{dt}$ acting on the test particle that travels at velocity v is $i_v\mathscr{b}$. There p is the momentum of the test particle, and thus $\frac{dp}{dt}$ is, in fact, the above force. Further, i_v is the contraction with the vector v. If we insert a second vector w, $\frac{dp}{dt} = i_v\mathscr{b}$ returns the virtual work done by this force, when the movement of the charge along any curve representing the vector w

is virtually variated by the vector v.

As we have noticed, separate electric and magnetic field concepts are not adequate for proper modeling of the electromagnetic phenomenon, because these fields depend on the frame used. These electric and magnetic phenomena cannot be separated, and thus we need a way to unify the two measurements. In the three-dimensional case, separate electric and magnetic phenomena build up a force on a charged particle based on the Lorentz force equation

$$\frac{dp}{dt} = q(e + i_v b) \tag{5.1}$$

for each test particle, where q is the charge of the particle. For this reason, it is natural to begin the search for a single basic electromagnetic force measurement from the Lorentz force.

Our aim is to combine the separate electric and magnetic fields into a unified spacetime quantity, a single electromagnetic field. The quantities in the Lorentz force equation are three-dimensional, observer-dependent objects. If we want to extend the equation into spacetime, we must replace its elements with their spacetime counterparts. Since the three-dimensional Lorentz equation must be included in the four-dimensional equation, we realize that we must have a machinery that takes in the particle's four-velocity and returns the four-force acting on the particle [51]. Let τ be the test particle's proper time and $x(\tau)$ its worldline. Then the four-velocity of the particle is $v = \frac{\partial x}{\partial \tau}$. Moreover, the four-force is the rate of change of the particle's momentum-energy over its proper time, i.e., $\frac{\partial p}{\partial \tau}$. Now, in the language of differential forms, we define the *pointwise electromagnetic field* to be a 2-form f, which obeys

$$\frac{\partial p}{\partial \tau} = q i_v f \tag{5.2}$$

for each test particle. Thus for a test particle, the time-rate of change of the momentum-energy 1-form is equal to the charge multiplied by the contraction of the electromagnetic field with the four-velocity.

As we know, in an arbitrary Lorentz frame[2], $p = -p_0\, dt + p$, where p_0 is the energy, p the momentum of the particle in this frame, and dt the basis 1-form associated with the time direction. It is well known that the energy satisfies

$$\frac{\partial p_0}{\partial t} = q(e|v). \tag{5.3}$$

[2]Henceforth, we assume that a frame-dependent, three-dimensional spatial manifold is embedded in the spacetime manifold; we can thus speak about three-dimensional quantities also in spacetime.

Thus if we want equation (5.2) to agree with (5.1) and (5.3) in every Lorentz frame for each test particle, we must have

$$f = \mathscr{e} \wedge dt + \mathscr{b}, \tag{5.4}$$

where $\mathscr{e}$ and $\mathscr{b}$ are the pointwise electric field and magnetic field flux in this Lorentz frame.

As we see, the intepretation of f is analogous with the magnetic field case, except that we are now in four dimensions. To obtain a one-to-one correspondence of the form f with its cochain counterpart, we must assume that f (and thus also $\mathscr{e}$ and $\mathscr{b}$) is a flat differential form. However, we will later require that f obey certain relations, which requisite its flatness, though it is not certain that such a flat form exists—a question that we, unfortunately, must leave open (see open question 5.1).

5.1.2 Maxwell-Faraday law

Now, we have a model for the pointwise electromagnetic field, i.e., the flat differential form f. But we also need a macroscopic counterpart for it, if we want to write the Maxwell laws in a macroscopic form. Fortunately, we have already developed techniques for this in sections 4.6.1 and 4.6.2. There we developed a one-to-one correspondence between flat cochains and flat differential forms. Hence we get the *electromagnetic field* flat cochain $F \in C^2_\flat(M)$ simply with

$$F = \mathcal{C}f \tag{5.5}$$

Alternatively, F is called the *Faraday cochain* and similarly f the *Faraday form*. Through a series of experiments, the electromagnetic field cochain F has been shown to obey the *Maxwell-Faraday law*:

$$(F|\partial\Sigma) = 0, \quad \forall \Sigma \in C^\flat_3(M). \tag{5.6}$$

By the definition of the exterior derivative, the Maxwell-Faraday law can also be written as

$$\mathrm{d}F = 0. \tag{5.7}$$

Remark 5.1. How is equation (5.6) or (5.7) connected to the three-dimensional Maxwell equations in the coordinates of an arbitrary observer? The easiest way to see this to interpret (5.6) with specially chosen Σ-cycles.

If we give a purely spatial (see section 4.3.7) chain Σ for the cochain F, we get magnetic flux through the chain, denoted $(\mathcal{B}|\Sigma)$. Being purely spatial, of

course, depends on the observer. Thus if the cycle Σ in (5.6) is purely spatial, the equation reduces to the source-freeness of the magnetic flux $\mathcal{B} \in C^2_\flat(M_\tau)$,

$$(\mathcal{B}|\partial\Sigma) = 0, \quad \forall\Sigma \in C^\flat_3(M_\tau),\ \forall\tau, \tag{5.8}$$

where M_τ is the purely spatial 3-dimensional submanifold with the proper time τ.

As a second case, we take a chain Σ that is an extrusion of a purely spatial 1-dimensional chain C along the observer's velocity field u, i.e., $\Sigma = \mathrm{extr}(C, u, t_1 - t_0)$. Here t_0 is such a time instant that Σ lies on M_τ with $\tau = t_0$ in the observer's point of view. This Σ is extruded along u until the (proper) time instant t_1. Now $(F|\Sigma)$ returns the time integral of the electromotive force along the chain C over the extrusion time period $[t_0, t_1]$ with the minus sign, i.e., $-\int_{t_0}^{t_1}(\mathcal{E}|C)dt$, where $\mathcal{E} \in C^1_\flat(M_\tau)$. The minus sign derives from the fact that the extruded 2-chain Σ is oriented disagreeing with the orientation of the boundary 1-chain C. The above integral quantity can also be called the *electromotive impulse* or *electric voltage impulse* [82].

Let us now extrude a purely spatial 2-dimensional chain S along u and take Σ to be the boundary of the resulting 3-cell (compare with figure 5.1), i.e., $\Sigma = \partial\mathrm{extr}(S, u, t_1 - t_0)$. Σ is now a 2-boundary and thus also a 2-cycle. We can now interpret that S and $\Psi_t(S)$ are the same purely spatial chain at different time instants—S is at $\tau = t_0$ and $\Psi_{t_1-t_0}(S)$ at $\tau = t_1$, where τ is the proper time of the observer. Now, on the bottom and top of Σ, $(F|\Sigma)$ reduces to $(F|S) = (\mathcal{B}|S)|_{\tau=t_0}$ and $(F|\Psi_{t_1-t_0}(S)) = (\mathcal{B}|\Psi_{t_1-t_0}(S)) = (\mathcal{B}|S)|_{\tau=t_1}$, respectively. On the side surface of Σ, $(F|\Sigma)$ gives $(F|\mathrm{extr}(\partial S, u, t_1 - t_0)) = -\int_{t_0}^{t_1}(\mathcal{E}|\partial S)dt$. Thus with the orientations described in equation (4.23) taken care of, the Maxwell-Faraday law (5.6) in this case reduces to (see figure 5.1)

$$(\mathcal{B}|S)|_{\tau=t_0} - (\mathcal{B}|S)|_{\tau=t_1} = \int_{t_0}^{t_1}(\mathcal{E}|\partial S)dt, \quad \forall S \in C^\flat_2(M_\tau),\ \forall\tau. \tag{5.9}$$

We can easily see that this is the three-dimensional Faraday law in a cochain form. Thus the Maxwell-Faraday law divides into the three-dimensional Faraday law and into the source-freeness of magnetic field.

From the three-dimensional point of view, if Σ is neither of the preceding two cases, $(F|\Sigma)$ returns a number that is a combination of the outputs of $\mathcal{B}$ and $\mathcal{E}$. By the topology of flat chains, an arbitrary Σ can always be approximated by spatial and time-extruded chains. In that way, we can always locally define F by $\mathcal{B}$ and $\mathcal{E}$, as well as divide the Maxwell-Faraday law (5.7) into its (local) observer-dependent counterparts (5.8) and (5.9).

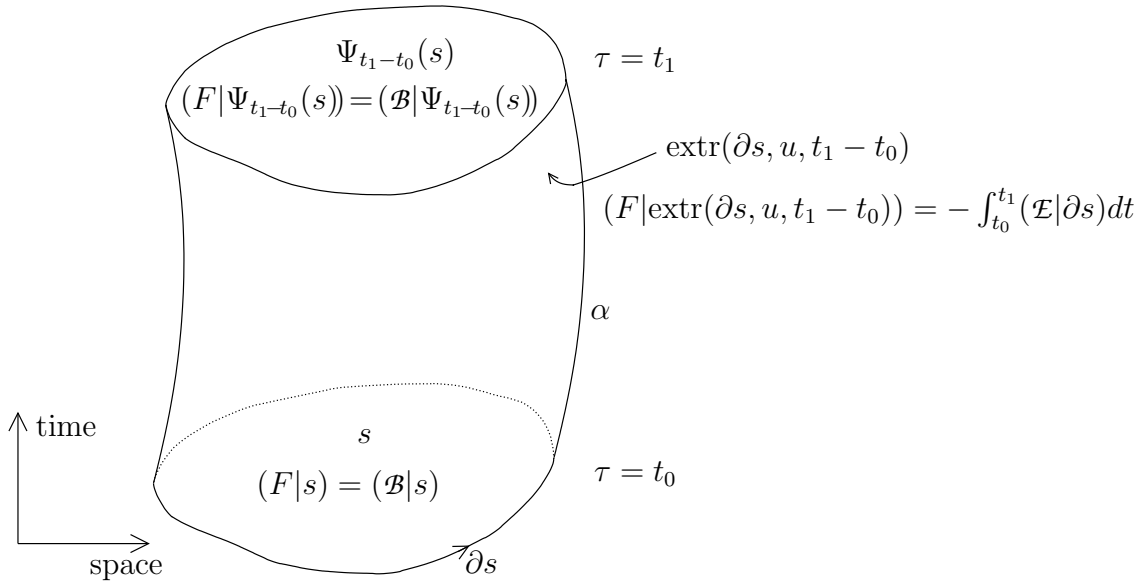


Figure 5.1: Application of the Maxwell-Faraday law in the case of a non-spatial 2-boundary Σ. Σ is construed by extruding the polyhedral 2-cell s along velocity field u (whose integral curve segment with arguments $[t_0, t_1]$ is α), and then by taking the boundary of the result, i.e., $\Sigma = \partial\mathrm{extr}(s, u, t_1 - t_0)$. Σ now consist of s on the bottom (at time $\tau = t_0$), $\Psi_{t_1-t_0}(s)$ on the top ($\tau = t_1$), and $\mathrm{extr}(\partial s, u, t_1 - t_0)$ on the side surface.

5.1.3 Electric charge-current

We have now found a model for the electromagnetic field and a Maxwell-Faraday law to describe it. However, in view of computation, they are not enough to fully describe the electromagnetic phenomena. We need some quantities to model the source terms and interaction with material.

The traditional source terms of the Maxwell equation include charge and current densities, which in an indiscrete way describe the placement and movement of charge carriers. In spacetime, both source terms combine into one quantity, the electric charge-current. Like charge and current, it is associated with outer oriented geometric objects.

Let us consider one outer oriented 3-cell $\tilde{\sigma}$, such as in figure 5.2. Each charge carrier has its own worldline, and some of them pierce $\tilde{\sigma}$. We build a weighted sum of these carriers' charges, with a positive sign for those whose direction maches with the outer orientation of $\tilde{\sigma}$, the rest with a negative sign. This is how we build the mapping $\tilde{\sigma} \mapsto \sum_i \pm q_i$. If we extend the

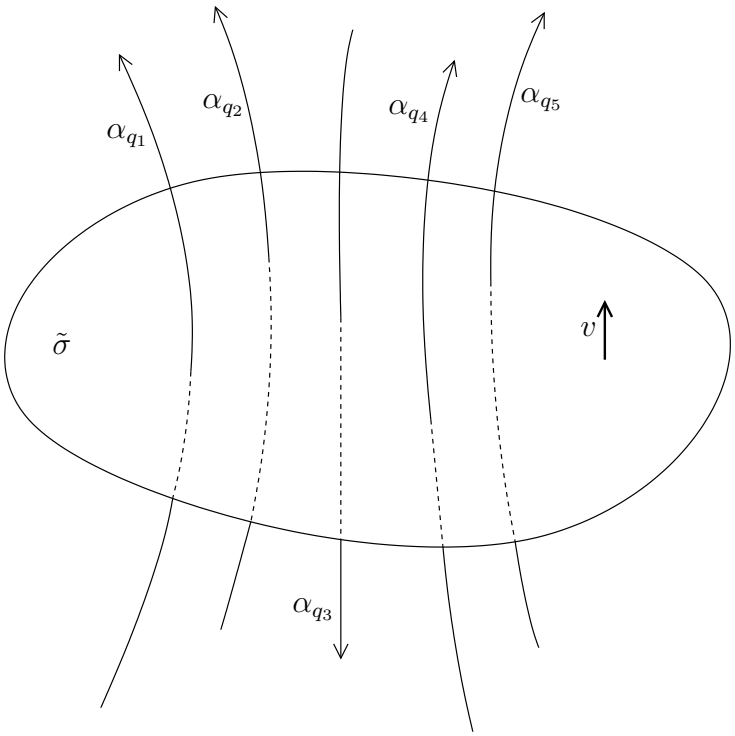


Figure 5.2: Worldlines of charge carriers $q_1, \ldots, q_5$ pass through the outer oriented 3-cell $\tilde{\sigma}$, which is oriented by the vector v. The charge carriers worldlines $\alpha_{q_1}, \ldots, \alpha_{q_5}$ pierce $\tilde{\sigma}$ according to the worldline orientation denoted in the figure. The charge-current through $\tilde{\sigma}$ is the weighted sum of the charges, where the sign is determined by whether the worldline orientation matches with v or not. Here $(J|\tilde{\sigma}) = q_1 + q_2 - q_3 + q_4 + q_5$.

mapping linearly to the flat chain space, we get a twisted flat 3-cochain

$$J : \tilde{C}_3^b(M) \to \mathbb{R} : \tilde{\sigma} \mapsto \sum_i \pm q_i, \tag{5.10}$$

where J is called the *electric charge-current*.

Remark 5.2. If an outer oriented chain $\tilde{\Sigma}$ inserted in a cochain J is purely spatial, $\tilde{\Sigma}$ is a volume in three-dimensional space, and J returns the charge inside the volume $\tilde{\Sigma}$, denoted $(\mathcal{P}|\tilde{\Sigma})$. If the chain $\tilde{\Sigma}$ is a purely spatial outer oriented 2-chain $\tilde{S}$ extruded along the observer's velocity field u (compare with figure 5.3), J returns the negative net current through the 2-chain integrated over the time period used in the extrusion, i.e., $(J|\tilde{\Sigma}(M)) = -\int_{t_0}^{t_1} (\mathcal{J}|\tilde{S}(M_\tau))dt$, where $\mathcal{J}$ is the *electric current* (for notation, see section 4.3.7). Thus the outcome is the charge penetrating through the 2-chain in

the time period (with the minus sign). Hence in both cases the outcome is a charge penetrating through some hypersurface in spacetime. The differential form counterpart $j = \mathcal{F}J$ divides into two parts relative to the observer: $j = \rho - j \wedge dt$, where ρ is a *charge density* and j is a *current density*.

It has been experimentally confirmed that the charge-current through any outer oriented 3-boundary $\partial\tilde{\Sigma}$ vanishes, i.e.,

$$(J|\partial\tilde{\Sigma}) = 0, \quad \forall\tilde{\Sigma} \in \tilde{C}_4^\flat(M). \tag{5.11}$$

This is called the *current continuity equation* or, alternatively, *charge conservation*. It can also be written as

$$\mathrm{d}J = 0, \tag{5.12}$$

which corresponds to the three-dimensional continuity equation $\mathrm{d}j + \frac{\partial\rho}{\partial t} = 0$ or in cochain form

$$(\mathcal{P}|\tilde{V})|_{\tau=t_1} - (\mathcal{P}|\tilde{V})|_{\tau=t_0} = -\int_{t_0}^{t_1} (\mathcal{J}|\partial\tilde{V})dt, \quad \forall\tilde{V} \in \tilde{C}_3^\flat(M_\tau),\ \forall\tau. \tag{5.13}$$

5.1.4 Maxwell field and constitutive law

So far we have described the electromagnetic field and its source term. Next we discuss how this source causes fields when electromagnetic phenomena interact with material. From experiments, via a lot of interpretation and theorizing, the following conclusion has been drawn

$$\mathsf{d} \star f = \sqrt{\frac{\mu_0}{\varepsilon_0}} j, \tag{5.14}$$

where j is the total electric charge-current density, possibly also including bound charge-currents. The coefficient $\sqrt{\frac{\mu_0}{\varepsilon_0}}$ stems from the units used and is here inherited from the choice of unit light speed. The bound charge-current contains bound charges and bound currents; i.e., it contains the charge density of bound electrons in the material, the polarization current density, and the magnetization current density [56].

Let us divide the total charge-current cochain $J = \mathcal{C}j$ into its free and bounded parts, i.e.,

$$J = J_{\text{free}} + J_{\text{bound}}. \tag{5.15}$$

Like the above constituents of J_{bound}, it can be written in the form

$$J_{\text{bound}} = dM, \tag{5.16}$$

where M is the *magnetization-polarization cochain*. Let $m = \mathcal{F}M$ be its differential form counterpart, a 2-form. Now, relative to some observer,

$$m = -\mathfrak{p} - \mathfrak{m} \wedge dt, \tag{5.17}$$

where $\mathfrak{p}$ is the three-dimensional *electric polarization* 2-form and $\mathfrak{m}$ the *magnetization* 1-form [85].

We define a new quantity,

$$g = \sqrt{\frac{\varepsilon_0}{\mu_0}} \star f - m, \tag{5.18}$$

and call it the *pointwise electromagnetic excitation* or *electromagnetic displacement*. Its cochain counterpart is $G = \mathcal{C}g \in \tilde{C}^2_\flat(M)$, which is called the *Maxwell cochain* or *electromagnetic excitation*. Notice that, in contrast to F, G is a twisted cochain. The pointwise version combines the observer-dependent quantities, the pointwise electric flux $\mathfrak{d}$ and the pointwise magnetic field $\mathfrak{h}$, into one observer-independent quantity

$$g = \mathfrak{d} - \mathfrak{h} \wedge dt. \tag{5.19}$$

For simplicity and tradition, we restrict ourselves to the simple case where m is a linear function of $\star f$. To simplify equation (5.18), we denote the function cleverly so that $m = \sqrt{\frac{\varepsilon_0}{\mu_0}} \star f - \star_M f$, where pointwisely

$$(\star_M f | v) = \varepsilon_r \sqrt{\frac{\varepsilon_0}{\mu_0}} (f | \star v_1) + \frac{1}{\mu_r} \sqrt{\frac{\varepsilon_0}{\mu_0}} (f | \star v_2), \tag{5.20}$$

where $v = v_1 + v_2$, v_1 is the purely spatial part of v, and v_2 is the part of v in the time direction. We can always uniquely divide an arbitrary $v \in T^x_2(M)$ into these parts relative to a chosen observer. Furthermore, ε_r and μ_r are the relative permittivity and permeability, which are scalar-valued functions of the spacetime location. Also ε_r and μ_r are observer-dependent, but $\star_M$ is not. If v is purely spatial or in the time direction, (5.20) becomes

$$(\star_M f | v) = \begin{cases} \varepsilon_r \sqrt{\frac{\varepsilon_0}{\mu_0}} (f | \star v) & \text{if } v \text{ is purely spatial,} \\ \frac{1}{\mu_r} \sqrt{\frac{\varepsilon_0}{\mu_0}} (f | \star v) & \text{if } v \text{ is in the time direction.} \end{cases} \tag{5.21}$$

If in the preceding form we now substitute m for equation (5.18), we get a *constitutive relation*

$$g = \star_M f. \tag{5.22}$$

Where does equation (5.20) come from, especially functions ε_r and μ_r? It comes from the comparison of (5.22) with three-dimensional constitutive relations.[3] Basically, $\star_M$ differs from the normal Hodge operator described in section 4.5.3 only in different scalar multipliers associated with different directions. Thus the change is only a minor technical detail. These different coefficients are inherited from material parameters. Furthermore, in our case, also the choice units, such as the speed of light, which has a unit value, contribute to these coefficients.

In order to simplify (5.20) and to clarify some concepts in numerical computation, we also define $\star_M$ for a 2-vector $v \in T_2^x(M)$. In the notation of equation (5.20), let us define

$$\star_M v = \varepsilon_r \sqrt{\frac{\varepsilon_0}{\mu_0}} \star v_1 + \frac{1}{\mu_r} \sqrt{\frac{\varepsilon_0}{\mu_0}} \star v_2, \tag{5.23}$$

which simplifies equation (5.20) into

$$(\star_M f|v) = (f| \star_M v). \tag{5.24}$$

In a vacuum, we have $\varepsilon_r = \mu_r = 1$, and thus $\star_M$ reduces to the normal Hodge operator $\star$, both in p-vector and p-form cases.

Remark 5.3. Let us look for a connection between equation (5.22) and three-dimensional constitutive relations. In the coordinates of one observer, f and

[3]We use a three-dimensional case where $\varepsilon = \varepsilon_r \varepsilon_0$ and $\mu = \mu_r \mu_0$ are scalar-valued functions of space, possibly changing with time. The only difference with a classical study is the unit of time. Because we use meter as the unit of time, the speed of light acquires a unit value. In addition, this changes the values of ε and μ, because $1/\sqrt{\varepsilon\mu}$ must be equal to the speed of light multiplied by the index of refraction. Now, let the quantities without a subscript be the traditional ones (with c as the speed of light) and those with the subscript m the modified ones with unit light speed. Now, $t_m = ct$, and we cleverly choose $\mathscr{d}_m = c\mathscr{d}$ and $\mathscr{b}_m = c\mathscr{b}$. If the variables with a traditional unit, i.e., $\mathscr{d}$, $\mathscr{b}$, $\mathscr{e}$, $\mathscr{h}$, and j, fulfil the Maxwell equations, also $\mathscr{d}_m$ and $\mathscr{b}_m$ (with $\mathscr{e}$, $\mathscr{h}$ and j) fulfil them with t_m:

$$\begin{aligned} \mathrm{d}\mathscr{e} &= -\partial_t \mathscr{b} = -c\partial_{t_m} \mathscr{b} = -\partial_{t_m} \mathscr{b}_m \\ \mathrm{d}\mathscr{h} &= -\partial_t \mathscr{d} + j = -c\partial_{t_m} \mathscr{d} + j = -\partial_{t_m} \mathscr{d}_m + j. \end{aligned}$$

With these modified variables, constitutive relations become

$$\begin{aligned} \mathscr{d}_m &= \varepsilon c\mathscr{e} = \frac{\varepsilon_r \varepsilon_0}{\sqrt{\varepsilon_0 \mu_0}} \mathscr{e} = \varepsilon_r \sqrt{\frac{\varepsilon_0}{\mu_0}} \mathscr{e}, \\ \mathscr{b}_m &= \mu c\mathscr{h} = \frac{\mu_r \mu_0}{\sqrt{\varepsilon_0 \mu_0}} \mathscr{h} = \mu_r \sqrt{\frac{\mu_0}{\varepsilon_0}} \mathscr{h}. \end{aligned}$$

Now, if we use these modified variables in place of the originals, and write the above equation in a spacetime format, we get (5.20).

g decompose according to equations (5.4) and (5.19). Furthermore, example 4.7 tells us how the Hodge operator of 2-vectors can be written in terms of the three-dimensional Riemannian Hodge-operator $\star_3$ in these coordinates. First, let v be an arbitrary purely spatial vector, i.e., $v \in T_2^x M_\tau$. Now

$$\begin{aligned}(\star_M f|v) &= \varepsilon_r\sqrt{\frac{\varepsilon_0}{\mu_0}}(f|\star v)\\ &= \varepsilon_r\sqrt{\frac{\varepsilon_0}{\mu_0}}(\mathfrak{b}+\mathfrak{e}\wedge dt|(\star_3 v)\wedge e_0)\\ &= \varepsilon_r\sqrt{\frac{\varepsilon_0}{\mu_0}}(\mathfrak{e}|\star_3 v)\\ &= \varepsilon_r\sqrt{\frac{\varepsilon_0}{\mu_0}}(\star_3\mathfrak{e}|v),\end{aligned} \tag{5.25}$$

since $\mathfrak{b}$ vanishes on timelike 2-vectors and $(dt|e_0) = 1$, where e_0 is the unit vector in the time direction. On the other hand,

$$(\star_M f|v) = (g|v) = (\mathfrak{d} - \mathfrak{h}\wedge dt|v) = (\mathfrak{d}|v). \tag{5.26}$$

Together, equations (5.25) and (5.26) equate to the constitutive relation of electric fields.

Second, we choose v to be an arbitrary 2-vector in the time direction; i.e., it can be written in the form $v = w \wedge e_0$, where $w \in T_1^x M_\tau$. We get

$$\begin{aligned}(\star_M f|v) &= \frac{1}{\mu_r}\sqrt{\frac{\varepsilon_0}{\mu_0}}(f|\star(w\wedge e_0))\\ &= \frac{1}{\mu_r}\sqrt{\frac{\varepsilon_0}{\mu_0}}(\mathfrak{b}+\mathfrak{e}\wedge dt|-\star_3 w)\\ &= -\frac{1}{\mu_r}\sqrt{\frac{\varepsilon_0}{\mu_0}}(\mathfrak{b}|\star_3 w)\\ &= -\frac{1}{\mu_r}\sqrt{\frac{\varepsilon_0}{\mu_0}}(\star_3\mathfrak{b}|w),\end{aligned} \tag{5.27}$$

and

$$(\star_M f|v) = (g|v) = (\mathfrak{d} - \mathfrak{h}\wedge dt|w\wedge e_0) = -(\mathfrak{h}|w), \tag{5.28}$$

because $\mathfrak{e}\wedge dt$ and $\mathfrak{h}\wedge dt$ are zero for purely spatial 2-vectors. Now equations (5.27) and (5.28) reduce to the constitutive relation of magnetic fields. In the coordinates of the chosen observer, an arbitrary vector v can always be uniquely divided into the above cases.

To sum up, $\star v$ is always Lorenzian orthogonal to v, and they are both 2-vectors. The orientation of the result depends on the causal character of the original v. In case of $\star_M$, there is also a different scalar coefficient depending on which component of the 2-form the operator $\star_M$ is acting on.

5.1.5 Maxwell-Ampère law

By combining equations (5.14), (5.15), (5.16), and (5.18) in a cochain form, we get

$$\mathrm{d}G = J_{\text{free}}, \tag{5.29}$$

or equally

$$(G|\partial\tilde{\Sigma}) = (J_{\text{free}}|\tilde{\Sigma}), \quad \forall\tilde{\Sigma} \in \tilde{C}_3^\flat(M). \tag{5.30}$$

These equations are called the *Maxwell-Ampère law*. Often the free-subscript of the charge-current is omitted for simplicity. Dropping the subscript is admissible, because, in our case, all the other parts of the current are included in the constitutive term $\star_M$. Additionally, because of (5.12), (5.15), and (5.16), also J_{free} satisfies the current continuity equation, i.e., $\mathrm{d}J_{\text{free}} = 0$.

Remark 5.4. We use the same technique as with the Maxwell-Faraday law to divide the Maxwell-Ampère law into its three-dimensional counterparts. If we give an outer oriented purely spatial 2-chain $\tilde{\Sigma}$ for G, we get the electric flux through the chain, i.e., $(G|\tilde{\Sigma}(M)) = (\mathcal{D}|\tilde{\Sigma}(M_\tau))$, where $\mathcal{D} \in \tilde{C}_\flat^2(M_\tau)$. If the purely spatial $\tilde{\Sigma}$ is a cycle, Maxwell-Ampère (5.30) reduces to

$$(\mathcal{D}|\partial\tilde{\Sigma}) = (\mathcal{P}|\tilde{\Sigma}), \quad \forall\tilde{\Sigma} \in \tilde{C}_3^\flat(M_\tau),\ \forall\tau. \tag{5.31}$$

This is the classical three-dimensional Maxwell source equation for the electric flux $\mathcal{D}$.

Second, we extrude a purely spatial outer oriented 1-chain $\tilde{C}$ along the velocity field u and get $\tilde{\Sigma} = \mathrm{extr}(\tilde{C}, u, t_1 - t_0)$. For this $\tilde{\Sigma}$, $(G|\tilde{\Sigma}(M)) = -\int_{t_0}^{t_1}(\mathcal{H}|C(M_\tau))dt$, where $\mathcal{H} \in \tilde{C}_\flat^1(M_\tau)$. The above time-integral of the magnetomotive force $\mathcal{H}$ along the 1-cell and over the time period $[t_0, t_1]$ is also called the *magnetomotive impulse*. Yet again, if the chain matches neither of the above cases, we get a combination of the above three-dimensional quantities.

The Maxwell-Ampère law states that the flux of G through the boundary of an outer oriented 3-chain equals the charge-current through the chain. Now, let us choose the 2-chain $\tilde{\Sigma}$ as the boundary of a chain that is produced by extruding a purely spatial 2-dimensional chain $\tilde{S}$ along the velocity field

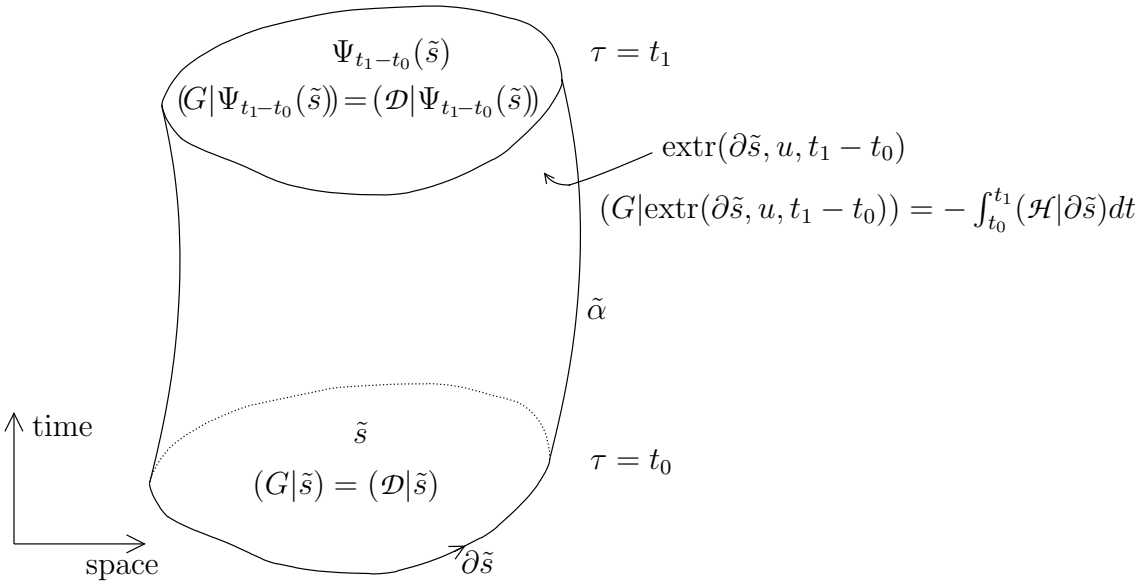


Figure 5.3: Application of the Maxwell-Ampère law in a case of non-spatial 2-boundary $\tilde{\Sigma}$. $\tilde{\Sigma}$ is construed by extruding the polyhedral 2-cell $\tilde{s}$ along the velocity field u (whose integral curve is α), and then by taking the boundary of the result, i.e., $\tilde{\Sigma} = \partial\mathrm{extr}(\tilde{s}, u, t_1 - t_0)$. $\tilde{\Sigma}$ now consist of $\tilde{s}$ on the bottom (at time $\tau = t_0$), $\Psi_{t_1-t_0}(\tilde{s})$ on the top ($\tau = t_1$) and $\mathrm{extr}(\partial\tilde{s}, u, t_1 - t_0)$ on the side surface.

u (compare with figure 5.3), i.e., $\tilde{\Sigma} = \partial\mathrm{extr}(\tilde{S}, u, t)$. As with the Maxwell-Faraday law case, the Maxwell-Ampère law (5.30) reduces, with this chosen $\tilde{\Sigma}$, to

$$(\mathcal{D}|S)|_{\tau=t_1} - (\mathcal{D}|S)|_{\tau=t_0} + \int_{t_0}^{t_1}(\mathcal{J}|S)dt = \int_{t_0}^{t_1}(\mathcal{H}|\partial S)dt, \quad \forall S \in \tilde{C}_2^\flat(M_\tau),\ \forall\tau. \tag{5.32}$$

This is the three-dimensional Ampère law in integral form. Thus, like the Maxwell-Faraday law, the Maxwell-Ampère law divides into the three-dimensional Ampère law and into the source equation for the electric flux.

Thus far we have reduced the Maxwell laws into two cochain laws and one constitutive relation that combines ε and μ. In fact, these relations are now enough for us, because we are not interested in conducting materials but in the electromagnetic wave equation for nonconducting materials. Before we can solve the equations, however, we need some interface and boundary conditions.

5.2 Interface and boundary conditions

In real-life electromagnetic analysis, we have domains with different materials; i.e., material parameters can change abruptly, when we cross the interface hypersurface between materials. Because this hypersurface also embodies a jump in electromagnetic fields, we need some additional conditions, e.g., the so-called interface conditions, to link cochains on its opposite sides.

Additionally, in many cases, e.g., in numerical computation, we cannot model the field in the whole spacetime, i.e., the whole universe. If we restrict our study to a finite piece of spacetime, we must know how the sources outside the domain affect the electromagnetic fields in the domain. All information about the electromagnetic phenomena exterior to the domain is packed in the boundary condition, i.e., in the known values of the fields on the boundary hypersurface.

5.2.1 Interface conditions

Let us study the electromagnetic fields on a hypersurface $S \subset M$. On one side of S, the material is described by the operator $\star_{M_1}$ and on the other by $\star_{M_2}$. Now, we examine how the Faraday cochain F behaves in the neighborhood of the hypersurface S. We apply the Maxwell-Faraday law (5.6) to a cylindrical

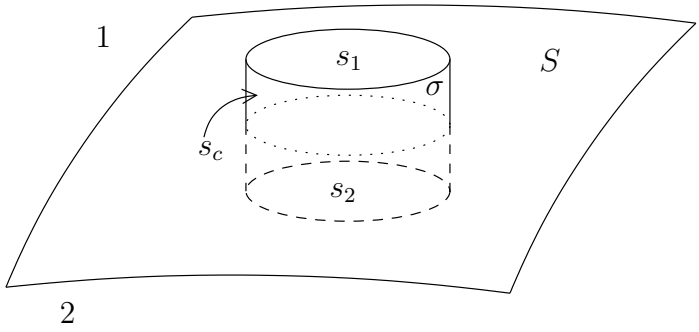


Figure 5.4: Cylindrical 3-cell σ penetrates the hypersurface S. In this figure, one dimension tangent to S is dropped, so S is shown as a normal 2-surface; however, σ remains a 3-cell. The idea is to apply the Maxwell-Faraday law (5.6) to the boundary $\partial\sigma = s_1 - s_2 - s_c$ and allow σ to be squashed so that s_1 and s_2 lie on the opposing sides of S and s_c vanishes during the limiting process.

boundary $\partial\sigma$, as in figure 5.4. If we allow the cylinder to shrink in its axial direction, yet require that its top and bottom lie on different sides of S, we get $(F|s_1) = (F|s_2)$. Moreover, this is valid for arbitrary shrinking cylindrical

shaped σ. If we denote $[(F|s)] = (F|s_1) - (F|s_2)$, i.e., the jump of F on the 3-cell $s \subset S$ across the hypersurface S so that $s_1 \to s$ and $s_2 \to s$ in the shrinking process, then in general we have

$$[(F|\Sigma)] = 0, \qquad \forall \Sigma \in C_2^\flat(S).$$

However, because $C_2(S)$ is a dense subspace of $C_2^\flat(S)$ and F is continuous, the above equation is equivalent to

$$[(F|\Sigma)] = 0, \qquad \forall \Sigma \in C_2(S). \tag{5.33}$$

Since the jump has zero value, the direction of the jump is irrelevant. In fact, if F is really a flat cochain, i.e., $F \in C_\flat^2(M)$, then (5.33) is always valid by the definition of flat cochains. In fact, this is one important reason why F must be a flat cochain.

Similarly, only by replacing the cells and chains with outer oriented ones, we get a similar result also for the Maxwell cochain G from the Maxwell-Ampère law (5.30). The only difference to the Maxwell-Faraday case is that the Maxwell-Ampère law has a nonzero right-hand side. In some modeling situations, the charge-current J of (5.30) does not vanish in the limiting process of a shrinking outer oriented cylinder $\tilde{\sigma}$ (compare with fig. 5.4). In these cases, we have a *hypersurface charge-current* J_S on S, i.e., $J_S \in \tilde{C}_\flat^2(S)$. Contrary to the above, we must also decide the direction of the jump. We assume that S is outer oriented[4] and that the jump direction is chosen according to the positive crossing direction of S. Consequently, we get a second interface condition

$$[(G|\tilde{\Sigma})] = (J_S|\tilde{\Sigma}), \qquad \forall \tilde{\Sigma} \in \tilde{C}_2(S). \tag{5.34}$$

If $J_S \neq 0$, the Maxwell cochain G cannot be a flat cochain. However, it can be piecewisely flat so that it is flat in the regions outside the interfaces with $J_S \neq 0$. On these interfaces, G is connected according to an additional condition (5.34). In fact, in our study, we focus on the case where $J_S = 0$ on every interface. Hence we can assume that G is flat; thus $[(G|\tilde{\Sigma})]$ vanishes on every surface automatically.

Remark 5.5. How are spacetime interface conditions (5.33) and (5.34) connected to the three-dimensional ones of a chosen observer? Let us assume that the hypersurface S is an extrusion of a spatial surface S_τ along the observer's velocity field. If we first assume that the 2-chains Σ and $\tilde{\Sigma}$ in

[4]The definition could be done without the requirement of outer orientation with a proper definition of twisted forms [31].

(5.33) and (5.34) are also in the time direction, they are then extrusions of spatial 1-chains C and $\tilde{C}$ along the observer's velocity field. In this case, $(J_S|\tilde{\Sigma}) = \int_{t_0}^{t_1} (\mathcal{J}_S|\tilde{C})dt$, where $\mathcal{J}_S \in \tilde{C}^1_\flat(S_\tau)$ is the *surface current* on the spatial surface S_τ. Now the interface conditions reduce to

$$[(\mathcal{E}|C)] = \quad 0, \qquad \forall C \in C_1(S_\tau), \tag{5.35}$$
$$[(\mathcal{H}|\tilde{C})] = (\mathcal{J}_S|\tilde{C}), \qquad \forall \tilde{C} \in \tilde{C}_1(S_\tau). \tag{5.36}$$

Second, if Σ and $\tilde{\Sigma}$ indeed are spatial chains in S_τ, the interface conditions become

$$[(\mathcal{B}|\Sigma)] = \quad 0, \qquad \forall \Sigma \in C_2(S_\tau), \tag{5.37}$$
$$[(\mathcal{D}|\tilde{\Sigma})] = (\mathcal{P}_S|\tilde{\Sigma}), \qquad \forall \tilde{\Sigma} \in \tilde{C}_2(S_\tau), \tag{5.38}$$

where $\mathcal{P}_S \in \tilde{C}^2_\flat(S_\tau)$ is the *surface charge cochain*. Thus if the surface S is in the time direction, the interface conditions (5.33) and (5.34) reduce to the familiar three-dimensional conditions (5.35)–(5.38).

In contrast, if S is purely spatial, Σ and $\tilde{\Sigma}$ must also be purely spatial, and the spacetime interface conditions then reduce to

$$[(\mathcal{B}|\Sigma)] = 0, \qquad \forall \Sigma \in C_2(S), \tag{5.39}$$
$$[(\mathcal{D}|\tilde{\Sigma})] = 0, \qquad \forall \tilde{\Sigma} \in \tilde{C}_2(S). \tag{5.40}$$

The right-hand side of (5.40) vanishes, because the time integral of the electric current vanishes, as the length of the integration period diminishes to zero. From the observer's point of view, the interface between different materials occurs on the hypersurface of simultaneity; i.e., at a fixed time, the material in some domain is replaced by some other material. The above conditions state that $\mathcal{B}$ and $\mathcal{D}$ keep their value in this change, and no other conditions are needed.

5.2.2 Boundary conditions

In spacetime electromagnetic problems, boundary conditions consist of the initial and boundary conditions of the equivalent three-dimensional problem. Let us first examine the initial conditions and later add the spatial boundaries.

For the sake of simplicity, we limit our investigation to the simplest boundary conditions, i.e., conditions which on the boundary hypersurface fix the value of the electromagnetic field F or excitation G, or both. In the case of scalar potentials, these correspond to the Dirichlet, Neumann, and Cauchy conditions, respectively.

Initial condition

Let us assume that we know F and G (and thus their pointwise counterparts f and g) on a spacelike surface S of a boundaryless spacetime M. Additionally, we assume that there are no source charge-currents in M. Take a connected subset S_0 of S. Since electromagnetic phenomena can proceed only at the speed of light (in a vacuum) or more slowly, f in the event $x \in M$ is influenced by f and g in $y \in S_0$, if $x > y$.[5] In other words, there exists a past-pointing causal curve from x to y. $f(x)$ depends *only* on the values of F and G in S_0 (and not at all on values in $S \setminus S_0$), if every past-pointing inextendible[6] causal curve through x also intersects S_0. All the events similar to x form the set called the *future domain of dependence* of S_0, denoted $D^+(S_0)$ (see fig 5.5). More precisely, $x \in D^+(S_0)$, if and only if every past-pointing inextendible causal curve through x intersects S_0. Thus the initial data of F and G on S_0 uniquely determines $f(x)$ for every $x \in D^+(S_0)$.

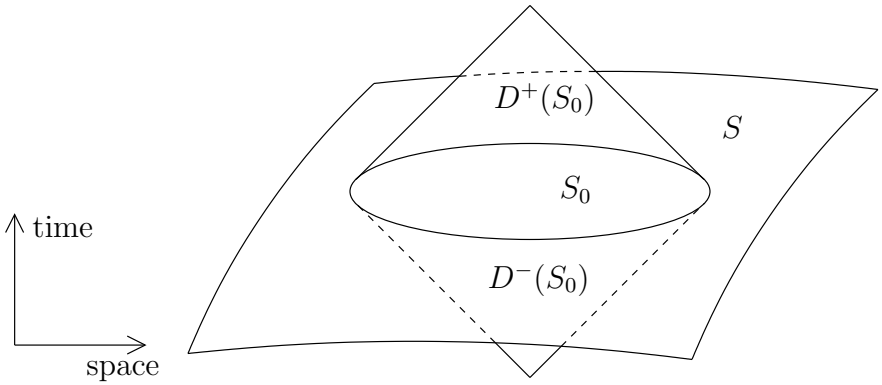


Figure 5.5: Future and past domains of dependence of S_0, denoted $D^+(S_0)$ and $D^-(S_0)$, respectively. The future domain of influence comprises points such that all information from the hypersurface S to this domain must originate from S_0. The past domain contains events whose information cannot reach $S \setminus S_0$.

By changing the past-pointing curve to a future-pointing one in the above, we can define the *past domain of dependence* of S_0, denoted $D^-(S_0)$ (see fig 5.5). If we know $F(\Sigma)$ and $G(\tilde{\Sigma})$ $\forall \Sigma \in C_2(S_0)$ and $\forall \tilde{\Sigma} \in \tilde{C}_2(S_0)$, we can solve the field F in the past domain $D^-(S_0)$ and predict F in the future domain

[5]Because we need to talk about pointwise values of the field, we must use the pointwise differential form counterpart f of the electromagnetic field cochain F.

[6]In this case, being inextendible means only that the curve continues forever and does not end at a finite point, e.g., at some singularity.

$D^+(S_0)$. Thus we can solve an *initial value problem* in $D^+(S_0)$ and a *final value problem* in $D^-(S_0)$. [13]

We can extend the domains of dependence to the whole spacelike hypersurface S. To fully solve the electromagnetic fields in the whole M, we must have $M = D^+(S) \cup D^-(S)$. This does not work, e.g., if we have closed causal curves or singularities (see [13] for details). These complications motivate us to rule out closed causal curves and to assume that spacetime is causal (see section 4.1.7), as in the theory of relativity. In the same fashion, we choose to leave black holes and other singularities outside of our investigation.

However, the initial or final data in S cannot be arbitrary. The given fields F and G must fulfil Maxwell equations on the hypersurface S. Because S is spacelike, there is—at least locally—an observer, who sees that the hypersurface is purely spatial. In the coordinates of this local observer, chains on S must also be purely spatial, and thus the initial (or final) data fixes only magnetic and electric fluxes $\mathcal{B}$ and $\mathcal{D}$. Thus in the coordinates of this observer, these constraints on the data mean that $\mathcal{B}$ and $\mathcal{D}$ must obey Maxwell source equations on S. These constraint equations are not needed for other time instants, because the other parts of Maxwell equation enforce them, if the constraints are realized in one time instant, and if we take the current continuity equation (5.11) for granted [10]. In chapter 6, we will return to this question in the context of dual volume trees.

But why do we have the value of both F and G on the hypersurface S for the initial or final value? Later, we see that only one condition is needed for spatial boundary conditions. In fact, there is no compelling reason for having two conditions in a single hypersurface. It is our modeling decision to give either two initial or final conditions, but not both in the same problem, and we are free to give one initial and one final condition instead of two initial or final conditions. In the latter event, we should have two non-crossing spacelike hypersurfaces S_1 and S_2 so that S_2 would be in the future of S_1 and the event $x \in M$ between them. Now, if we know the value of F or G on $J^-(x) \cap S_1$ and on $J^+(x) \cap S_2$, we can solve $f(x)$. However, both final value problems and combined initial-final value problems are so rare in electromagnetics that henceforth we concentrate rather on pure initial value problems. After all, we want to predict the future and not to trace the history of the fields before a time instant or between two instants. However, with minor changes, these other alternatives could also be possible.

Spatial boundary conditions

Because we aim to solve electromagnetic fields numerically, we must restrict our problem to a bounded domain Ω. Ω is a bounded n-dimensional subman-

ifold of the spacetime manifold M, and its boundary $\partial\Omega$ has both spacelike and timelike parts. For the spacelike parts, we set initial (or final) conditions similar to those in the previous section, and for the timelike parts we set spatial boundary conditions. First, let us examine one simple case for illustration.

Let us assume that we know the initial field distributions of F and G on a spacelike hypersurface S, and that the hypersurface crosses with a timelike hypersurface Q. S and Q locally enclose the domain Ω, and thus Q and a part of S belong to the boundary $\partial\Omega$ according to figure 5.6. By a spatial boundary condition, the value of F or G is fixed on Q. The hypersurface Q cuts S into two parts and leaves a part of (a connected subset) $S_0 \subset S$ outside of Ω, as in figure 5.6. Now, in order to solve F in $D^+(S_0) \cap \Omega$, we

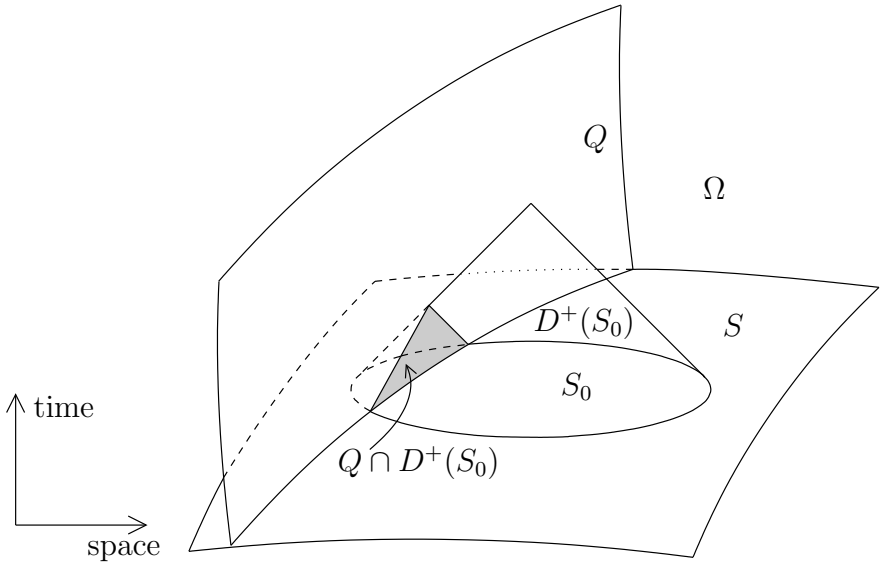


Figure 5.6: Ω is the spacetime domain of examination, and its boundary $\partial\Omega$ contains a (part of the) spacelike boundary S and a timelike boundary Q. If F and G are given on $S_0 \cap \Omega$ and F or G on $Q \cap D^+(S_0)$, we can solve F in $D^+(S_0) \cap \Omega$.

must fix F and G on $S_0 \cap \Omega$ and F or G on $Q \cap D^+(S_0)$. Thus we have two conditions on the initial boundary S and only one condition on the spatial boundary Q.

For a general boundary, we must divide the boundary into two separate parts, one for a spatial boundary condition and the other for an initial condition. First, a spatial boundary condition is used for all timelike boundaries, denoted Γ_{boundary}. Second, spacelike boundaries must be divided into initial and final boundaries; the initial conditions are placed on the first ones. We

have noticed earlier that using lightlike submanifolds involves difficulties such as those with chain norms in section 4.3.4. Furthermore, in electromagnetic computation, lightlike boundaries might disturb the causality. Hence we assume that there is no lightlike part in $\partial\Omega$, which in practise means that there is always a sharp corner between adjacent spacelike and timelike parts of the boundary.

For timelike boundaries, F or G is fixed as spatial boundary data on Γ_{boundary}; for the spacelike part, we need a way to divide them into initial and final boundaries. The initial boundary Γ_{initial} is that part of the spacelike boundary whose inside normal vector is future-pointing. The rest of that spacelike boundary belongs to the final boundary, denoted Γ_{final}. In the initial boundary Γ_{initial}, we fix F and G as initial data. In contrast, in the final boundary we leave the fields unfixed, because we already have two initial conditions. Generally, we have $\partial\Omega = \Gamma_{\text{boundary}} \cup \Gamma_{\text{initial}} \cup \Gamma_{\text{final}}$, where Γ_{boundary}, Γ_{initial}, and Γ_{final} are separate.

Henceforth, in many occasions, we use a certain model geometry with certain boundary conditions, a prototype problem geometry shown in figure 5.7.

Open question 5.1. It would be essential to check the existence and uniqueness of a solution with chosen boundary conditions, but unfortunately we must bypass this question in this thesis. Especially, whether or not such a *flat* 2-cochain exists—which satisfies the Maxwell equations (5.6) and (5.30), the constitutive relation (5.22), and the boundary conditions—is not self-evident.

For the timelike (i.e., spatial) boundary Γ_{boundary}, it is locally always possible to find an observer whose worldlines are tangential to the boundary. Then fixing F on Γ_{boundary} would mean imposing the electric field $\mathcal{E}$ of this local observer on Γ_{boundary}. In contrast, specifying G would mean fixing $\mathcal{H}$ on the boundary.

Modeling boundary conditions

How do we know the boundary conditions in an electromagnetic modeling situation? Placing initial conditions assumes that at some time instant (of some observer) we know the fields. For example, this information can derive from measurements, or we can place the initial boundary at so early a time instant that fields are still zero or at a constant known value. However, the knowledge of the fields can be purely speculative, and only if we got some positive results from simulations, would we be able to determine the physical feasibility of these boundary conditions.

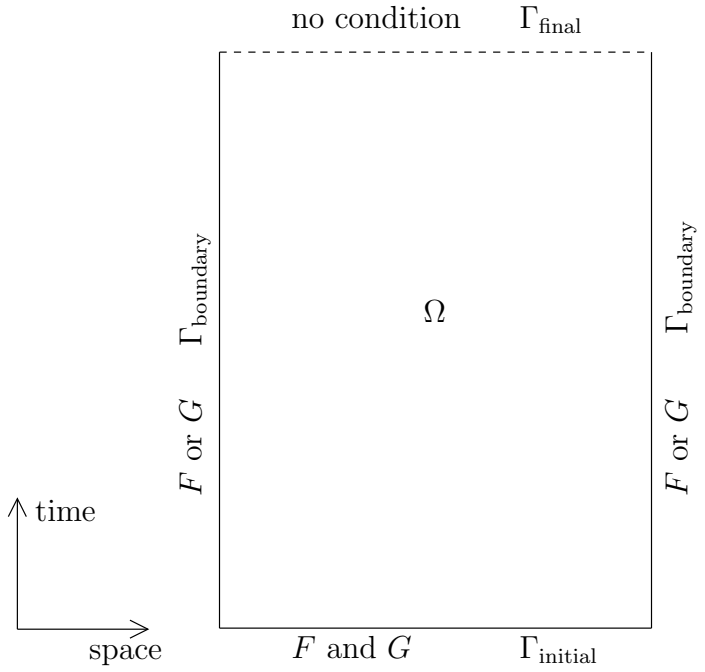


Figure 5.7: Our prototype problem geometry, shown here in two dimensions. The boundary of our domain $\partial\Omega$ divides into timelike parts Γ_{boundary}, a spacelike initial part Γ_{initial}, and a final part Γ_{final}. On the initial boundary Γ_{initial}, we impose the values of both F and G, on Γ_{boundary} only one of them. On the final boundary Γ_{final} no condition is imposed.

Similarly, spatial boundaries can sometimes be placed so far from the sources that we can assume fields to vanish there. Another typical way is to place them on the boundaries of such materials where F or G can be known in advance, though perhaps only approximately. For example, inside good conductors, F is very small, except for a thin surface region. However, good conductors are often modeled as *perfect electric conductors*, which by definition stands for a material where $F = 0$—also on the surface. Hence we place the spatial boundary Γ_{boundary} on the surface of a perfect electric conductor and place the condition $F = 0$ there. There is also a dual counterpart, called the *perfect magnetic conductor*, where by definition $G = 0$. Both these conditions are used henceforth throughout this work.

Relative cochain spaces

Mathematically, boundary conditions are placed by restricting the cochain spaces. Let us begin with relative chain spaces modulo the boundary. Let Γ be a part of $\partial\Omega$ (possibly even $\Gamma = \partial\Omega$) and $C_p^\flat(\Gamma)$ the space of flat p-chains

on Γ. With the inclusion $i : C_p^\flat(\Gamma) \to C_p^\flat(\Omega)$, we can define the space of *relative flat p-chains modulo the boundary* Γ as $C_p^\flat(\Omega,\Gamma) = C_p^\flat(\Omega)/i[C_p^\flat(\Gamma)]$. By means of i and the mapping $j : C_p^\flat(\Omega) \to C_p^\flat(\Omega,\Gamma)$, we have a short exact sequence

$$0 \to C_p^\flat(\Gamma) \xrightarrow{i} C_p^\flat(\Omega) \xrightarrow{j} C_p^\flat(\Omega,\Gamma) \to 0. \tag{5.41}$$

A relative chain $\sigma \in C_p^\flat(\Omega,\Gamma)$ is an equivalence class of $C_p^\flat(\Omega)$ with a relation $\sigma \sim \sigma'$ if $(\sigma - \sigma') \in C_p^\flat(\Gamma)$. The boundary operator for $C_p^\flat(\Omega,\Gamma)$ is defined by ∂ of the representative chain in $C_p^\flat(\Omega)$; by the definitions of ∂ and relative chains, it is independent of the choice of representative. Consequently, we can define the spaces of relative cycles and boundaries modulo Γ, denoted $\mathcal{Z}_p^\flat(\Omega,\Gamma)$ and $\mathcal{B}_p^\flat(\Omega,\Gamma)$, respectively. [30]

The *relative p-cochain space modulo* Γ is defined as the dual space of $C_p^\flat(\Omega,\Gamma)$ and denoted $C_\flat^p(\Omega,\Gamma)$ (or $C_\flat^p(\Omega-\Gamma)$). By definition, $C_\flat^p(\Omega,\Gamma)$ forms the space of p-cochains that vanish on $C_p^\flat(\Gamma)$. If we denote the space of p-cochains on Γ by $C_\flat^p(\Gamma)$, we get a short exact sequence

$$0 \leftarrow C_\flat^p(\Gamma) \xleftarrow{t_\Gamma} C_\flat^p(\Omega) \xleftarrow{k} C_\flat^p(\Omega,\Gamma) \leftarrow 0. \tag{5.42}$$

The mapping t_Γ is the *trace* of a p-cochain on Γ, and k is the inclusion of $C_\flat^p(\Omega,\Gamma)$ in $C_\flat^p(\Omega)$. Like with chain counterparts, we can also define the exterior derivative of relative cochains and, consequently, the spaces of relative cocycles and coboundaries modulo Γ, denoted $\mathcal{Z}_\flat^p(\Omega,\Gamma)$ and $\mathcal{B}_\flat^p(\Omega,\Gamma)$.

If we want to enforce a boundary condition on a field $F \in C_\flat^p(\Omega)$ on Γ, we fix the value of $t_\Gamma F \in C_\flat^p(\Gamma)$. Thus by the boundary condition we know the value of the part of F that lies in $C_\flat^p(\Gamma)$. By (5.42), if the $C_\flat^p(\Gamma)$- and $C_\flat^p(\Omega,\Gamma)$-parts of F are known, we also know the whole $F \in C_\flat^p(\Omega)$ [33]. Hence, after asserting the boundary condition, a field in $C_\flat^p(\Omega,\Gamma)$ remains to be solved.

We denote by Γ_F the part $\partial\Omega$ where the Faraday field $F \in C_\flat^2(\Omega)$ is fixed and by Γ_G the part where the Maxwell field $G \in \tilde{C}_\flat^2(\Omega)$ is fixed. Note that generally $\Gamma_F \cap \Gamma_G \neq \emptyset$ and $\Gamma_F \cup \Gamma_G \neq \partial\Omega$. For the Faraday field F we know by boundary conditions $t_{\Gamma_F} F \in C_\flat^2(\Gamma_F)$ and the unknown is $F' \in C_\flat^p(\Omega,\Gamma_F)$. Similarly for the Maxwell field G, we know $t_{\Gamma_G} G \in \tilde{C}_\flat^2(\Gamma_G)$ and $G' \in \tilde{C}_\flat^p(\Omega,\Gamma_G)$ remains unknown.

5.3 Potentials

Our next step is to ponder how to solve the electromagnetic problem: the Maxwell equations (5.6) and (5.30), the constitutive relation (5.22), and the

boundary conditions. In many electromagnetic models, the use of potentials simplifies the formulation and also the solution process; this applies also to our case.

The existence of potentials is always connected to the topology of the domain. Because we are not interested in topologically complicated domains, let us make a quite rough assumption that guarantees the existence of a potential. Our domain Ω is assumed to be starshaped, bound, or non-bound. Now, since the exterior derivative of F is zero (i.e., F is a cocycle or a closed cochain), there is a flat 1-cochain A in Ω that satisfies [91]

$$F = \mathrm{d}A. \tag{5.43}$$

This 1-cochain A is called the *electromagnetic potential* or *four-potential*. Let Σ be an arbitrary flat 2-chain. By the definition of the exterior derivative, (5.43) is equal to

$$(A\,|\partial\Sigma) = (F\,|\,\Sigma), \quad \forall\Sigma \in C_2^\flat(M). \tag{5.44}$$

Thus the flux of F on a cochain Σ equals the circulation of A around $\partial\Sigma$.

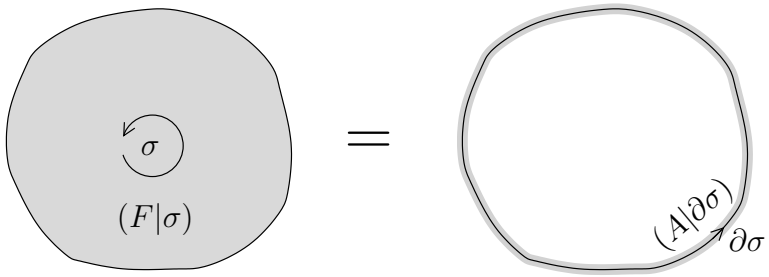


Figure 5.8: Action of F on cell σ is identical with that of the electromagnetic potential A on boundary $\partial\sigma$.

Remark 5.6. If the cochain Σ is purely spatial relative to the observer, $(F\,|\,\Sigma) = (\mathcal{B}\,|\,\Sigma)$, and thus also $(A\,|\,\partial\Sigma) = (\mathcal{A}\,|\,\partial\Sigma)$, where $\mathcal{A}$ is a three-dimensional *magnetic* (vector) *potential* satisfying $d\mathcal{A} = \mathcal{B}$. Thus for a purely spatial chain, the four-potential is the same as the magnetic potential.

Second, if $\Sigma = \mathrm{extr}(C, u, t_1 - t_0)$, for a purely spatial 1-chain C and for an observer with a velocity field u and a period $[t_0, t_1]$, we have $(F\,|\,\Sigma) = -\int_{t_0}^{t_1}(\mathcal{E}|C)dt$. Now, in the notation of section 4.3.7, $\partial\Sigma = \Psi_{t_1-t_0}(C) - C - \mathrm{extr}(\partial C, u, t_1 - t_0)$. If we denote $(A\,|\,\mathrm{extr}(\partial C, u, t_1 - t_0)) = -(\Phi\,|\,\mathrm{extr}(\partial C, u, t_1 - t_0))$, then $(A\,|\,\partial\Sigma) = (\mathcal{A}\,|\Psi_t(C)) - (\mathcal{A}\,|C) + (\Phi\,|\,\mathrm{extr}(\partial C, u, t_1 - t_0))$. If we compare this with the standard three-dimensional case, we find that the

1-cochain Φ must be the time integral of the *electric scalar potential* cochain ϕ that satisfies $\mathcal{E} = -\partial_t \mathcal{A} - \mathrm{d}\phi$ in the spatial submanifold.

In differential form format, $\mathcal{F}A = a$, and, for this observer, $a = \mathcal{a} - \varphi dt$, where $\mathcal{a} = \mathcal{F}\mathcal{A}$ (the magnetic 1-form potential) and $\varphi dt = \mathcal{F}\Phi$ (φ is the electric scalar potential differential form).

If we use the potential described in (5.43) for F, Maxwell-Faraday law (5.7) is satisfied automatically. Thus solving an electromagnetic problem simplifies into the following: find a potential A so that (5.30), (5.22), and the boundary conditions are satisfied.

5.3.1 Gauge conditions

For a known F, the electromagnetic potential $A \in C^1_\flat(M)$ described by (5.43) is not unique. We can always add the exterior derivative of a scalar potential $\xi \in C^0_\flat(M)$ to A without affecting F, because $\mathrm{d}(A + \mathrm{d}\xi) = \mathrm{d}A = F$. In fact, in our topologically simple case, choosing this ξ fixes all the degrees of freedom of the potential A.

Choosing ξ, or making A unique, is called *gauging*, and the condition required of A in that case is called the *gauge condition*. There are many different gauge conditions, from which we can obtain different values of A, but we always obtain the same F.

Another point of view on gauging is associated with the following short exact sequence

$$0 \to \mathcal{Z}^p_\flat(\Omega, \Gamma) \xrightarrow{i} C^p_\flat(\Omega, \Gamma) \xrightarrow{d} \mathcal{B}^{p+1}_\flat(\Omega, \Gamma) \to 0, \tag{5.45}$$

where $\mathcal{Z}^p_\flat(\Omega, \Gamma)$ and $\mathcal{B}^{p+1}_\flat(\Omega, \Gamma)$ are the spaces of relative cocycles and coboundaries modulo Γ, respectively. As a corollary of exactness of sequence (5.45) we get $C^p_\flat(\Omega, \Gamma)/i(\mathcal{Z}^p_\flat(\Omega, \Gamma)) \cong \mathcal{B}^{p+1}_\flat(\Omega, \Gamma)$ and $C^p_\flat(\Omega, \Gamma) \cong \mathcal{Z}^p_\flat(\Omega, \Gamma) \oplus C^p_{\flat,\mathrm{d}}(\Omega, \Gamma)$, where d is an isomorphism from $C^p_{\flat,\mathrm{d}}(\Omega, \Gamma)$ onto $\mathcal{B}^{p+1}_\flat(\Omega, \Gamma)$. Here $\oplus$ is the direct sum of spaces. Thus, loosely speaking, $C^p_\flat(\Omega, \Gamma)$ is divided into two spaces: the cocycles in $\mathcal{Z}^p_\flat(\Omega, \Gamma)$ and the preimage of (the nonzero elements of) the coboundary $\mathcal{B}^{p+1}_\flat(\Omega, \Gamma)$ via d.

The members of the space $C^p_\flat(\Omega, \Gamma)/i(\mathcal{Z}^p_\flat(\Omega, \Gamma))$ are cosets $[A]$ so that $d[A] = F$ is unique. Thus an ungauged electromagnetic potential is a coset in $C^1_\flat(\Omega, \Gamma)/i(\mathcal{Z}^1_\flat(\Omega, \Gamma))$, and the gauge is a mapping

$$g : C^1_\flat(\Omega, \Gamma)/i(\mathcal{Z}^1_\flat(\Omega, \Gamma)) \to C^1_\flat(\Omega, \Gamma), \tag{5.46}$$

such that g chooses a unique representative A from each coset $[A]$.

For wave problems, the best known gauge condition is the *Lorentz gauge*, which is described by

$$\mathrm{d}\,\mathcal{C} \star \mathcal{F} A = 0. \tag{5.47}$$

In differential form notation, with $a = \mathcal{F}A$, condition (5.47) simplifies into $\mathsf{d} \star a = 0$. Unfortunately, the Lorentz gauge is an incomplete gauge, because it does not fully fix A by itself, as seen in the following. Let A satisfy Lorentz gauge (5.47) and let f be a flat 0-form satisfying $\mathsf{d}\star\mathsf{d}f = 0$. Now, also $A+d\mathcal{C}f$ satisfies the Lorentz gauge condition (5.47) and produces the same F with (5.43). The condition $\mathsf{d} \star \mathsf{d}f = 0$ is, in fact, a homogenous wave equation for f. Notice that if we combine (5.43), (5.30), and (5.22) in differential form format, we get also a wave equation, namely $\mathsf{d} \star_M \mathsf{d}a = j$. Since the solution of a wave equation with proper boundary conditions is unique, we can fix f by applying boundary conditions.

Another possible gauge condition is the *temporal gauge* (or the *Weyl gauge*) [5, 35, 49]. The domain Ω must be covered by the charts of one observer (or a group of observers). In the coordinates of the observer(s), the electric scalar potentinal vanishes, i.e., $\Phi = 0$. Thus the gauge is heavily observer-dependent and, like the Lorentz gauge, also incomplete. To prove the latter, let B be an arbitrary spatial 0-chain. Then if A satisfies the temporal gauge condition, so does $A' = A+\mathrm{d}B$ [35]. In fact, $\mathrm{d}B$ describes the gauge freedom of A. Hence we could use an additional spatial three-dimensional gauge condition to fix B, e.g., *Coulomb's gauge* $\mathrm{d} \star \mathcal{A} = 0$.

All the above gauge conditions are so-called *metric gauge conditions*, because the Hodge operator takes part in all of these. In contrast, numerical computation embraces gauging techniques that do not depend on the metric, e.g., *the spanning-tree-extraction technique* [1]. We return to this topic in the following chapter.

5.4 Lower dimensional cases

Four-dimensional geometry and four-dimensional objects are on the edge of human capability to understand. In many situations, we must turn to lower-dimensional spacetimes for examples and interpretation. Additionally, in many situations symmetry allows us to model the problem in a lower-dimensional domain. Lower-dimensionality means that one or two spatial dimensions are suppressed so that only one or two spatial dimensions remain. These situations correspond to one- and two-dimensional time-dependent cases.

5.4.1 One spatial dimension: plane wave

Spacetime with the lowest possible dimensions has one spatial and one temporal dimension; i.e., it corresponds to a one-dimensional purely spatial manifold, where entities are time-dependent. In electromagnetics, this two-dimensional spacetime can be used, e.g., to model the TEM-planewave, where pointwise fields $\mathscr{e}$ and $\mathscr{b}$ both have only one component and, the components are perpendicular and, additionally, orthogonal to the direction of propagation.

In the above two-dimensional spacetime $\mathcal{E}$ and $\mathcal{B}$ are 1-cochains, returning nonzero values only on temporal and spatial 1-chains, respectively. Similarly also $\mathcal{D}$ and $\mathcal{H}$ (and $\mathcal{J}$) are 1-cochains. Thus in the observer-independent point of view, the Faraday and Maxwell fields F and G are 1-cochains, and, accordingly, A is a 0-cochain, i.e., a scalar potential.

The Maxwell-Faraday and Maxwell-Ampère laws reduce to laws on one-dimensional boundaries

$$(F|\partial S) = 0, \quad \forall S \in C_2^\flat(M) \tag{5.48}$$

for $F \in C_\flat^1(M)$ and

$$(G|\partial \tilde{S}) = (J|\tilde{S}), \quad \forall \tilde{S} \in \tilde{C}_2^\flat(M). \tag{5.49}$$

for $G \in \tilde{C}_\flat^1(M)$ and $J \in \tilde{C}_\flat^2(M)$. Similarly, the potential $A \in C_\flat^0(M)$ in the two-dimensional case must fulfil

$$(A\,|\partial\Sigma) = (F\,|\,\Sigma), \quad \forall \Sigma \in C_1^\flat(M). \tag{5.50}$$

The constitutive equation is of the same form as (5.20), but now f and $g = \star_M f$ are 1-forms, v is a 1-vector, and $\star$ is the Hodge-operator of a two-dimensional Lorentz manifold acting on a 1-vector. Further, interface and boundary conditions are placed on a curve in a two-dimensional spacetime manifold. Because the potential A is a scalar potential, it is made unique by proper boundary conditions with no gauging necessary. This is a major benefit in the two-dimensional electromagnetic wave case. Figure 5.7 shows the prototype problem geometry.

5.4.2 Two spatial dimensions: long objects

Another possible lower-dimensional spacetime has two spatial and one temporal dimensions; i.e., in the electromagnetic case it models a two-dimensional time-dependent problem. This three-dimensional spacetime can model, e.g., the radiation of a uniformly fed very long straight antenna. The radiation

pattern is now modeled to be independent of the axial direction of the antenna. By symmetry, the same model fits also the case of a straight antenna between two metal layers made from good conductors, i.e., the case inside a parallel plate waveguide. In these examples, $\mathscr{e}$ has one component—along the axis of the source antenna—and $\mathscr{b}$ has two components—on the perpendicular plane. For this reason, we choose such a three-dimensional spacetime model that spatially everything happens on a two-dimensional plane with $\mathscr{e}$ perpendicular to the plane and $\mathscr{b}$ parallel to the plane. Also possible is the dual case where $\mathscr{h}$ is perpendicular to the plain and $\mathscr{d}$ has two components on the plane, but that is not our concern in this work.

In the above three-dimensional spacetime, $\mathcal{E}$ and $\mathcal{B}$ are 1-cochains, as in the two-dimensional case, but $\mathcal{D}$ and $\mathcal{H}$ are now 2-cochains. $\mathcal{D}$ returns zero on every plane made by extrusion in the time direction, and $\mathcal{H}$ has a null value on every purely spatial 2-chain. Accordingly, F and G are a 1-cochain and 2-cochain, respectively, and the potential A is yet again a 0-cochain, i.e., a scalar potential. The above dual case would have F as a 2-cochain, and thus A would be a 1-cochain potential that, like the four-dimensional case, would need gauge fixing.

Because F is now a 1-cochain and A a 0-cochain, the situation resembles the two-dimensional case. The difference lies in the Maxwell-Ampère law, which is now placed on a 2-boundary, as in the four-dimensional case (5.30). In fact, everything related to straight cochains is similar to the two-dimensional spacetime case and everything related to twisted cochains resembles the four-dimensional case. Objects connected with the constitutive law are specific to this three-dimensional spacetime case. The constitutive law has, again, the same form as (5.20), but now f is a 1-form and $g = \star_M f$ is a 2-form, v is a 2-vector, and $\star$ is the Hodge-operator of three-dimensional Lorentz manifold acting on a 1-vector. Boundary and interface conditions are placed on surfaces for the 1-cochain F (or for the 0-cochain A). The prototype problem geometry is the three-dimensional counterpart of figure 5.7, and, as in the two-dimensional spacetime case, no gauging is needed.

Two-dimensional waves are troublesome, because a point source with impulse time dependence does not create a sharp spherical wave front even on a homogenous and isotropic lossless medium. The wave front is always followed by a tail of exponentially diminishing fields. This is because in the equivalent three-dimensional case, the source is infinitely long, and on a perpendicular plane the wave is thus generated, apart from the source point on the plane, from every point on the infinite source axis. The wavefronts from off-plane source points arrive later and create a tail to the wave front.

Chapter 6

Numerical electromagnetic computation in spacetime

Now we turn our attention to our main application, numerical electromagnetic computation in spacetime. In the above chapters, we introduced mathematical and physical tools for the computation. Therefore, it is quite straightforward to write our formulation for electromagnetic computation.

In the first section below, we introduce a numerical method but illustrate it fully somewhat later. Second, we demonstrate the method with fairly easy two- and three-dimensional spacetime problems. After we have gained some insight into the computation, we turn to our primary case, computation in four-dimensional spacetime. Our numerical method is implemented with MATLAB®.

Even if the numerical demonstrations are mainly carried out with meshes constructed from four-dimensional bricks, the method itself can tolerate more general meshes. The austerity of the validation was due to the lack of proper four-dimensional mesh generators. After all, like the whole thesis project, also the implementation was a one man project with a limited schedule.

6.1 Formulation of computational electromagnetics in spacetime

For numerical computation, we need to formulate the electromagnetic problem with a finite number of variables and equations. For the Maxwell laws, the finite counterparts are easy to find, because we have already introduced the finite cochain and chain spaces: we simply restrict the Maxwell equations to these finite subspaces. In addition, the finite counterpart of the electromagnetic potential is considered in the first section.

In contrast, finding the finite counterparts for the constitutive relation requires some work. In fact, choosing the finite constitutive relation is a major modeling decision, which depends on the method to be used. In the second section, we introduce the constitutive relation of geometric methods with an orthogonal pair of staggered meshes.

In the end, we consider how boundary conditions are imposed on the finite problem. Additionally, according to these conditions, we must restrict the space where the finite Maxwell laws are imposed. In this section, we assume that spacetime is four-dimensional—apart from a few graphical illustrations—even if the first examples in the following section are lower-dimensional.

6.1.1 Finite Maxwell laws

In the above chapters, we formulated our electromagnetic computation problem as follows. For a known $J \in \tilde{C}^3_\flat(M)$, find $F \in C^2_\flat(M)$ and $G \in \tilde{C}^2_\flat(M)$ so that

$$
\begin{aligned}
(F|\partial\Sigma) &= 0, & \forall\Sigma \in C^\flat_3(M), && (6.1)\\
(G|\partial\tilde{\Sigma}) &= (J|\tilde{\Sigma}), & \forall\tilde{\Sigma} \in \tilde{C}^\flat_3(M), && (6.2)\\
g &= \star_M f && (6.3)
\end{aligned}
$$

are satisfied, where $f \in F^2_\flat(M)$ and $g \in \tilde{F}^2_\flat(M)$ are such that

$$
\begin{aligned}
f &= \mathcal{F}F, && (6.4)\\
G &= \mathcal{C}g. && (6.5)
\end{aligned}
$$

Alternatively, in our topologically simple case, the Maxwell-Faraday law (6.1) can be replaced by the potential equation, i.e.,

$$(A\,|\partial\Sigma) = (F\,|\,\Sigma), \quad \forall\Sigma \in C^\flat_2(M). \tag{6.6}$$

Now the problem is to solve $A \in C^1_\flat(M)$ so that (6.2)-(6.6) are satisfied.

We want to solve the problem (6.1)-(6.5) (or equally (6.2)-(6.6)) numerically. We must then have a finite number of unknows and equations. For Maxwell equations (6.1) and (6.2), the finite counterparts are easy to find: we merely replace the chain and cochain spaces with finite ones. For $F \in C^2(K)$, $G \in \tilde{C}^2(\tilde{K})$ and the known $J \in \tilde{C}^3(\tilde{K})$, we get

$$
\begin{aligned}
(F|\partial\Sigma) &= 0, & \forall\Sigma \in C_3(K), && (6.7)\\
(G|\partial\tilde{\Sigma}) &= (J|\tilde{\Sigma}), & \forall\tilde{\Sigma} \in \tilde{C}_3(\tilde{K}). && (6.8)
\end{aligned}
$$

Notice that we need two cellular mesh complexes on M: a inner oriented K for F and an outer oriented $\tilde{K}$ for G and J.[1] Thus after we have build K and $\tilde{K}$, we have the proper finite counterparts of the Maxwell equations. For the potential equation (6.6), the finite version is

$$(A\,|\partial\Sigma) = (F\,|\,\Sigma), \quad \forall\Sigma \in C_2(K) \tag{6.9}$$

for a $A \in C^1(K)$.

In section 4.4.5, we noticed that by choosing a basis for the finite cochain spaces $C^p(K)$, with $p = 0, \ldots, n$, allows an arbitrary $\Omega \in C^p(K)$ to be represented by a real array $\mathbf{\Omega} \in \mathbb{R}^{\kappa_p}$, where κ_p is the number of elements in the basis of $C^p(K)$. For $\tilde{C}^p(\tilde{K})$, the situation is similar—now the basis has $\tilde{\kappa}_p$ elements. At the same time, we get a real array representative (matrix) $\mathbf{d}_p$ for the exterior derivative restricted to the finite cochain space $C^p(K)$ and $\tilde{\mathbf{d}}_p$ for $\tilde{C}^p(\tilde{K})$. If now $\boldsymbol{F} \in \mathbb{R}^{\kappa_2}$, $\boldsymbol{G} \in \mathbb{R}^{\tilde{\kappa}_2}$, $\boldsymbol{J} \in \mathbb{R}^{\tilde{\kappa}_3}$, and $\boldsymbol{A} \in \mathbb{R}^{\kappa_1}$ are the real array counterparts of F, G, J, and A, equations (6.7)–(6.9) can be written

$$\mathbf{d}_2\boldsymbol{F} = 0, \tag{6.10}$$

$$\tilde{\mathbf{d}}_2\boldsymbol{G} = \boldsymbol{J}, \tag{6.11}$$

and

$$\mathbf{d}_1\boldsymbol{A} = \boldsymbol{F}. \tag{6.12}$$

6.1.2 Finite constitutive relations

A more difficult task is to obtain a finite counterpart for the constitutive equation. In fact, our main problem is not the constitutive relation (6.3) but rather the mappings (6.4) and (6.5) between the chain and cochain spaces. In the finite Maxwell laws, i.e., (6.7) (or (6.9)) and (6.8), we deal only with cochains over the chains of the mesh complexes K and $\tilde{K}$. That is, we know only the values of F (or A) and G on a finite number of cells. Because mappings $\mathcal{F}$ and $\mathcal{C}$ need a sequence of diminishing cells, it is not straightforward or sometimes even possible to work out the pointwise values of the differential forms f and g, even if we know $F \in C^2(K)$ and $G \in \tilde{G}^2(\tilde{K})$ accurately.

There are several ways to deal with the finite counterparts of constitutive equations. In the *finite element method*, Whitney interpolators are used

[1] In certain versions of the three-dimensional geometric method (e.g., methods with a barycentric dual), the dual mesh do not form a cellular mesh complex. However, the dual mesh yet forms a chain complex, and thus this causes only minor changes in the following discussion.

for pointwise values, but as mentioned in chapter 2, this method is not the optimal for us. Instead, we use a *geometric method*, where we use only the first elements of the sequences that define $\mathcal{F}$ and $\mathcal{C}$.

Approximative relation

How could we transform the mapping $\mathcal{C}$, i.e., a map from flat forms to flat cochains, into a finite case? Let us start with the definition of $\mathcal{C}$ in section 4.6.1. There we reduced the definition to the integration of a differential form ω' over a convex p-cell c in $\mathbb{R}^n$. We arrived at the statement

$$\int_c \omega' = \lim_{k\to\infty} \sum_{s_i \in \mathfrak{S}_k c} (\omega'(\boldsymbol{s}_i)|\{s_i\}), \tag{6.13}$$

where $\mathfrak{S}_k c$ is the simplex subdivision of c, $\boldsymbol{s}_i$ is the barycenter of the convex simplex s_i and $\{s_i\}$ is the p-vector corresponding to s_i. We choose to approximate the integral with the first element of sequence (6.13), i.e.,

$$\int_c \omega' \approx (\omega'(\boldsymbol{c})|\{c\}), \tag{6.14}$$

where $\boldsymbol{c}$ is the barycenter of c. If we use the above approximation in the definition of $\mathcal{C}$, we get a finite version $\mathcal{C}_f$. Let us define $\mathcal{C}_f$ precisely. We only need to define $(\mathcal{C}_f\omega|\sigma)$ for every p-cell σ in K, because then $\mathcal{C}_f\omega$ on an arbitrary p-chain $\Sigma \in C_p(K)$ follows from linearity. For this, we use pullback (4.56) and approximation (6.14) to integrate ω over the cell σ. If we denote $\{\sigma\} = (\chi^{-1})_*\{\chi(\sigma)\}$ with the pushforward $(\chi^{-1})_*$, $\mathcal{C}_f\omega$ is defined as follows:

$$(\mathcal{C}_f\omega|\sigma) = (\omega(x_\sigma)|\{\sigma\}), \tag{6.15}$$

where $x_\sigma = \chi^{-1}\boldsymbol{x}_\sigma$, if $\boldsymbol{x}_\sigma$ is the barycenter of $\chi(\sigma)$ [74].

Similarly, we define an approximation $\mathcal{F}_f$ for the mapping $\mathcal{F}$ by taking the first element of the sequence in (4.58),

$$(\mathcal{F}_f\Omega(x)|v) = (\Omega|\sigma_v), \tag{6.16}$$

where σ_v is a p-cell corresponding to the simple p-vector v, as in section 4.6.2. Because we use $\mathcal{F}_f$ only for the finite dimensional chain $\Omega \in C^p(K)$, σ_v must be included in the cell complex K. Thus for $\sigma \in K$, (6.16) reads

$$(\mathcal{F}_f\Omega(x_\sigma)|\{\sigma\}) = (\Omega|\sigma). \tag{6.17}$$

Fortunately, the approximative mappings $\mathcal{C}_f$ and $\mathcal{F}_f$ are inverses of each other, like their exact counterparts $\mathcal{F}$ and $\mathcal{C}$.

By substituting (6.15) and (6.17) as well as (6.3) with definition (5.24) for mappings (6.4) and (6.5), we get

$$(f|\{\sigma\}) \;=\; (F|\sigma), \qquad \forall \sigma \in K, \tag{6.18}$$

$$(G|\tilde{\sigma}) \;=\; (f|\star_M \{\tilde{\sigma}\}), \qquad \forall \tilde{\sigma} \in \tilde{K}. \tag{6.19}$$

In order to obtain a solvable numerical formulation, we need a connection between $(F|\sigma)$ and $(G|\tilde{\sigma})$. Therefore, we must choose the mesh complexes K and $\tilde{K}$ so that we have a relation between $\{\sigma\}$ and $\star_M\{\tilde{\sigma}\}$ with certain requirements.

For now, for simplicity, let us assume that our domain is a vacuum, i.e., $\varepsilon_r = \mu_r = 1$. With (5.20), $\star_M = \sqrt{\frac{\varepsilon_0}{\mu_0}}\star$. So we need a relation between $\{\sigma\}$ and $\star\{\tilde{\sigma}\}$. If we want an efficient numerical method, this relation needs to be a linear function. In the most straightforward and feasible method, we choose $\{\sigma\}$ and $\star\{\tilde{\sigma}\}$ as identical except for norm, i.e., $x_\sigma = x_{\tilde{\sigma}}$ and

$$\frac{\star\{\tilde{\sigma}\}}{|\tilde{\sigma}|_m} = \frac{\{\sigma\}}{|\sigma|_m}, \tag{6.20}$$

because $\|\{\tilde{\sigma}\}\| = \| \star \{\tilde{\sigma}\}\|$ and $|\sigma|_m = \|\{\sigma\}\|$. By combining (6.18), (6.19) and (6.20), we get

$$(G|\tilde{\sigma}) = \sqrt{\frac{\varepsilon_0}{\mu_0}}\frac{|\tilde{\sigma}|_m}{|\sigma|_m}(F|\sigma). \tag{6.21}$$

The above approximation of the constitutive relation is very simple; it is based on averaging the cochain fields over the 2-cells, which yields us pointwise approximates of the fields. Consequently, the approximation is the better, the smaller the 2-cells—i.e., the denser the mesh. Accordingly, unformally speaking, we can make this approximate as accurate as necessary by taking a dense enough mesh. For equation (6.21), we need only the mass of the 2-cells in question, nothing else.

The finite Maxwell laws (6.7) (or (6.9)) and (6.8) together with (6.21) gives us a numerical algorithm for the electromagnetic wave problem in a vacuum. Additionally, we need only the boundary (and initial) conditions, which are assigned similarly to the continuous case in section 5.2.

Let us return to the general medium, which is somewhat more complicated because $\|\{\tilde{\sigma}\}\| \neq \| \star_M \{\tilde{\sigma}\}\|$. Consequently, we define

$$\kappa'(\star_M, \{\tilde{\sigma}\}) = \frac{\| \star_M \{\tilde{\sigma}\}\|}{\|\{\tilde{\sigma}\}\|}. \tag{6.22}$$

The above κ' is a function of $\{\tilde{\sigma}\}$ and $\star_M$; i.e., in a fixed frame it is a function of ε_r and μ_r. The material parameters ε_r and μ_r are evaluated at

the barycenter of $\tilde{\sigma}$. If we divide $v \in T_2^x$ into a purely spatial part v_1 and a part in the time direction v_2, we have $v = v_1 + v_2$ and

$$\kappa'(\star_M, v) = \frac{\| \star_M v\|}{\|v\|} = \sqrt{\frac{\varepsilon_0}{\mu_0}} \frac{\left|\varepsilon_r^2 \|v_1\|^2 - \frac{1}{\mu_r^2}\|v_2\|^2\right|^{\frac{1}{2}}}{\|v\|}. \tag{6.23}$$

In a vacuum, $\varepsilon_r = \mu_r = 1$ and $\kappa' = \sqrt{\varepsilon_0/\mu_0}$. In the general case, when v is purely spatial, we have $\kappa' = \sqrt{\varepsilon_0/\mu_0}\,\varepsilon_r$. Similarly, if v is in the time direction, we have $\kappa' = \sqrt{\varepsilon_0/\mu_0}\,\frac{1}{\mu_r}$. If v does not fall into either of the above categories, we must use equation (6.23).

With $\kappa' = \kappa'(\star_M, \{\tilde{\sigma}\})$ we have a counterpart for equation (6.20),

$$\frac{\star_M\{\tilde{\sigma}\}}{\kappa'|\tilde{\sigma}|_m} = \frac{\{\sigma\}}{|\sigma|_m}. \tag{6.24}$$

This requires that $\{\sigma\}$ and $\star_M\{\tilde{\sigma}\}$ be equal after normalization. Also in this case, $x_\sigma = x_{\tilde{\sigma}}$. By substituting (6.24) for (6.18) and (6.19), we conclude

$$(G|\tilde{\sigma}) = \kappa' \frac{|\tilde{\sigma}|_m}{|\sigma|_m}(F|\sigma). \tag{6.25}$$

This equation closely resembles the vacuum case; in fact, in a vacuum (6.25) reduces to equation (6.21). Likewise, loosely speaking, the approximation error diminishes when the mesh is made denser. For the finite constitutive law (6.25), we need the values of the material parameters at the common barycenter of σ and $\tilde{\sigma}$, in addition to the masses of the 2-cells.

Because the finite constitutive relation (6.25) includes a ratio of the mass norms of the 2-cells, we do not want 2-cells with the zero norm. As mentioned in section 4.3.4, this happens on lightlike submanifolds. Therefore, we discard meshes that have 2-cells that lie on a two-dimensional lightlike manifold. Such a 2-cell would be Lorentzian orthogonal to itself.

The finite Maxwell laws (6.7) (or (6.9)) and (6.8) together with (6.25) and boundary conditions give us a numerical algorithm for the electromagnetic wave problem in the presence of a non-vacuum medium.

Primal and dual meshes

Orthogonality of meshes

To use the above algorithm, we must build two complexes K and $\tilde{K}$ so that each 2-cell σ in K has a *dual* counterpart in $\tilde{K}$, and so that $\{\sigma\}$ and $\star_M\{\tilde{\sigma}\}$ are equidirectional, and that $x_\sigma = x_{\tilde{\sigma}}$. This means that $\{\sigma\}$ and $\{\tilde{\sigma}\}$ are

orthogonal in terms of $\star_M$, and that their centers coincide. However, the orthogonality of $\{\sigma\}$ and $\{\tilde{\sigma}\}$ is not enough, because if $\{\sigma\}$ and $\{\tilde{\sigma}\}$ are orthogonal, also $\{-\tilde{\sigma}\}$ is orthogonal to $\{\sigma\}$; we must require equidirectionality. Nevertheless, we consider orthogonality first and deal with the signs later in the context of orientations.

In a vacuum, orthogonality means Lorentzian orthogonality, because the Hodge operator $\star$ is equivalent to the scalar product of a Lorentzian manifold. If we denote $v = \{\tilde{\sigma}\}$ and use the above notation, in the material case $\star_M$-orthogonality means that $\varepsilon_r v_1 + \frac{1}{\mu_r} v_2$ is Lorentzian orthogonal to $\{\sigma\}$. Thus in the general case, $\star_M$-orthogonality corresponds to Lorentzian orthogonality, where material parameters are taken into account. We call this $\star_M$-orthogonality *electromagnetic orthogonality*. Notice that electromagnetic orthogonality equals the Lorentzian one in the vacuum case—and also when $\{\tilde{\sigma}\}$ is purely spatial or, equally, if it is in the time direction.

Remark 6.1. The concept of electromagnetic orthogonality is related to the so-called *optical metric* (or effective metric), first introduced by W. Gordon in [27]. The optical metric differs from the Lorentzian metric in its choice of the speed of light; it uses the speed of light in the local dielectric medium instead of in a vacuum.

In comparison, the Lorentzian metric is based on the vacuum speed regardless of the effects of the medium on the speed of light—or, more generally, on the speed of an electromagnetic wave. Thus the geodesic of a light ray is always lightlike in the case of optical metrics. With the Lorentz metric, the rays can also be timelike in dielectric materials. However, the optical metric also has a Lorentzian signature.

Anisotropic or dispersive materials are problematic for the optical metric, because the propagation speed of an electromagnetic wave in such a medium depends then on the traveling direction or frequency of the wave. In these cases, the optical metric becomes unusable. Apart from these cases, the optical metric could be used instead of the Lorentzian one, and we would have a more straightforward approach, because all the effects of the medium would then be on the metric.

The most vital aspect of the optical metric is that it changes the concepts of causality—e.g., an optically lightlike curve can be Loretzially merely timelike. With the optical metric, the division between boundary and initial conditions by causality would perhaps be more real. Similarly, the measure on a manifold should be related to this metric instead of the Lorentzian. We leave open the question of the effect of the above on the formulation.

As mentioned above, the optical metric is related to electromagnetic orthogonality: $\star_M$ is the Hodge-operator inherited from the optical metric;

moreover, electromagnetic orthogonality equates with the concept of orthogonality in this metric.

Mesh complexes

One important question remains: where do we get the mesh complexes K and $\tilde{K}$? They must fill the spacetime domain $\Omega \subset M$ so that every 2-cell in K has a dual 2-cell counterpart in $\tilde{K}$; i.e., the 2-cells are electromagnetically orthogonal and their centers coincide. For simplicity, let us first assume that our domain Ω has no boundaries. In a vacuum, and also in a purely spatial or time-directional cell, the dual meshes of the geometric methods—i.e., the meshes of the finite integration technique and the cell method—fulfill our requirements. In three-dimensional time-dependent cases, geometric methods have two cuboidal three-dimensional meshes, one primal mesh and one dual. Each 2-cell (or each *facet*) of the primal mesh has a unique dual 1-cell (or a dual *edge*) orthogonal to the primal facet. Similarly, each dual facet has an orthogonal primal edge counterpart. Nothing forbids us to generalize these meshes to four-dimensional spacetime. Furthermore, they can also be generalized to the non-vacuum case.

In our method, K is called the *primal mesh complex* and $\tilde{K}$ the *dual mesh complex*. Each facet f of K is electromagnetically orthogonal to a dual facet $\tilde{f} \in \tilde{K}$ and vice versa. Similarly, each primal edge $e \in K$ has an orthogonal dual 3-cell (i.e., a *volume*) $\tilde{v} \in \tilde{K}$ and, further, a primal volume has an orthogonal dual edge. Further, each 0-cell (or each *node*) has a dual 4-cell (i.e., a *hypervolume*) counterpart. In our case, because orthogonality is needed only in the constitutive relation, orthogonality of the primal facet–dual facet -pair is the only relevant requirement (see figure 6.1 for a two-dimensional example).

Additionally, we require that if a $(p-1)$-cell $\tau \in K$ lies on the boundary of a p-cell $\sigma \in K$, then $\tilde{\sigma}$ lies on the boundary of $\tilde{\tau}$, where $\tilde{\sigma} \in \tilde{K}$ is the $(n-p)$-cell dual to σ, and $\tilde{\tau} \in \tilde{K}$ is the $(n-p+1)$-cell dual to τ. That is, if τ is a *face* of σ, then $\tilde{\sigma}$ is a face of $\tilde{\tau}$.

A second additional requirement is that the outer orientation of the dual $(n-p)$-cell $\tilde{\sigma}$ be determined by the inner orientation of the primal p-cell σ. This is possible, because by the definition of cell orientation (section 4.3.1), orientations can be compared via two charts, and because the outer orientation of $\tilde{\sigma}$ and the inner orientation of σ are defined by the inner orientation of a p-dimensional affine space. By the orthogonality of $\{\tilde{\sigma}\}$ and $\{\sigma\}$, the orientations of the affine spaces can be compared with a method similar to the equation (4.11).

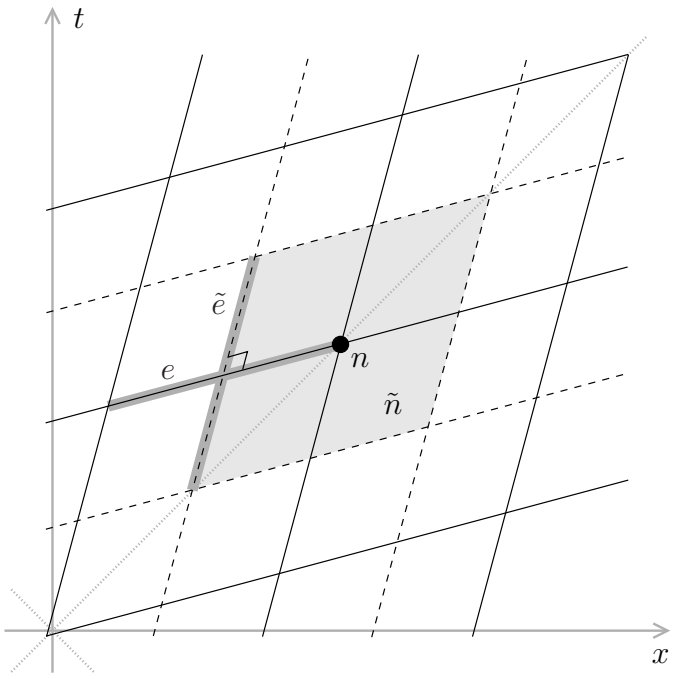


Figure 6.1: Primal and dual meshes in a flat two-dimensional spacetime. The primal mesh is plotted with a solid line and the dual with a dashed line. The primal edge e and the dual edge $\tilde{e}$ are duals to each other and orthogonal in Minkowski geometry. Similarly, the primal node n is dual to the dual facet $\tilde{n}$ (gray area). The finite constitutive relation of the electromagnetic wave problem in two-dimensional spacetime is needed for the e–$\tilde{e}$-pair, whose orthogonality is thus the most crucial one.

Equidirectionality of meshes

The above choice of orientations does not take into account the equidirectionality of $\{\sigma\}$ and $\star_M\{\tilde{\sigma}\}$. Consequently, because the relation (6.24) between a primal and dual 2-cell is not valid for every σ–$\tilde{\sigma}$-pair of mesh complexes, we must modify (6.25) accordingly. By the definition of Hodge operator in section 4.5.3, the orientations agree with (6.24) if $\star_M\{\tilde{\sigma}\}$ is timelike and disagrees if it is spacelike. Therefore, we define

$$\kappa(\star_M, \{\tilde{\sigma}\}) = \begin{cases} \kappa'(\star_M, \{\tilde{\sigma}\}), & \text{if } \{\tilde{\sigma}\} \text{ timelike,} \\ -\kappa'(\star_M, \{\tilde{\sigma}\}), & \text{if } \{\tilde{\sigma}\} \text{ spacelike.} \end{cases} \tag{6.26}$$

As explained before, a lightlike case is forbidden. In our primal and dual meshes oriented in the above manner, (6.25) changes into

$$(G|\tilde{\sigma}) = \kappa \frac{|\tilde{\sigma}|_m}{|\sigma|_m} (F|\sigma). \tag{6.27}$$

We have two important special cases of (6.27),

$$(G|\tilde{\sigma}) = \begin{cases} \varepsilon_r \sqrt{\frac{\varepsilon_0}{\mu_0}} \frac{|\tilde{\sigma}|_m}{|\sigma|_m} (F|\sigma) & \text{if } \tilde{\sigma} \text{ purely spatial,} \\ -\frac{1}{\mu_r} \sqrt{\frac{\varepsilon_0}{\mu_0}} \frac{|\tilde{\sigma}|_m}{|\sigma|_m} (F|\sigma) & \text{if } \tilde{\sigma} \text{ in the time direction.} \end{cases} \tag{6.28}$$

Construction of meshes

Spacetime meshing has been studied over the last few years [19, 78], but the meshes have been made for the discontinuous Galerkin method. Thus they have only one mesh instead of a pair of dual meshes. In most cases, it is impossible to add an orthogonal dual mesh afterwards to a ready-made (primal) mesh. Therefore, since involved mesh generation is beside our topic, we concentrate on easily constructible meshes. In the most austere case, the 4-cells of both meshes are *hypercubes*, also called *tesseracts*. A hypercube is a four-dimensional generalization of a cube; i.e., all of its faces are cubes. However, since we can use hypercube-meshes only in some rare cases, we must at least be able to stretch cubes along their axial direction. Like a cuboid is a stretched cube, a *hypercuboid* is a stretched hypercube. This could also be generalized to a free-form *hyper-octahedron*, but we keep to hypercuboidal meshes. However, we take some freedom with the concept of orthogonality, because we allow the adjacent faces of a hypercuboid to be orthogonal in the Lorentzian metrics, not in the customary Riemannian way. Additionally, electromagnetic orthogonality is also possible.

Constructing a pair of hypercuboidal dual meshes—often called staggered meshes—is straightforward in a vacuum and with isotropic materials. With anisotropic materials, it is also simple, if we can choose the mesh axes along the anisotropy axes. If that is not possible, the material parameters must be taken into account in mesh construction via the required electromagnetic orthogonality between primal and dual faces. Of course, if we can choose a metric so that we get rid of the anisotropy [55], this case turns dramatically simpler. Notice that the spacetime constitutive relation (6.28) is inheritedly anisotropic, because we have different material parameters for purely spatial and time directions.

The whole concept of electromagnetic orthogonality becomes unusable with a dispersive anisotropic material, because material parameters are then

frequency-dependent. In this case, the only possibility is to choose mesh faces so that they are either along or orthogonal to the anisotropy axes. This is not a significant limitation, because in any case it is difficult to deal with dispersive materials in the context of non-timeharmonic problems. In these problems, constitutive relations must embrace the whole history of the fields, rather than present values only. Consequently, we do not discuss the dispersive medium any further in this work, and avoid thus any insuperable problems with electromagnetic orthogonality.

In the following sections, we give a few examples of mesh complexes and explain in depth how to realize discrete constitutive relations.

Notice that the finite Maxwell laws (6.7) (or (6.9)) and (6.8) are exact—they do not produce any additional inaccuracy in numerical computation, except for floating point errors. Only the finite constitutive law (6.27) generates inaccuracy in results with errors inherited from the fact that only the first elements of sequences (6.13) and (4.58) were included. Error diminishes when the mesh is made denser.

Because of the wide scope of error-bound analysis, we cannot discuss convergence and stability results any further here (see [74] for some results). Moreover, since our cochain spaces do not hold a proper norm, but a plain topology (or uniform structure), error-bound analysis cannot be done in the standard way with norms. However, since convergence and stability are key issues development of a new formulation, their omission is quite unfortunate. Exploration of this subject could be one of the most important possible continuations of this work.

Real array counterparts

Let us look for a real array counterpart to the finite constitutive relation. Because of the duality of the mesh complexes K and $\tilde{K}$, notation can be simplified. The basis of $C^p(K)$ is connected to the basis of the finite chain space $C_p(K)$ by equation (4.31). Further, we use the canonical basis of $C_p(K)$, composed of the p-cells of K. Additionally, because of duality there are as many dual $(n-p)$-cells as primal p-cells, and thus $\kappa_p = \tilde{\kappa}_{n-p}$.

Moreover, since $\tilde{\mathbf{d}}_p = \tilde{\boldsymbol{\partial}}_{p+1}^T$, where $\tilde{\boldsymbol{\partial}}_{p+1}$ is the real counterpart of the boundary operator of $\tilde{K}$, and, further, since the boundary operator of $\tilde{K}$ is connected to the boundary operator of K, $\tilde{\mathbf{d}}_p$ can also be simplified. If $[\boldsymbol{\partial}_{n-p}]_{ij} = 1$, the i^{th} basis $(n-p-1)$-cell lies on the boundary of the j^{th} basis $(n-p)$-cell and their orientations match. If the dual basis cells are arranged in accordance with the primal cells, we have also $[\tilde{\boldsymbol{\partial}}_{p+1}]_{ji} = 1$, because the outer orientation of a dual cell is determined by the inner orientation of the corresponding primal cell. Continuing this logic, we get $\tilde{\boldsymbol{\partial}}_{p+1} = \boldsymbol{\partial}_{n-p}^T$ and,

consequently, $\tilde{\mathbf{d}}_p = \mathbf{d}_{n-p-1}^T$. Especially in four-dimensional spacetime, the exterior derivative (incidence matrix) of (6.11) $\tilde{\mathbf{d}}_2$ changes to $\mathbf{d}_1^T$.

It is quite straightforward to represent the finite constitutive relation (6.27) with real arrays. Since the equation is applied to all 2-cells in K and their dual counterparts, and since the dual basis cells $\{\tilde{\sigma}_i\}_i$ are numbered by the numbering of the primal basis $\{\sigma_i\}_i$, we have $[\boldsymbol{F}]_i = (F|\sigma_i)$ and $[\boldsymbol{G}]_i = (G|\tilde{\sigma}_i)$.[2] Thus we can build a *diagonal material matrix* $\mathbf{H}_M \in \mathbb{R}^{\kappa_2 \times \kappa_2}$ so that

$$[\mathbf{H}_M]_{ii} = \kappa(\star_M, \{\tilde{\sigma}_i\}) \frac{|\tilde{\sigma}_i|_m}{|\sigma_i|_m}. \tag{6.29}$$

Equation (6.27) can now be written in real array form

$$\boldsymbol{G} = \mathbf{H}_M \boldsymbol{F}. \tag{6.30}$$

If we now want to have a potential-based numerical electromagnetic wave problem, we combine (6.11), (6.12), and (6.30). The problem then takes the form: for a known $\boldsymbol{J} \in \mathbb{R}^{\kappa_1}$, find a $\boldsymbol{A} \in \mathbb{R}^{\kappa_1}$ so that

$$\mathbf{d}_1^T \mathbf{H}_M \mathbf{d}_1 \boldsymbol{A} = \boldsymbol{J}, \tag{6.31}$$

and the boundary conditions are satisfied.

6.1.3 Boundary conditions

To impose boundary conditions on a finite electromagnetic problem, we must discuss first how a dual mesh is built on the boundary. After all, we did not consider boundaries in the above discussion. Then we can determine boundary conditions by using proper relative cochain spaces.

Dual mesh on a boundary

Boundary mesh

We need to consider how a dual mesh is built on the boundary. The primal mesh complex K covers the whole domain Ω, and its subcomplex covers the boundary $\partial\Omega$—at least approximatively. In contrast, the elements of $\tilde{K}$ are transversal to the primal ones. Now, should the boundary cells of the dual complex go over the boundary $\partial\Omega$? Moreover, we want to impose a boundary and initial conditions for both F and G on certain parts of $\partial\Omega$. However, this is not possible, if no elements of $\tilde{K}$ lie on $\partial\Omega$. Consequently, we must modify

[2] Here $\{\tilde{\sigma}_i\}_i$ does not denote a multivector corresponding to $\tilde{\sigma}_i$, but the (ordered) set of basis cells.

the above definition of dual mesh on the boundary. Additionally, we must solve another problem: if the dual—as described in the above section—is a real cellular mesh complex in the bounded case, it must have extra cells that are not dual to any primal cell (see figure 6.2, left).

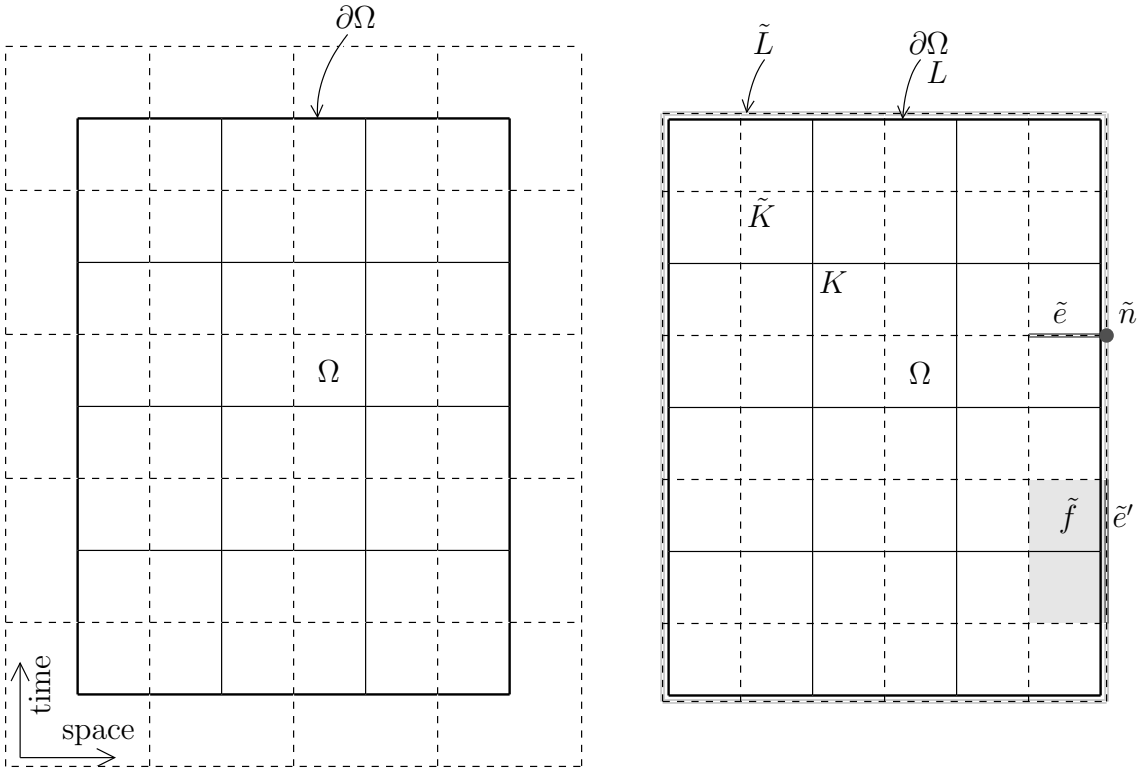


Figure 6.2: Primal and dual meshes in a two-dimensional spacetime domain Ω with a boundary. The horizontal axis is the only spatial dimension and the vertical one is time. The domain is plotted with a thick solid line, the primal mesh with a thin solid line, and the dual one with a dashed line. Left: a dual mesh is constructed without the boundary taken into account. The primal mesh fills the domain Ω, but the dual cells go outside the domain. The dual cells on the boundary of the dual mesh are not dual to any primal cell. Right: The dual mesh outside Ω is truncated into a mesh complex on the boundary. The boundary complex is denoted by $\tilde{L}$ and highlighted in gray. For illustration, the primal and dual meshes L and $\tilde{L}$ on the boundary $\partial\Omega$ are plotted separately; in reality they are one on the other. The 1-cell $\tilde{e}$ now ends at $\tilde{n}$ of $\tilde{L}$ and the 2-cell $\tilde{f}$ has $\tilde{e}' \in \tilde{L}$ on its boundary.

First, we construct a dual mesh complex for K as described in the above

section. Then we truncate the portion of the dual complex $\tilde{K}$ outside K—i.e., outside Ω. This procedure adds a $(p-1)$-cell to the location where a p-cell penetrates $\partial\Omega$ (see figure 6.2, right). Let us denote this new dual complex $\tilde{K}'$ temporarily. For example, a 1-cell $\tilde{e} \in \tilde{K}'$, which originally passed through $\partial\Omega$, now ends at a new 0-cell $\tilde{n} \in \tilde{K}'$ on $\partial\Omega$. Similarly, a 2-cell $\tilde{f} \in \tilde{K}'$ has a new 1-cell $\tilde{e}' \in \tilde{K}'$ on its boundary, and this $\tilde{e}'$ lies on $\partial\Omega$, as in figure 6.2, right. These new dual cells on $\partial\Omega$ form a cellular mesh complex $\tilde{L}$ so that $\tilde{L} \subset \tilde{K}'$. We adopt new dual chain spaces $\tilde{C}_p(\tilde{K}')$ (for $p = 0, \dots, n$), and $\tilde{C}_p(\tilde{L})$ (for $p = 0, \dots, n-1$), and with a boundary operator we construct finite chain complexes $\tilde{C}_\times(\tilde{K}')$ and $\tilde{C}_\times(\tilde{L})$ [40, 68]. Henceforth, we drop the apostrophe from $\tilde{K}'$.

Relative cochains

The space of relative dual p-chains modulo $\partial\Omega$ is defined to be $\tilde{C}_p(\tilde{K}, \tilde{L}) = \tilde{C}_p(\tilde{K})/i[\tilde{C}_p(\tilde{L})]$, where i is the inclusion $i : \tilde{C}_p(\tilde{L}) \to \tilde{C}_p(\tilde{K})$. Loosely speaking, a relative chain from $\tilde{C}_p(\tilde{K}, \tilde{L})$ omits the portion of the chain that lies on $\partial\Omega$. Now $C_p(K)$ and $\tilde{C}_{n-p}(\tilde{K}, \tilde{L})$ have a one-to-one correspondence such that each p-cell in K has a unique dual $(n-p)$-cell counterpart on $\tilde{K} \setminus \tilde{L}$, and vice versa.[3] Thus $\tilde{C}_p(\tilde{L})$ is needed only when we impose boundary conditions for the Maxwell field G with $p = 2$. The unfixed part of G lies on the dual space of $\tilde{C}_2(\tilde{K}, \tilde{L})$, denoted $G' \in \tilde{C}^2(\tilde{K}, \tilde{L})$. Now we can connect $F \in C^2(K)$ and $G' \in \tilde{C}^2(\tilde{K}, \tilde{L})$ with the discrete Hodge operator (6.27).

Real arrays

With the above choices, both $C^2(K)$ and $\tilde{C}^2(\tilde{K}, \tilde{L})$ have κ_2 elements in their basis. The real array counterparts of $F \in C^2(K)$ and $G' \in \tilde{C}^2(\tilde{K}, \tilde{L})$, denoted $\boldsymbol{F}, \boldsymbol{G}' \in \mathbb{R}^{\kappa_2}$, satisfy

$$\boldsymbol{G}' = \mathbf{H}_M \boldsymbol{F}, \tag{6.32}$$

with $\mathbf{H}_M \in \mathbb{R}^{\kappa_2 \times \kappa_2}$ defined by (6.29). The only difference to the boundaryless case (6.30) is that $\boldsymbol{G}'$ now covers only G in the relative space $\tilde{C}^2(\tilde{K}, \tilde{L})$ but no part of G in $\tilde{C}^2(\tilde{L})$ (denoted $G_{\tilde{L}}$).

Let us denote by $\boldsymbol{G} \in \mathbb{R}^{\tilde{\kappa}_2}$ the real counterpart of the whole $G \in \tilde{C}^2(\tilde{K})$ so that $\boldsymbol{G}'$ contains those elements of $\boldsymbol{G}$ that are associated with a 2-cell in $\tilde{K} \setminus \tilde{L}$. In linear algebra, this can be written as follows: we have a projection matrix $\mathbf{P}_G \in \mathbb{R}^{\kappa_2 \times \tilde{\kappa}_2}$ so that $\boldsymbol{G}' = \mathbf{P}_G \boldsymbol{G}$, where $\mathbf{P}_G$ is an integer matrix with only zeros and ones for its elements. Similarly, if $\tilde{C}^2(\tilde{L})$ has $\tilde{\kappa}_2(\tilde{L})$ elements

[3] Notice that in a boundaryless situation there was directly a one-to-one correspondence of $C_p(K)$ and $\tilde{C}_{n-p}(\tilde{K})$. In contrast, in the boundary case, we resort to relative chain spaces modulo a boundary on the dual mesh side to establish a one-to-one connection.

in its basis, $\boldsymbol{G}_{\tilde{L}} \in \mathbb{R}^{\tilde{\kappa}_2(\tilde{L})}$ represents the $G_{\tilde{L}} \in \tilde{C}^2(\tilde{L})$, and we have a matrix $\mathbf{Q}_G \in \mathbb{R}^{\tilde{\kappa}_2(\tilde{L})\times\tilde{\kappa}_2}$ so that $\boldsymbol{G}_{\tilde{L}} = \mathbf{Q}_G\boldsymbol{G}$. We can easily see that

$$\boldsymbol{G} = \mathbf{P}_G^T\boldsymbol{G}' + \mathbf{Q}_G^T\boldsymbol{G}_{\tilde{L}}. \tag{6.33}$$

In the end, $\boldsymbol{G}_{\tilde{L}}$ is described by the boundary condition of G and $\boldsymbol{G}'$ remains to be solved.

In a bounded dual mesh, we have dual p-cells on the boundary that are not dual to any primal $(n-p)$-cell. In this case, we thus have $\tilde{\mathbf{d}}_p \neq \mathbf{d}_{n-p-1}^T$, in contrast to the unbounded case. The $\tilde{\mathbf{d}}_p$ has extra columns compared with $\mathbf{d}_{n-p-1}^T$ that are associated with the above extra dual cells on the boundary. Especially, $\mathbf{d}_1^T \neq \tilde{\mathbf{d}}_2$ in four-dimensional spacetime. However, if we want to operate $\boldsymbol{G}'$ with $\tilde{\mathbf{d}}_2$, we use the form $\tilde{\mathbf{d}}_2\mathbf{P}_G^T\boldsymbol{G}'$. This new operator $\tilde{\mathbf{d}}_2\mathbf{P}_G^T$ contains only those columns of $\tilde{\mathbf{d}}_2$ that are associated with inner dual 2-cells—the columns identified with $\tilde{C}^2(\tilde{L})$ are dropped. Thus $\tilde{\mathbf{d}}_2\mathbf{P}_G^T = \mathbf{d}_1^T$.

As in the above case, also the primal cells on $\partial\Omega$ form a cellular complex L and a chain complex $C_\times(L)$. The boundary chain space $C_2(L)$ is used to fix the boundary condition for the Faraday field F.

Boundary conditions

After the above definitions, imposing boundary conditions for the finite electromagnetic wave problem seems to be an easy task. We must fix the trace of the field on proper boundaries and take the latter into account in the computational solution. Additionally, we must take into account that the boundary conditions of F also fix some parts of G that are connected to the trace of F via the constitutive relations.

We fix the value of the finite Faraday field $F \in C^2(K)$ on a boundary section $\Gamma_F \subset \partial\Omega$. The primal cells on Γ_F form the cellular mesh complex $L_F \subset K$ and the finite chain spaces $C_p(L_F)$ for $p = 0, \ldots, n-1$. The boundary condition fixes the value of $t_{\Gamma_F}F \in C^2(L_F)$. By constitutive relations, if we know the value of F in a p-cell $\sigma \in L_F$, we know the value of G (approximately) in the dual $(n-p)$-cell of σ, denoted $\tilde{\sigma}$. Therefore, we must exclude such a part from G that is associated with those dual cells. We form a subcomplex of $\tilde{K}$ for the free parts of G by removing all the cells that touch L_F, i.e., cells that have a vertex on Γ_F, and obtain a cellular mesh complex $\tilde{K}_F$ and a chain complex $\tilde{C}_\times(\tilde{K}_F)$.

The boundary condition for $G \in \tilde{C}^2(\tilde{K})$ is established similarly to the above. The condition applies to a part of the boundary $\Gamma_G \subset \partial\Omega$, and the dual cells lying on Γ_G form a cellular mesh complex $\tilde{L}_G \subset \tilde{K}$ and finite chain spaces $\tilde{C}_p(\tilde{L}_G)$ for $p = 0, \ldots, n-1$. The boundary condition fixes

the value of $t_{\Gamma_G} G \in \tilde{C}^2(\tilde{L}_G)$. However, this does not fix any part of F, because there are no primal cells dual to the cells of $\tilde{L}_G$, as seen in the previous section. This situation, where the boundary conditions of F and G are assigned differently, resembles somewhat the finite element method, where some (primal) boundary conditions are imposed on a strong form and others (dual) on a weak form.

Similarly to the non-finite case in equation (5.42), we get the following short exact sequence for the finite relative cochain spaces:

$$0 \to C^p(K, L_F) \xrightarrow{k} C^p(K) \xrightarrow{t_{L_F}} C^p(L_F) \to 0. \tag{6.34}$$

With the boundary condition of F, we fix the value of $t_{L_F} F \in C^p(L_F)$. Sequence (6.34) ensures that if $F_{L_F} \in C^p(L_F)$ and $F' \in C^p(K, L_F)$ are known, we also know the whole $F \in C^p(K)$, on condition that $t_{L_F} F = F_{L_F}$ and $kF' = F$. Thus after the boundary condition has been assigned, a field in $C^p(K, L_F)$ remains to be solved. A similar sequence and result could be written also for $G \in \tilde{C}^2(\tilde{K})$.

For the Faraday field $F \in C^2(K)$, we know by the boundary condition the value of $t_{\Gamma_F} F \in C^2(L_F)$, and $F' \in C^2(K, L_F)$ remains a free variable (see figure 6.3, lower left). Second, for the Maxwell field $G \in \tilde{C}^2(\tilde{K})$, we know the $\tilde{C}^2(\tilde{L}_G)$-part by the boundary condition of G and the part of G on $\tilde{K} \setminus \tilde{K}_F$ by the boundary condition of F. Thus after boundary conditions have been imposed, $G^\dagger \in \tilde{C}^2(\tilde{K}_F, \tilde{L}_G \cap \tilde{K}_F)$ remains a free variable.[4] At lower right in figure 6.3, $\tilde{K}_F \setminus \tilde{L}_G$ contains besides the dashed parts (denoted later by $\tilde{K}_F \setminus \tilde{L}'_G$) also the final boundary subcomplex $\tilde{L}_{\text{final}}$, i.e., the subcomplex of the dual cells on Γ_{final}. As we can see, there is no primal counterpart to $\tilde{L}_{\text{final}}$; thus no one-to-one connection can exist between $C^2(K, L_F)$ and $\tilde{C}^2(\tilde{K}_F, \tilde{L}_G \cap \tilde{K}_F)$. To work this out, we should remove G on $\tilde{L}_{\text{final}}$ from the free variable $G^\dagger$. Fortunately, in the next section we realize that we do not need to know G on $\tilde{L}_{\text{final}}$, if we are only to solve F or A. Thus we define $\tilde{L}'_G = (\tilde{L}_G \cup \tilde{L}_{\text{final}}) \cap \tilde{K}_F$ so that $F' \in C^2(K, L_F)$ and $G'' \in \tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$ have a one-to-one connection by the constitutive law (6.27) (see figure 6.3, lower right). Additionally, since the boundary operator generalizes naturally to the corresponding relative chain spaces, we get also the exterior derivative and its real array counterpart for the relative cochain spaces.

If we use the potential $A \in C^1(K)$ instead of the Faraday field F, we must impose a boundary condition on A. If we know $t_{\Gamma_F} F \in C^2(L_F)$ by boundary

[4] We must use the intersection of $\tilde{L}_G$ with $\tilde{K}_F$, because $\tilde{L}_G$ is not a subset of $\tilde{K}_F$. $G^\dagger$ is truly a free part, since, in the next section, we will see how the values of G in $\Gamma_F \setminus \Gamma_G$ are dispensable. Moreover, we will see how the (initial) boundary values of G on $\Gamma_F \cap \Gamma_G$ are included in the solving process, even if the cells on $\Gamma_F \cap \Gamma_G$ do not belong to $\tilde{K}_F$.

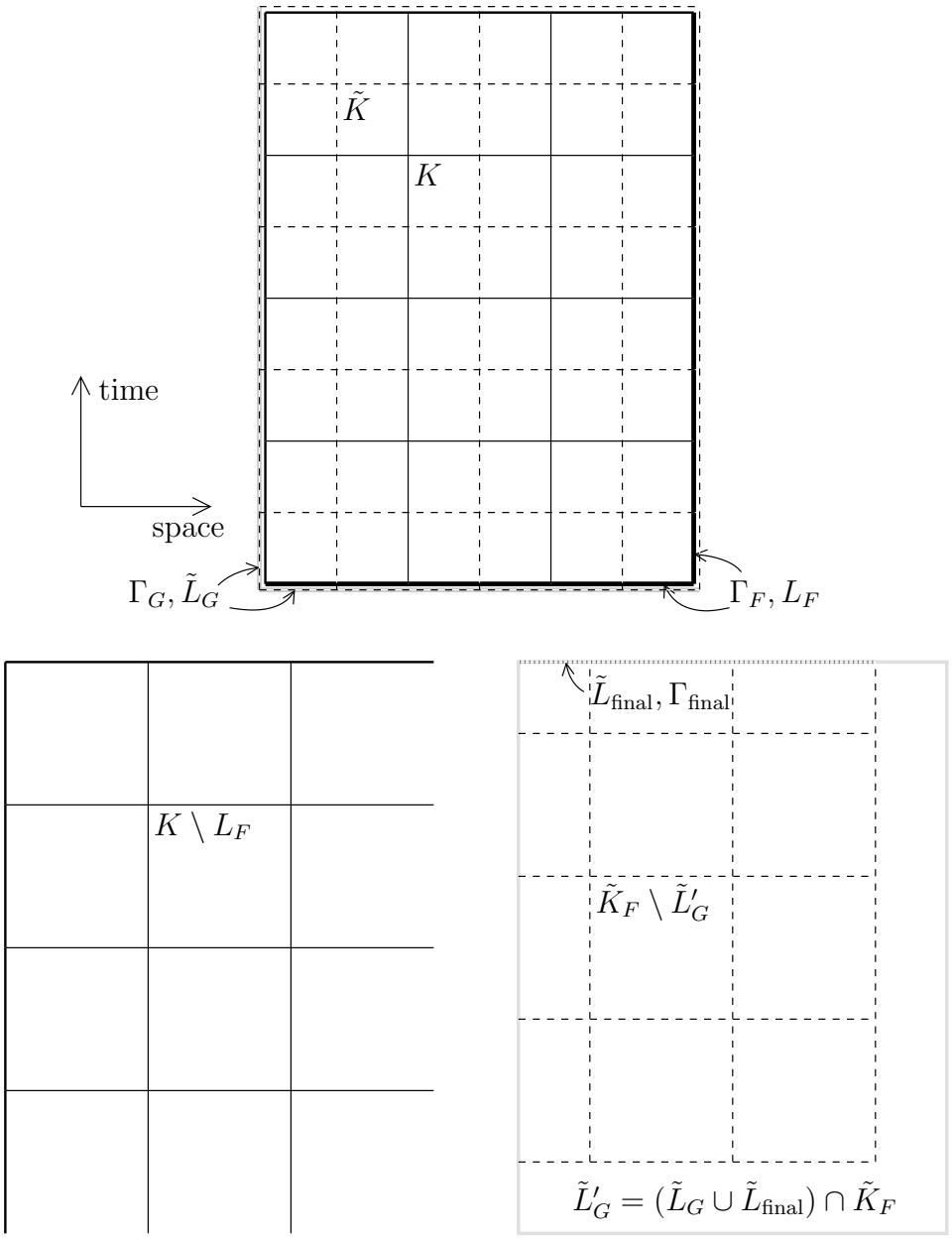


Figure 6.3: A two-dimensional example of the subcomplexes K, $\tilde{K}$, L_F, $\tilde{L}_G$, $\tilde{K}_F$, and $\tilde{L}_{\text{final}}$ and the sets $K \setminus L_F$ and $\tilde{K}_F \setminus \tilde{L}'_G$. Above: on the initial boundary both F and G are fixed; the left side holds a G-boundary condition and the right side an F-condition. Bottom left: after the F-condition has been imposed, only F in $K \setminus L_F$ is left as a free variable. Bottom right: the F-condition also fixes G on the dual 2-cells of the 2-cells of L_F; thus only G on $\tilde{K}_F$ is left as a free variable. Additionally, the G-condition fixes more variables, and the G-solution need not be known on Γ_{final}. Thus finally only G on $\tilde{K}_F \setminus \tilde{L}'_G$ is left as a free variable.

conditions, this fixes also $t_{\Gamma_F} A \in C^1(L_F)$ by using (6.9) in L_F. Thus only $A' \in C^1(K, L_F)$ is left as free variable, as in figure 6.4, right.

We must connect these new boundary parts Γ_F and Γ_G to the old ones Γ_{boundary}, Γ_{initial}, and Γ_{final}; i.e., we must see how Γ_F and Γ_G are composed from the old ones. Especially we must have $\Gamma_F \cap \Gamma_G = \Gamma_{\text{initial}}$ and $\Gamma_F \cup \Gamma_G = \Gamma_{\text{initial}} \cup \Gamma_{\text{boundary}}$, because the initial boundary has two conditions but the spatial boundary only one. In addition, the final boundary has no condition. To generate a formulation so that we gain an equal number of unknowns and linearly independent equations, we must require that the initial and final boundaries Γ_{initial} and Γ_{final} have the same number of 2-cells on them. This is not a restrictive additional requirement for the mesh, because our hypercuboidal (or even the more general hyper-octahedron) mesh naturally fulfils this conditions, assuming no element is degenerated.

6.1.4 Restricting the finite Maxwell laws

By boundary conditions, we know the fields F (or A) and G in some parts of a domain. As we saw above, the relative spaces of free variables (i.e., $C^2(K, L_F)$, $\tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$, and $C^1(K, L_F)$) are smaller than the entire finite chain spaces $C^2(K)$, $\tilde{C}^2(\tilde{K})$, and $C^1(K)$. Thus since the number of free variables is reduced in the finite Maxwell equations (6.7), (6.8), and (6.9), some parts of the equations become redundant. To be precise, the Maxwell-Ampère law (6.8) does not need to be applied on every boundary of dual 3-cells in $\tilde{C}_3(\tilde{K})$, because G is known in $\tilde{K} \setminus \tilde{K}_F$ and $\tilde{L}_G$ in advance by the boundary conditions of F and G.

Then which equations should be removed to make the rest of them linearly independent? Additionally, we want no relevant condition to be unenforced. The reasoning behind these choices would make a fascinating study, but, unfortunately, we must skip any further rumination on the matter. It would involve a series of different decompositions of relative cochain and chain spaces, which would then tell us which parts of the finite Maxwell laws could be omitted. Here we deal only with the results.

In two-dimensional spacetime, no part of the finite Maxwell-Faraday law should be omitted. In contrast, three- and four-dimensional spacetimes hold equations that deal only with $F \in C^2(L_F)$ (or $F \in C^1(L_F)$ in the three-dimensional case) known in advance by boundary conditions. Thus the equations dealing only with known variables are not needed. However, boundary conditions should be placed so that the Maxwell-Faraday law is satisfied also there. This imposes a constraint on the initial and spatial boundary conditions, as mentioned in section 5.2.2 for the initial case. Then (6.7) is reduced

to

$$(F|\partial\Sigma) = 0 \quad \forall\Sigma \in C_3(K) \setminus C_3(L_F). \tag{6.35}$$

In two- and three-dimensional cases, the space $C_3(K) \setminus C_3(L_F)$ is replaced by $C_2(K) \setminus C_2(L_F)$; in the two-dimensional case, the boundary is one-dimensional, which causes $C_2(L_F) = \emptyset$. However, the alternative approach with the potential equation (6.9) stay the same, because it is valid regardless of whether F is a free variable or fixed by a boundary condition.

However, the Maxwell-Ampère law (6.8) is affected by boundary conditions in all dimensions. In classical two- and three-dimensional formulations, we remove the equations connected to the duals of the boundary cells on which the variables of the primal mesh are imposed. However, in our spacetime case the situation is not so straightforward.

On spatial boundaries where F is fixed, i.e., at $\Gamma_F \cap \Gamma_{\text{boundary}}$, the situation resembles the classical case. All the equations of the dual 3-cells that touch $\Gamma_F \cap \Gamma_{\text{boundary}}$ must be removed, because they would solve only the value of G on those boundaries. These values have no effect on fields elsewhere, because those boundary dual 3-cells do not have any primal cells dual to them. However, we cannot remove the equations associated with the dual 3-cells that touch the initial boundary Γ_{initial}, because, in addition to the F-condition, there is also a G-boundary condition on the initial boundary. If the equations were removed, the G-initial condition would be forgotten, and we should not allow that. Moreover, G is only an auxiliary quantity to solve F, and we are interested only in solving F in every 2-cell in $K \setminus L_F$, and the knowledge of G is not necessary everywhere. Thus we can remove the equations associated with the dual 3-cells that touch the final boundary Γ_{final}, because they are only needed to solve G on Γ_{final} after F is already known everywhere (see figure 6.4, right).

The subcomplex where all the dual cells that touch $\Gamma_{\text{boundary}} \cap \Gamma_F$ or Γ_{final} are removed from $\tilde{K}$ is denoted by $\tilde{K}_{\text{eq}}$. Additionally, analogous to the Maxwell-Faraday law, a part of the Maxwell-Ampère law deals only with the values of G that are known in advance by boundary conditions. Because those equations are associated with 3-cells in $\tilde{L}_G$, they are associated with the parts of G on $\tilde{C}^2(\tilde{L}_G)$. Consequently, the Maxwell-Ampère law (6.8) is reduced to

$$(G|\partial\tilde{\Sigma}) = (J|\tilde{\Sigma}), \quad \forall\tilde{\Sigma} \in \tilde{C}_3(\tilde{K}_{\text{eq}}) \setminus \tilde{C}_3(\tilde{L}_G). \tag{6.36}$$

In two- and three-dimensional spacetime cases, the space $\tilde{C}_3(\tilde{K}_{\text{eq}}) \setminus \tilde{C}_3(\tilde{L}_G)$ is replaced by $\tilde{C}_2(\tilde{K}_{\text{eq}})$ and $\tilde{C}_3(\tilde{K}_{\text{eq}})$, respectively, because in lower-dimensional cases the $\tilde{L}_G$-related part vanishes. In the real array form (6.36) becomes

$$\mathbf{P}_{\text{eq}}\tilde{\mathbf{d}}_2\boldsymbol{G} = \mathbf{P}_{\text{eq}}\boldsymbol{J}, \tag{6.37}$$

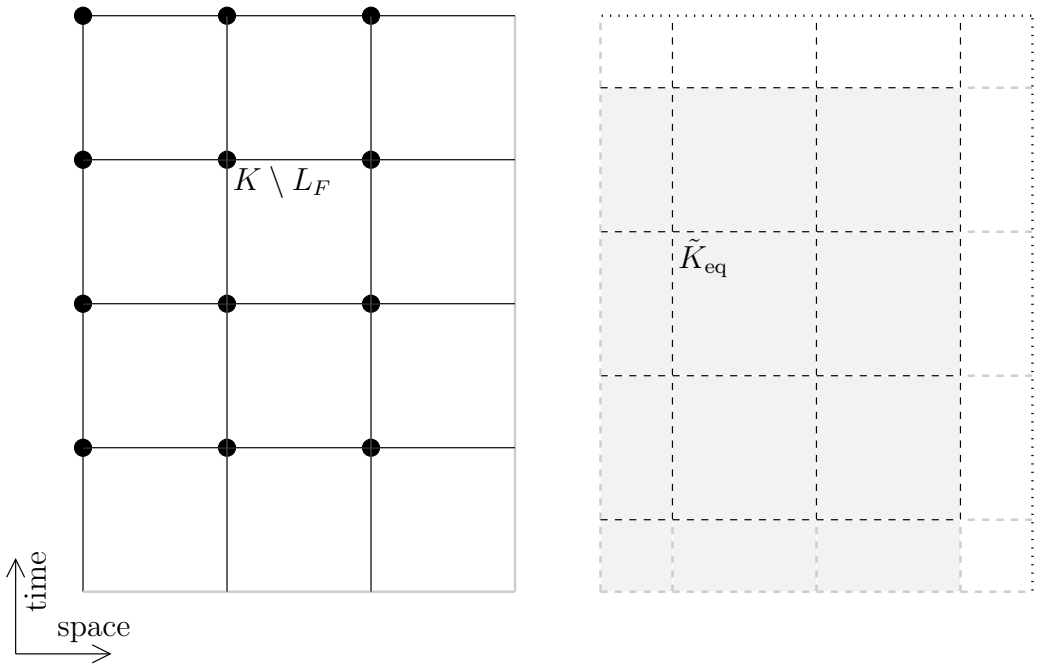


Figure 6.4: A two-dimensional example of subcomplexes for the potential and the Maxwell-Ampère law. Left: in the set of cells $K \setminus L_F$, the potential A is a free variable. The degrees of freedom of the free part A' lies on the nodes plotted with dots. On the boundary plotted in gray, the potential A (and field F) is known by boundary conditions. Right: the subcomplex $\tilde{K}_{\mathrm{eq}}$, where the Maxwell-Ampère is used in computation, is denoted by a light gray background. In the gray dual cells, G is known by boundary conditions, and on the dotted boundaries, it is not necessary to know the solution of G.

where $\mathbf{P}_{\mathrm{eq}} \in \mathbb{R}^{\kappa_1^{\mathrm{eq}} \times \tilde{\kappa}_3}$ is a projection matrix (with only zeros and ones as its elements), which chooses those rows of $\tilde{\mathbf{d}}_2$ and $\boldsymbol{J} \in \mathbb{R}^{\tilde{\kappa}_3}$ that are associated with dual 3-cells in $\tilde{C}_3(\tilde{K}_{\mathrm{eq}}) \setminus \tilde{C}_3(\tilde{L}_G)$. Here κ_1^{eq} denotes the number of elements in $\mathbf{P}_{\mathrm{eq}}\boldsymbol{J}$. Remember that $\mathbf{d}_1^T \neq \tilde{\mathbf{d}}_2$ because of the extra dual cells on the boundary. With these choices, we have now as many independent equations as we have unknowns. In the next section, the potential formulation shows this more clearly.

6.1.5 Formulation with the potential

If we want a formulation that could be solved directly with standard linear algebra codes, we must combine all the finite equations into one matrix equa-

tion. As we saw above, we can do this with the electromagnetic potential, as in equation (6.31) and by now taking into account the influences of boundary conditions. Therefore, we must write boundary conditions in real array form.

Boundary conditions in real array form

In section 6.1.3, we divided the cochain spaces of F, G and A according to boundary conditions. In each case, one part was determined by boundary conditions and the rest remained as a free variable. We must now divide the real array counterparts of G and A according to the above cochain spaces. As before, the real array counterparts of the undivided spaces $F \in C^2(K)$ and $G \in \tilde{C}^2(\tilde{K})$ are $\boldsymbol{F} \in \mathbb{R}^{\kappa_2}$ and $\boldsymbol{G} \in \mathbb{R}^{\tilde{\kappa}_2}$.

In the previous sections, we concluded that $G \in \tilde{C}^2(\tilde{K})$ is known by boundary conditions in $\tilde{K} \setminus \tilde{K}_F$ and $\tilde{L}_G$, and that the solution does not need to be known in Γ_{final}. The set of cells where G is known is denoted $\tilde{K}_G = \tilde{L}'_G \cup (\tilde{K} \setminus \tilde{K}_F)$ (where $\tilde{L}'_G = (\tilde{L}_G \cup \tilde{L}_{\text{final}}) \cap \tilde{K}_F$, as in section 6.1.3). The part of G we want to solve lies in the relative cochain space $\tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$ (see figure 6.3, bottom right). With the above choices, both $C^2(K, L_F)$—i.e., the space of the free part of F—and $\tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$ have the same number of elements in their bases, and there is a one-to-one duality connection between the cells in $K \setminus L_F$ and $\tilde{K}_F \setminus \tilde{L}'_G$. Consequently, the elements of $C^2(K, L_F)$ and $\tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$ are one-to-one connected by the discrete Hodge operator (6.27).

We have already defined a division for the real array $\boldsymbol{G}$ in (6.33), which can be used to impose the boundary condition of G. $\boldsymbol{G}_{\tilde{L}}$ contains such a part of $\boldsymbol{G}$ that is known by the boundary condition of G (on $\tilde{L}_G$), but it also contains other parts of $\boldsymbol{G}$. Nonetheless, we can assume that $\boldsymbol{G}_{\tilde{L}}$ is zero outside of $\tilde{L}_G$. This assumption forces zero values for G on Γ_{final} and $\Gamma_F \cap \Gamma_{\text{boundary}}$, which is yet acceptable, because we are not interested in these parts of G, and because they have no effect on the solution elsewhere. Note that $\boldsymbol{G}'$ is not truly a free part of G, because it contains quantities that are below fixed by the boundary condition of A.

The real array counterpart of the (undivided) potential $A \in C^1(K)$ is $\boldsymbol{A} \in \mathbb{R}^{\kappa_1}$. The free part of A, i.e., the relative cochain $A' \in C^1(K, L_F)$, has κ_1^A elements in its basis and thus corresponds to the real array $\boldsymbol{A'} \in \mathbb{R}^{\kappa_1^A}$. Similarly, the boundary condition part $A_{L_F} \in C^1(L_F)$ corresponds to $\boldsymbol{A}_{L_F} \in \mathbb{R}^{\kappa_1(L_F)}$. We have projection matrices $\mathbf{P}_A \in \mathbb{R}^{\kappa_1^A \times \kappa_1}$ and $\mathbf{Q}_A \in \mathbb{R}^{\kappa_1(L_F) \times \kappa_1}$ such that $\boldsymbol{A}' = \mathbf{P}_A \boldsymbol{A}$ and $\boldsymbol{A}_{L_F} = \mathbf{Q}_A \boldsymbol{A}$ and

$$\boldsymbol{A} = \mathbf{P}_A^T \boldsymbol{A}' + \mathbf{Q}_A^T \boldsymbol{A}_{L_F}. \tag{6.38}$$

By the exterior derivative, an arbitrary $A' \in C^1(K, L_F)$ determines a unique

$F' = \mathrm{d}A' \in C^2(K, L_F)$ and, by the discrete Hodge operator (6.27), a unique $G'' \in \tilde{C}^2(\tilde{K}_F, \tilde{L}'_G)$. Similarly, $A_{L_F} \in C^1(L_F)$ determines uniquely an $F_{L_F} = \mathrm{d}A_{L_F} \in C^2(L_F)$ and fixes a part of G in $\tilde{K}_G$. The rest of G in $\tilde{K}_G$ is fixed by the boundary condition of G—or left open in Γ_{final} and $\Gamma_F \cap \Gamma_{\text{boundary}}$, as explained before.

Formulation

We are now ready to write a numerical formulation with boundary conditions taken care of. By combining (6.12), (6.32), (6.33), (6.37), and (6.38), we get

$$\mathbf{C}'^T \mathbf{H}_M \mathbf{C} \boldsymbol{A}' = \boldsymbol{J}', \tag{6.39}$$

where

$$\mathbf{C}'^T = \mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_2 \mathbf{P}_G^T \tag{6.40}$$

from equation (6.37),

$$\mathbf{C} = \mathbf{d}_1 \mathbf{P}_A^T, \tag{6.41}$$

and

$$\boldsymbol{J}' = \mathbf{P}_{\text{eq}} \boldsymbol{J} - \mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_2 \mathbf{Q}_G^T \boldsymbol{G}_{\tilde{L}} - \mathbf{C}'^T \mathbf{H}_M \mathbf{d}_1 \mathbf{Q}_A^T \boldsymbol{A}_{L_F}. \tag{6.42}$$

Note that the incidence matrices $\mathbf{C}'$ and $\mathbf{C}$ are truly different: $\mathbf{C}'$ in (6.40) contains the columns of $\mathbf{d}_1 = \mathbf{P}_G \tilde{\mathbf{d}}_2^T$ that are associated with dual 3-cells in $\tilde{C}_3(\tilde{K}_{\text{eq}}) \setminus \tilde{C}_3(\tilde{L}_G)$, whereas $\mathbf{C}$ in (6.41) contains columns related to primal 0-cells in $K \setminus L_F$—and these sets are not elementwise duals with each other. Thus the system matrix $\mathbf{C}'^T \mathbf{H}_M \mathbf{C}$ is not symmetric like its analogues in most lower-dimensional non-spacetime formulations.

The modified source current term $\boldsymbol{J}'$ in (6.42) contains, in addition to the source currents in $\tilde{C}_3(\tilde{K}_{\text{eq}}) \setminus \tilde{C}_3(\tilde{L}_G)$, also nonhomogenous parts of boundary conditions. The wave problem (6.39) is identical with (6.31) when boundary conditions are taken into account. The part of $\boldsymbol{A}$ that is known by the boundary condition of A is moved to the left hand side and constitutes the $-\mathbf{C}'^T \mathbf{H}_M \mathbf{d}_1 \mathbf{Q}_A^T \boldsymbol{A}_{L_F}$-part into $\boldsymbol{J}'$. The $-\mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_2 \mathbf{Q}_G^T \boldsymbol{G}_{\tilde{L}}$-part is generated by the known part of G and is likewise brought to the left hand side.

With our choices of primal and dual meshes, boundary conditions, and restriction of Maxwell equations, we have the same number of equations and unknowns in equation (6.39), i.e., $\kappa_1^{\text{eq}} = \kappa_1^A$ (cf. figure 6.4). The numerical electromagnetic wave problem with the potential is now reduced to form: for a known $\boldsymbol{J}' \in \mathbb{R}^{\kappa_1^{\text{eq}}}$, find a $\boldsymbol{A}' \in \mathbb{R}^{\kappa_1^A}$ so that (6.39) is satisfied. After that, the full $\boldsymbol{A}$ is computed from (6.38), because both $\boldsymbol{A}'$ and $\boldsymbol{A}_{L_F}$ are now known. Because in lower-dimensional cases this formulation requires no gauging methods, we deal with gauging in the context of computation in four-dimensional spacetime.

6.2 Spacetime algorithms in two- and three-dimensional spacetimes

First, we demonstrate the algorithm with two- and three-dimensional spacetime electromagnetic wave problems. In both cases, the potential is a 0-cochain, i.e., a scalar potential. In fact, it was a modeling decision to choose it so. Consequently, since we need no gauging in these cases, the algorithm becomes somewhat more simple. For a reason, we postpone introducing gauging techniques to the next section, where we first truly need them.

6.2.1 Spacetime of one spatial dimension

First, we formulate a two-dimensional wave problem and then work out to find an efficient numerical solution to the resulting linear system of equations. In the end, we give a few numerical examples of the uses of this formulation.

Formulation

In two-dimensional spacetime, we obtain the finite equivalents of electromagnetic laws in the same manner as above in the four-dimensional case. For $F \in C^1(K)$, $G \in \tilde{C}^1(\tilde{K})$ and the known $J \in \tilde{C}^2(\tilde{K})$, we obtain the finite Maxwell laws

$$(F|\partial\Sigma) = 0, \qquad \forall\Sigma \in C_2(K), \tag{6.43}$$

$$(G|\partial\tilde{\Sigma}) = (J|\tilde{\Sigma}), \quad \forall\tilde{\Sigma} \in \tilde{C}_2(\tilde{K}). \tag{6.44}$$

For the potential equation (5.50), the finite version is

$$(A\,|\partial\Sigma) = (F\,|\,\Sigma), \quad \forall\Sigma \in C_1(K), \tag{6.45}$$

for $A \in C^0(K)$. Thus we need a pair of two-dimensional meshes, i.e., two mesh complexes K and $\tilde{K}$. The 1-cells (or edges) of the two meshes must be Lorentzian orthogonal to each other, as in figure 6.1; then we can use the finite Hodge operator (6.27) (or especially (6.28)) between F and G, where now cells σ and $\tilde{\sigma}$ are one-dimensional. We have already shown in figures 6.3 and 6.4 how boundary conditions restrict chain and cochain spaces. Now, the Maxwell-Ampère 1-cochain G has a real counterpart $\boldsymbol{G} \in \mathbb{R}^{\tilde{\kappa}_1}$ divided into two parts with equation (6.33), where, in this case, $\boldsymbol{G}' \in \mathbb{R}^{\kappa_1}$, $\boldsymbol{G}_{\tilde{L}} \in \mathbb{R}^{\tilde{\kappa}_1(\tilde{L})}$, $\mathbf{P}_G \in \mathbb{R}^{\kappa_1 \times \tilde{\kappa}_1}$, and $\mathbf{Q}_G \in \mathbb{R}^{\tilde{\kappa}_1(\tilde{L}) \times \tilde{\kappa}_1}$. The real array equivalent for the potential 0-cochain A is $\boldsymbol{A} \in \mathbb{R}^{\kappa_0}$, and it is partitioned identically to (6.38), where

$\boldsymbol{A}' \in \mathbb{R}^{\kappa_0^A}$, $\boldsymbol{A}_{L_F} \in \mathbb{R}^{\kappa_0(L_F)}$, $\mathbf{P}_A \in \mathbb{R}^{\kappa_0^A \times \kappa_0}$, and $\mathbf{Q}_A \in \mathbb{R}^{\kappa_0(L_F)\times\kappa_0}$. In real array form, equation (6.45) and the restricted version of (6.44) become

$$\mathbf{d}_1 \boldsymbol{A} = \boldsymbol{F}, \tag{6.46}$$

$$\mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_1 \boldsymbol{G} = \mathbf{P}_{\text{eq}} \boldsymbol{J}, \tag{6.47}$$

where $\boldsymbol{F} \in \mathbb{R}^{\kappa_1}$, $\boldsymbol{J} \in \mathbb{R}^{\tilde{\kappa}_2}$, and $\mathbf{P}_{\text{eq}} \in \mathbb{R}^{\kappa_0^{\text{eq}} \times \tilde{\kappa}_2}$.

Analogously to (6.39), we get an algorithm for the two-dimensional spacetime problem: for known $\boldsymbol{J}$, $\boldsymbol{A}_{L_F}$ and $\boldsymbol{G}_{\tilde{L}}$, find $\boldsymbol{A}' \in \mathbb{R}^{\kappa_0^A}$ so that

$$\mathbf{C}'^T \mathbf{H}_M \mathbf{C} \boldsymbol{A}' = \boldsymbol{J}', \tag{6.48}$$

where

$$\mathbf{C}'^T = \mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_1 \mathbf{P}_G^T, \tag{6.49}$$

$$\mathbf{C} = \mathbf{d}_0 \mathbf{P}_A^T, \tag{6.50}$$

$$\boldsymbol{J}' = \mathbf{P}_{\text{eq}} \boldsymbol{J} - \mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_1 \mathbf{Q}_G^T \boldsymbol{G}_{\tilde{L}} - \mathbf{C}'^T \mathbf{H}_M \mathbf{d}_0 \mathbf{Q}_A^T \boldsymbol{A}_{L_F}. \tag{6.51}$$

Numerical linear algebra

We noticed earlier that, in contrast to its counterparts in typical non-spacetime problems, the system matrix $\mathbf{C}'^T \mathbf{H}_M \mathbf{C}$ is not symmetric. This asymmetry forbids us from using many of the standard algorithms to solve systems of linear equations, e.g., the Cholesky decomposition and the conjugate gradient method. Moreover, we naturally begin to wonder how solving this one big system of linear equations is connected to the time-stepping algorithms—especially to methods in which each time-step is explicit. Answers could be found by examining the sparsity pattern of the matrix $\mathbf{C}'^T \mathbf{H}_M \mathbf{C}$.

Figure 6.5 shows the sparsity pattern of the system matrix $\mathbf{C}'^T \mathbf{H}_M \mathbf{C}$ in our prototype geometry of figure 5.7 with a mesh similar to the upper part of figure 6.3. In the first examples below, the problem is defined more accurately. The mesh is made from a node grid of 9×14 nodes in spatial and temporal directions, respectively. The grid density is equal in spatial and temporal directions (with $c = 1$); i.e., the mesh facets are squares. As we can see, the matrix is lower triangular. Thus the system of equations (6.48) can be solved with forward substitution without any additional approximations or iterations. The solution process closely resembles the leap-frog time-stepping scheme: the potential A at one event is computed from the values of A at previous time instants in the neighboring nodes. In this example, the resemblance results from the structure of our mesh: the nodes lie in layers of fixed time instants and are enumerated continuously layer by layer from the initial surface. In fact, in this geometry and mesh, solving (6.48) is

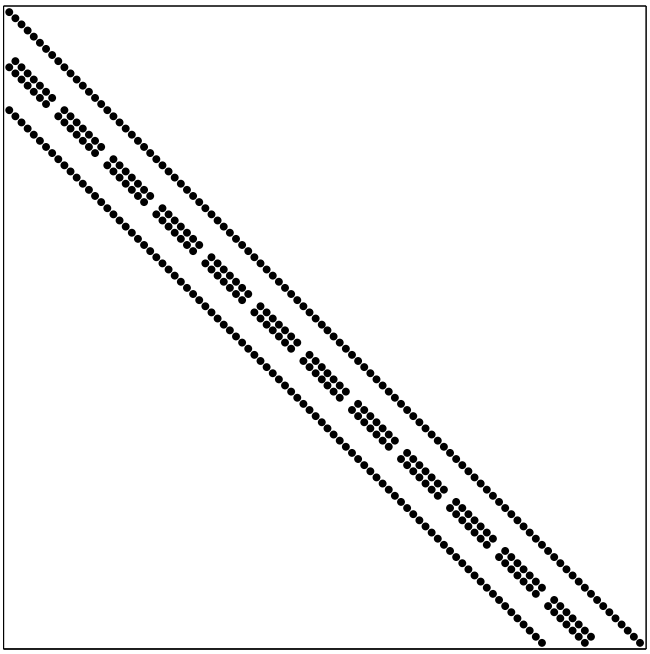


Figure 6.5: Sparsity pattern of the system matrix $\mathbf{C}'^T\mathbf{H}_M\mathbf{C}$. The dots show the places of the nonzero elements in the matrix. This system matrix is lower diagonal.

operation by operation identical with the leap-frog time-stepping method (see section 2.2.2). In general geometries and meshes, the situation becomes more complicated and time-stepping interpretation ambiguous. However, as long as the mesh is construed from layers of spacelike elements, this interpretation can be put forward. Thus the whole leap-frog time-stepping scheme is, in our case, realized by the standard forward substitution process of a single lower triangular matrix.

Examples

Now, we can find a solution to an electromagnetic wave problem in two-dimensional spacetime by solving equation (6.48) in a given geometry. First, we look for a solution in a rather simple geometry—in our prototype problem geometry of figure 5.7. We give zero boundary conditions similar to the upper part of figure 6.3; i.e., F vanishes on the edges of the right side, and G vanishes on the dual edges of the left side. Materials and the type and placing of source are left for choice. We have three basic types of sources: a source current inside the domain, a nonvanishing initial or boundary condition for F, and the

latter also for G. In the first three examples, we use the prototype geometry and the three possible source terms in turn. Additionally, we demonstrate the effect of a dielectric material inside the domain. The last two examples show a few possible atypical geometries for spacetime problems.

Example 6.1. Figure 6.6 shows an example where a nonvanishing initial condition is given to the potential A and a homogenous initial condition to G. The potential A on the initial surface is forced into the form of a bell curve—i.e., a Gaussian function—centered left of the middle of the spatial domain with a width of several mesh edges. The domain consists of a vacuum. The upper part of figure 6.6 shows the solved potential. The horizontal axis is the spatial dimension and the vertical the time direction. There are 101 nodes in the spatial direction and 131 in the temporal direction. The pointwise values of the potential are interpolated from the node values of A by a bilinear interpolator.

The initial distribution of A begins to travel in both spatial directions at the speed of light—here units are chosen so that $c = 1$. These wavefronts form a V-shaped pattern at the bottom of the spacetime field distribution. When the wavefront collides with the left side, it reflects back with the same sign as the original wavefront. In contrast, a reflection from the right side changes the sign. The difference is caused by the different boundary conditions: a homogenous G-condition of the left side and a homogenous F-condition on the right side. When the wavefronts meet again at a later time instant, the total potential cancels out.

The lower parts of figure 6.6 show both components of the Faraday field $F = \mathrm{d}A$. The left shows the temporal component, i.e., the electric field $\mathcal{E}$. Additionally, the right shows the spatial component, i.e., the magnetic flux $\mathcal{B}$. Their pointwise values are interpolated with the edge elements of van Welij [84]. The two initial wavefronts have opposite $\mathcal{B}$-values but identical $\mathcal{E}$-values. When the wavefront reflects from the perfect electric conductor on the right, the sign of $\mathcal{E}$ changes.

Example 6.2. Let us change the source term in the above example into a nonvanishing initial condition for G; i.e., G serves now as the source instead of A. Additionally, we change the distribution of the initial conditions into a period of the sin-function. The source is placed again left of the middle of the spatial domain and its width is 40 mesh edges. The upper part of figure 6.7 shows the potential. With this source, the two initial wavefronts have opposite signs. The lower part of figure 6.7 shows the electric field and magnetic flux.

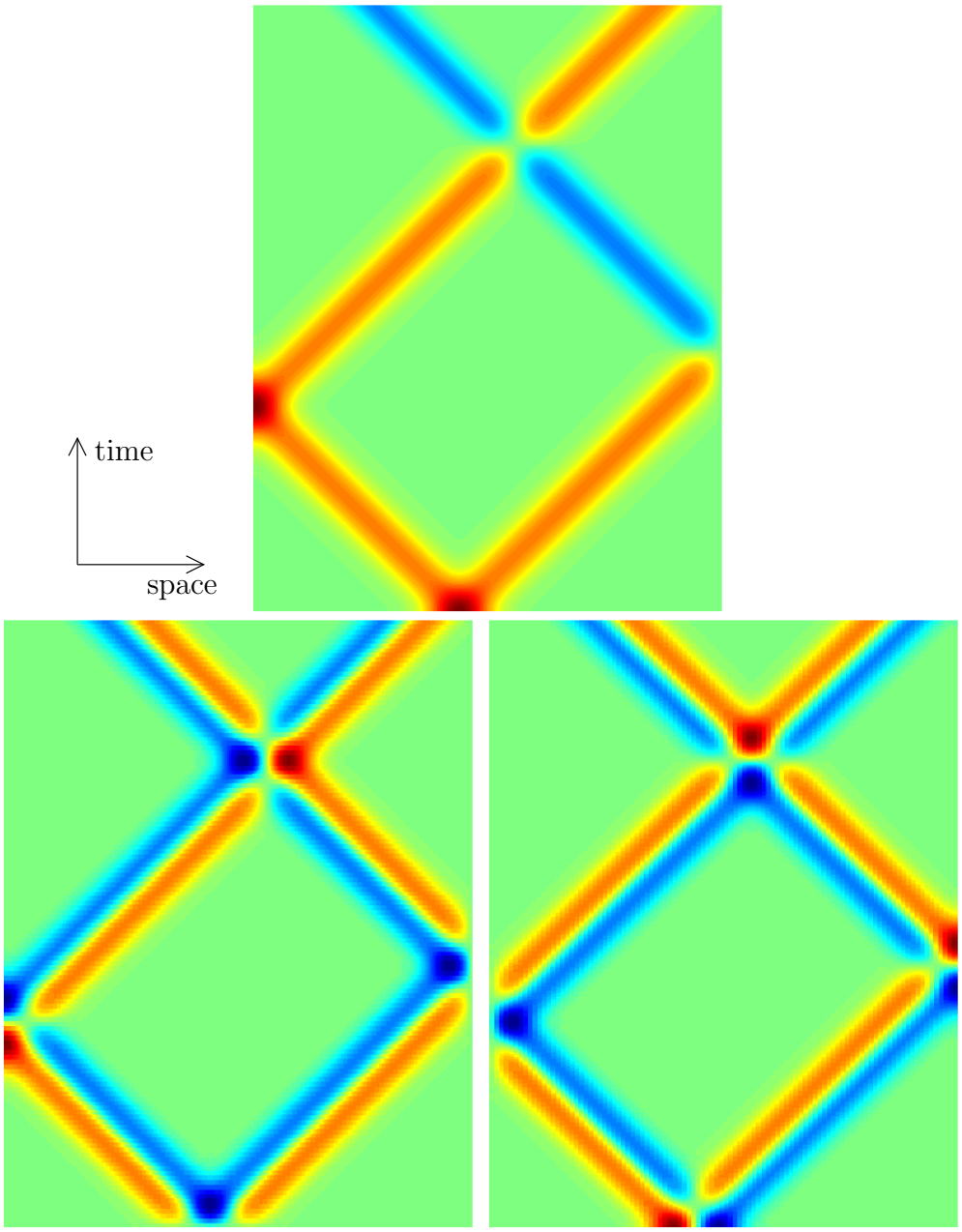


Figure 6.6: A two-dimensional example, where the horizontal axis is the spatial dimension and the vertical is the time direction. The domain consists of a vacuum without source charge-currents; a zero boundary condition for A is given on the right side and a zero G-condition in the initial and left boundaries. A nonvanishing initial A-condition is given as the source term: $\mathrm{t}_{\Gamma_{\mathrm{initial}}} A$ is a Gaussian function, i.e., a bell curve. Above: the distribution of the potential A. Bottom left: the electric field $\mathcal{E}$. Bottom right: the magnetic flux $\mathcal{B}$.

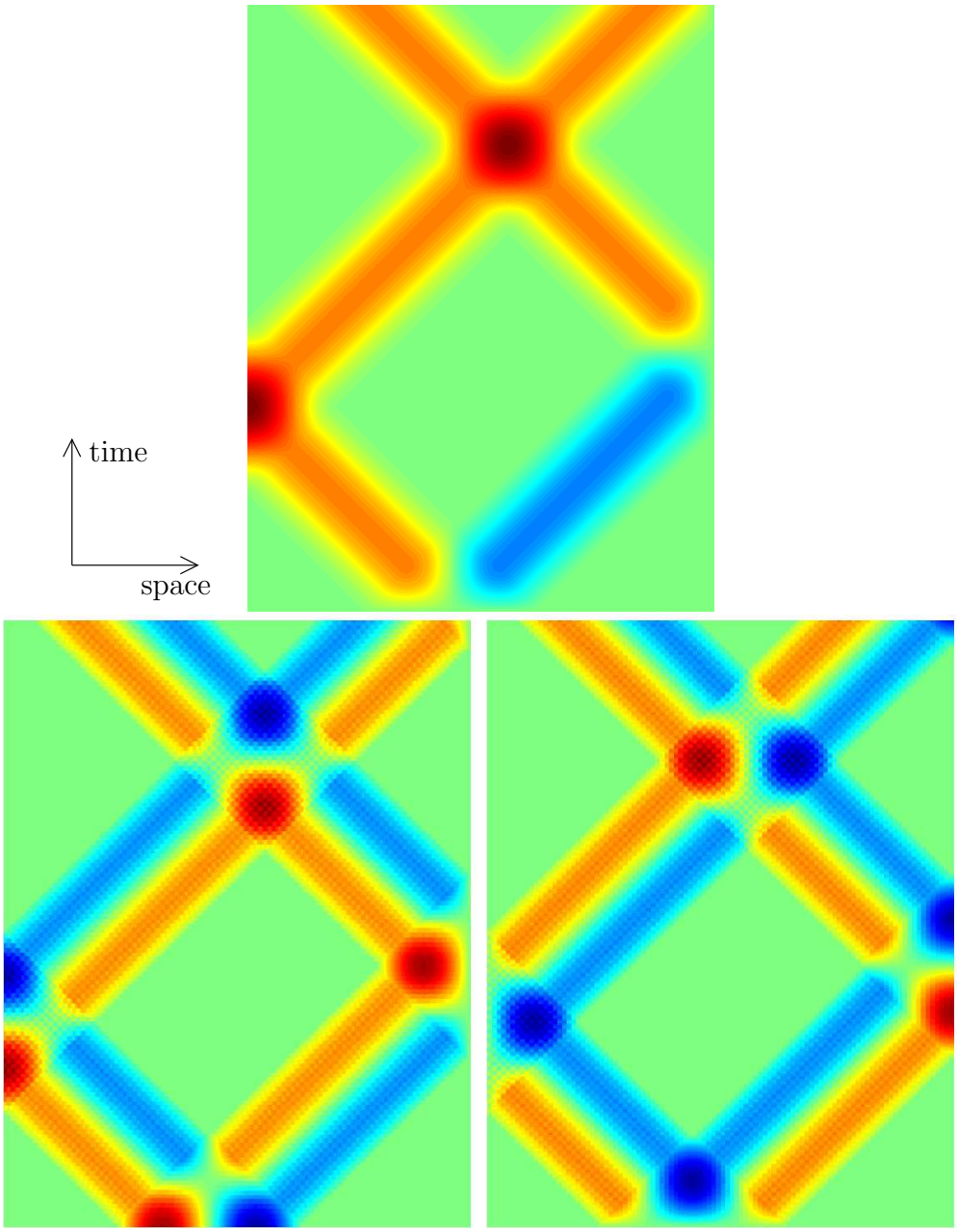


Figure 6.7: A two-dimensional example: a vacuum without source charge-currents and with a homogenous boundary condition for A on the initial and right boundaries and a homogenous G-condition on the left side. A nonvanishing initial G-condition is given as the source term: $\mathrm{t}_{\Gamma_{\mathrm{initial}}} G$ is a period of the sin-function. Above: the distribution of the potential A. Bottom left: the electric field $\mathcal{E}$. Bottom right: the magnetic flux $\mathcal{B}$.

Example 6.3. In the third example, the source is given by a charge-current distribution inside the domain. The charge-current is, in fact, a true current integrated over a time period, because the two-dimensional spacetime of our model cannot have any charge distributions. The source is placed yet again left of the middle of the spatial domain in a time-directional row of facets with a length of 40 facets. The source current is a period of the sin-function as a function of time. The upper part of figure 6.8 shows the potential, and the lower part the electric field and magnetic flux. Except for the source region, the wavefronts are similar to those in the previous example.

Example 6.4. In first three examples, the domain was a vacuum. In figure 6.9, we demonstrate the effects of different materials: the right part (80% of the width of the spatial domain) is a vacuum, and the left part (20%) is made from a dielectric material with the dielectric constant $\varepsilon_r = 2$. Otherwise, the situation is identical to example 6.1 and figure 6.6. That is, a Gaussian function is given as the initial condition for the potential A and a zero initial condition for G. The upper part of figure 6.9 gives the solved potential. A part of the right wavefront is refracted, and a second part is reflected from the interface between the two materials. In the dielectric material, the speed of light is $c/\sqrt{2}$; thus the wavefront turns more steeply in the time direction. The lower parts of figure 6.9 show the electric field and the magnetic flux.

Example 6.5. In the previous examples, we did not fully use Lorentzian (or Minkowskian) orthogonality. In each case, the primal and dual edges were either purely spatial or in the time direction relative to the observer, in whose coordinates the mesh was represented. Now, we are taking a slightly more complicated case, where the domain is a *rhomboid*, i.e., a parallelogram with sides of equal length. Both sides of the rhomboid tilt the same angle toward the lightcone (see figure 6.10), and the facets of the mesh are of the same shape, as in figure 6.1. In fact, there exists another observer, in whose coordinates the spacelike edges are purely spatial, and the timelike ones are in the time direction. However, for demonstration purposes, we choose to solve this problem in the coordinates of the observer who sees the domain as a rhomboid. Otherwise, the situation is identical to example 6.1 and figure 6.6; the bottom spacelike boundary acts as the initial boundary, and the spatial boundary conditions are placed on the timelike boundaries.

The upper part of figure 6.10 gives the solved potential, and the lower parts show the electric field and magnetic flux. If we perform the Lorentz transformation from the coordinates of the rhomboid observer to those of the square observer, the field patterns become identical to those in figure 6.6.

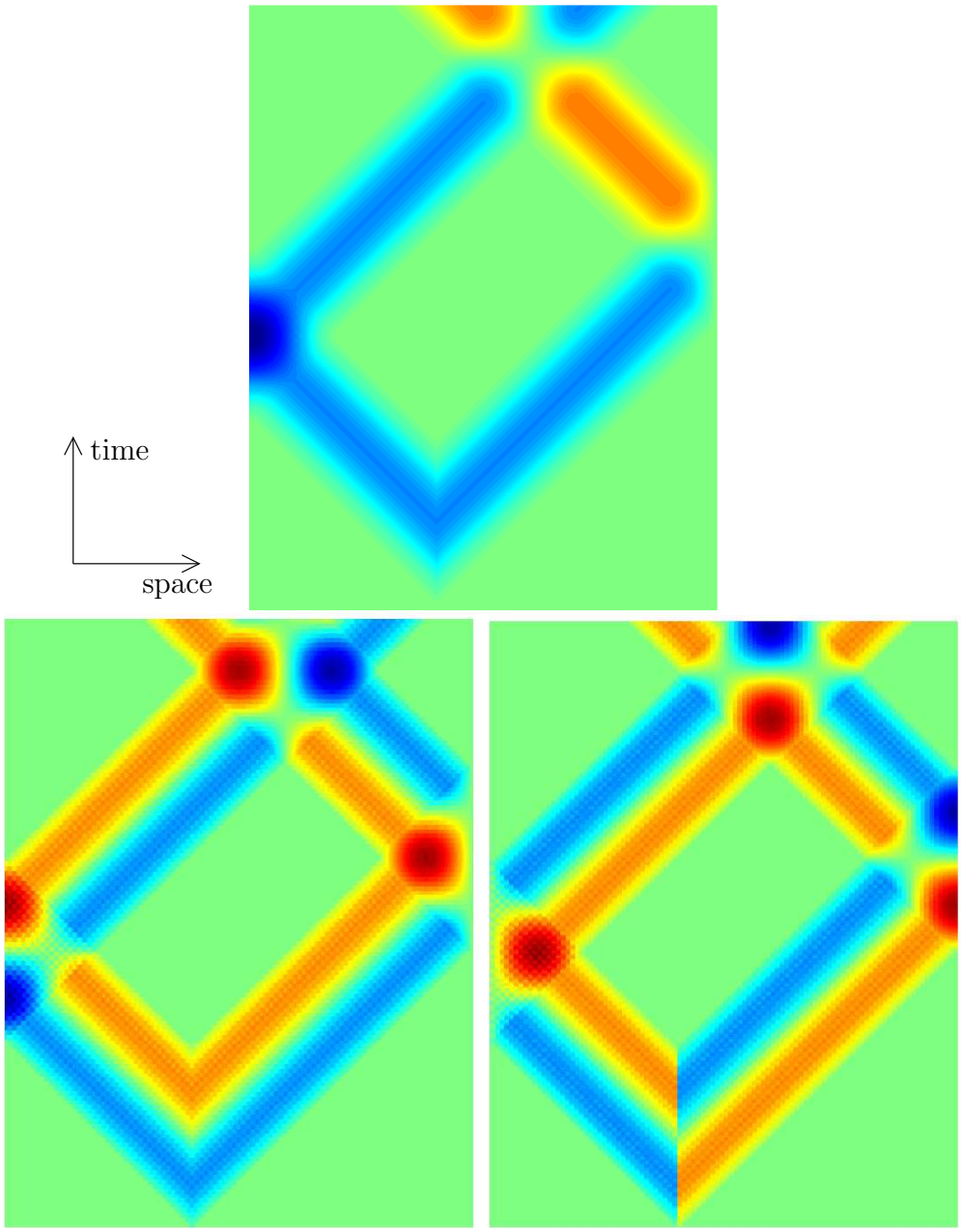


Figure 6.8: A two-dimensional example: a vacuum with a homogenous boundary condition for A on the initial and right boundaries and a homogenous G-condition on the initial and left boundaries. A nonvanishing source charge-current is given as the source term inside the domain at one point: At this spatial point, J is a period of sin as a function of time. Above: the distribution of the potential A. Bottom left: the electric field $\mathcal{E}$. Bottom right: the magnetic flux $\mathcal{B}$.

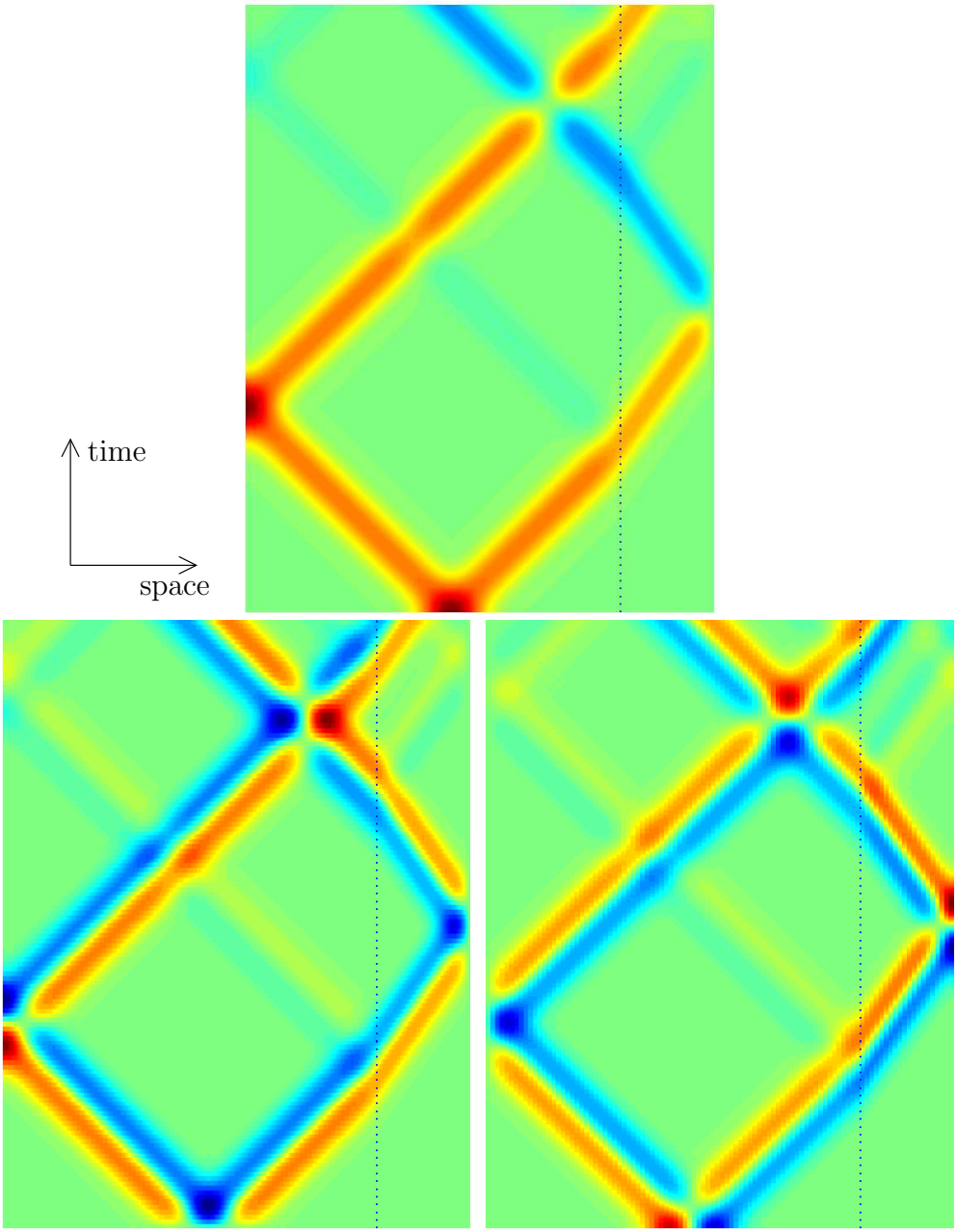


Figure 6.9: A two-dimensional example, where the domain consists of two different materials. The left part is a vacuum and the right part is made from a dielectric material with the dielectric constant $\varepsilon_r = 2$. Otherwise, the situation is identical to that in figure 6.6. Above: the distribution of the potential A; the place of the material interface is shown with a dotted line. Bottom left: the electric field $\mathcal{E}$. Bottom right: the magnetic flux $\mathcal{B}$.

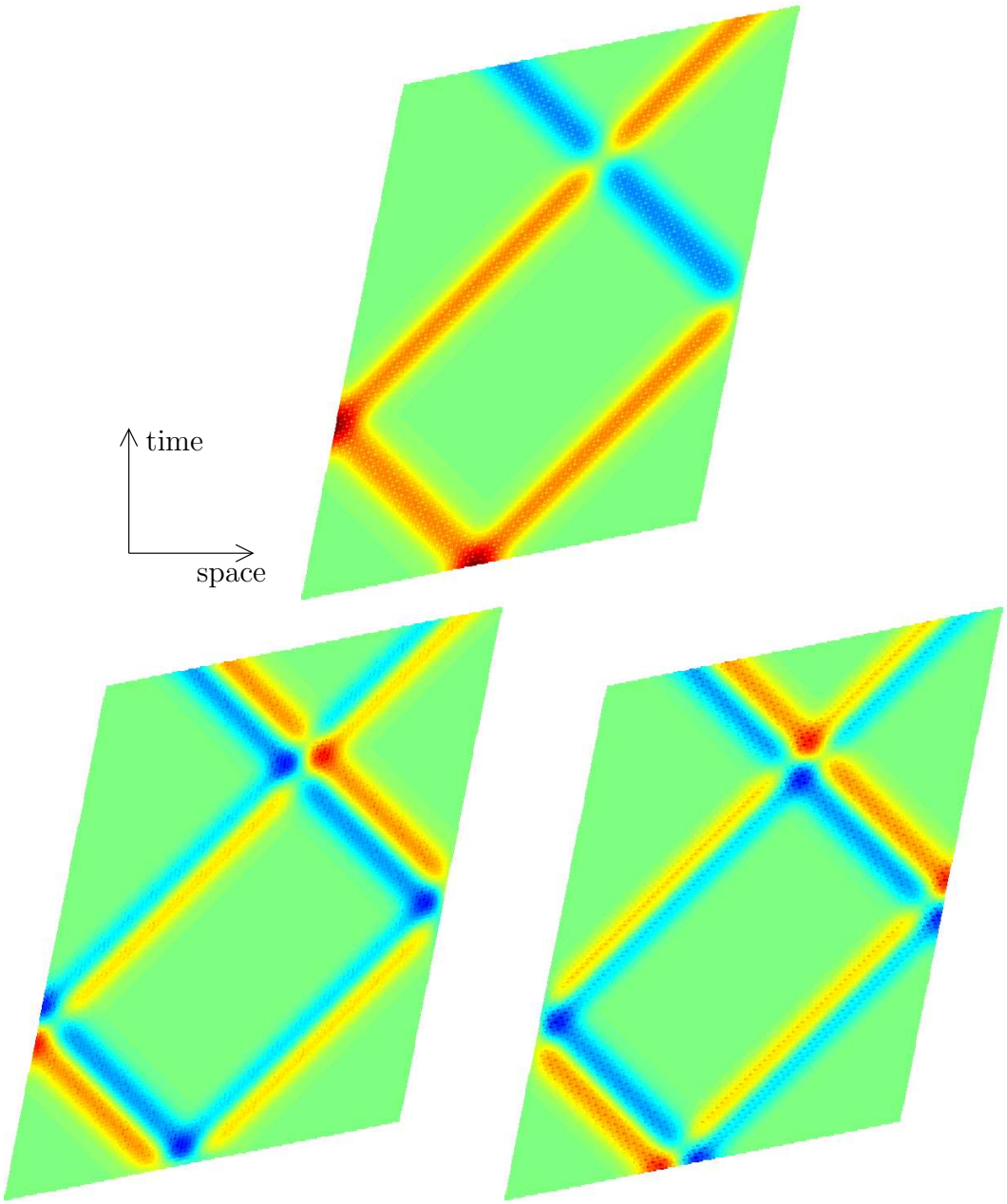


Figure 6.10: A two-dimensional example of a rhomboidal domain. Apart from the geometry, the situation is identical to that in figure 6.6. Above: the distribution of the potential A. Bottom left: the electric field $\mathcal{E}$. Bottom right: the magnetic flux $\mathcal{B}$.

Example 6.6. Until now, all examples have more or less dealt with the Cartesian product of a spatial domain with a time interval. However, this is not the only alternative as long as we have as many unknowns as independent equations after the boundary conditions and the restriction of the equations.

Let us consider a potential produced by the point source of A on the initial boundary, as shown in the upper part of figure 6.11. There, the domain is a vacuum, and homogenous F-conditions are placed on the spatial boundaries. As we can see, the potential vanishes everywhere except in certain thin lines. These lines consist of the future lightcone of the source and the future ligthcones of the wavefronts reflected from the boundary. Naturally a question arises: why do we need to solve the potential in regions where, by causality, nonzero fields never exist. We could remove the mesh outside the domain of influence of the source and place a homogenous A-condition on the new boundaries. In this example, we have gone even further. Because we know that the fields always advance at the speed of light in a vacuum and never slower, we have removed the mesh also from the inside region of the source's lightcone. The lower part of figure 6.11 shows the results of two cases with a different number of elements left around the lightcone. In the bottom left figure, only a stripe of three facets nearest to the lightcones are present. The right side is the ultimate version; only the facets lying on the lightcone remain in the mesh. In the first one, the new boundary nodes have a homogenous A-condition; in the second one, every node is a boundary node, and the homogenous A-condition is placed only for the outermost boundary nodes. The results agree superbly with each other and with the full mesh results in the upper part of figure 6.11—even in the rather harsh mesh of the bottom right of figure 6.11.

The significance of point source results lies in the theory of Green's functions. If we know the analytic solution with Dirac's delta source, i.e., the problem where $J'(x,t) = \delta(x - x_{\text{source}}, t - t_{\text{source}})$ centered at coordinates $(x_{\text{source}}, t_{\text{source}})$, we obtain a solution to a general source with a convolution-type integral. Naturally, if a boundary exists close to the source, the situation is more complicated; but in a boundaryless case or far from boundaries, point source solutions are very helpful. In numerical computation, we have an analogue of Green's functions: a solution to a problem with a point source at one time instant—i.e., we have a source at one node. Finite Green's functions can be used like their analytic counterparts.

This example is a solution to one of the challenge problems introduced in chapter 2. We return to lightcone meshes also in the three-dimensional spacetime case.

In each above example, the mesh facets were squares or rhomboids; i.e.,

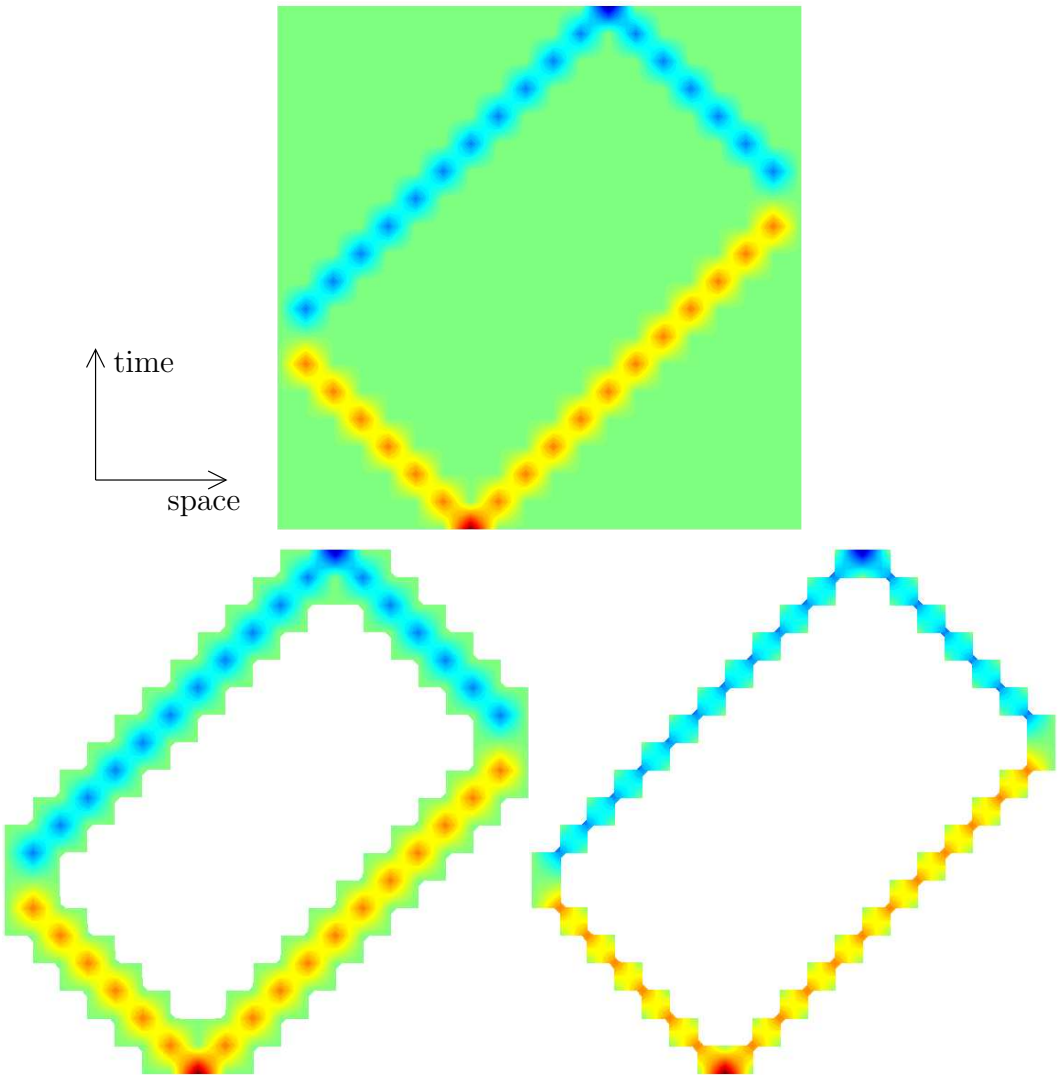


Figure 6.11: A two-dimensional example of lightcone meshes. A point source for the potential A is placed on the initial boundary. It creates two wavefronts, which advance at the speed of light; i.e., the wavefronts travel along the lightcone of the source until they collide with the spatial boundaries. In this example, all the spatial boundaries have a homogenous F-boundary condition. Above: for comparison, the problem is solved with a full mesh. Below: only the vicinity of the lightcone of the source point is meshed—and additionally, the lightcone regions of the reflected wavefronts. The distribution of the potential A is solved in those lightcone meshes. In the bottom left figure, a largish region around the lightcones is meshed; in the bottom right figure, only the facets on the lightcone are present. Note how the results agree perfectly.

they were of equal size in spatial and temporal directions (with unit speed of light). This choice is in accordance with the magic time-step (or the Courant condition) of the Yee scheme [72, 92] and the geometric methods [2, 88]. In the vacuum case, computation becomes unstable, if the facets are taller than the magic time-step in the temporal direction. Additionally, if we diminish the temporal dimension below the maximum value, the numerical error in the result increases. Thus in a vacuum, square elements are the best choice for the two-dimensional spacetime method.

6.2.2 Spacetime of two spatial dimensions

In three-dimensional spacetime, there are two possible choices: either F is a 1-cohain and G a 2-cochain or F is a 2-cochain and G a 1-cochain. The choices model two different situations. Because the computation remains more straightforward when the potential A is a 0-cochain, we choose the first one of the two options.

First, we formulate the three-dimensional spacetime electromagnetic problem and examine the method's computational aspects. Second, we illustrate the method with several numerical examples.

Formulation and computation

As discussed in section 5.4.2, everything related to the primal variables F and A is identical to the two-dimensional spacetime case. In contrast, all dual variables, e.g., G and J, are raised by one dimension. Thus for $F \in C^1(K)$, $G \in \tilde{C}^2(\tilde{K})$, and the known $J \in \tilde{C}^3(\tilde{K})$, we obtain the finite Maxwell laws

$$(F|\partial\Sigma) \;=\; 0, \qquad \forall\Sigma \in C_2(K), \tag{6.52}$$

$$(G|\partial\tilde{\Sigma}) \;=\; (J|\tilde{\Sigma}), \quad \forall\tilde{\Sigma} \in \tilde{C}_3(\tilde{K}). \tag{6.53}$$

The finite equation for the potential $A \in C^0(K)$ is

$$(A\,|\partial\Sigma) = (F\,|\,\Sigma), \quad \forall\Sigma \in C_1(K). \tag{6.54}$$

Now the mesh complexes K and $\tilde{K}$ must be three-dimensional. For the constitutive relations, the primal 1-cells must be Lorentzian orthogonal to the dual 2-cells. If this holds, the finite Hodge operator (6.27) works between F and G with a 1-cell σ and its dual 2-cell $\tilde{\sigma}$. On the boundary, the dual mesh is construed as described in section 6.1.3. Thus like in the two-dimensional case, the real counterpart $\boldsymbol{G} \in \mathbb{R}^{\tilde{\kappa}_2}$ of the 2-cochain G is divided into two parts according to equation (6.33). Now, $\boldsymbol{G}' \in \mathbb{R}^{\kappa_1}$, $\boldsymbol{G}_{\tilde{L}} \in \mathbb{R}^{\tilde{\kappa}_2(\tilde{L})}$, $\mathbf{P}_G \in \mathbb{R}^{\kappa_1 \times \tilde{\kappa}_2}$, and $\mathbf{Q}_G \in \mathbb{R}^{\tilde{\kappa}_2(\tilde{L}) \times \tilde{\kappa}_2}$. The real array equivalent of the potential 0-cochain A is

$\boldsymbol{A} \in \mathbb{R}^{\kappa_0}$, and it is partitioned according to (6.38), as in the two-dimensional spacetime case. Similarly, the real array versions of (6.53) and (6.54)—with proper restrictions—are identical to (6.37) and (6.46), respectively.

Let us pull everything together. In three-dimensional spacetime, the finite electromagnetic wave problem is as follows: for known $\boldsymbol{J} \in \mathbb{R}^{\kappa_1^{\text{eq}}}$, $\boldsymbol{A}_{L_F} \in \mathbb{R}^{\kappa_0(L_F)}$, and $\boldsymbol{G}_{\tilde{L}} \in \mathbb{R}^{\tilde{\kappa}_2(\tilde{L})}$, find $\boldsymbol{A}' \in \mathbb{R}^{\kappa_0^A}$ so that

$$\mathbf{C}'^T \mathbf{H}_M \mathbf{C} \boldsymbol{A}' = \boldsymbol{J}', \tag{6.55}$$

where

$$\mathbf{C}'^T = \mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_2 \mathbf{P}_G^T, \tag{6.56}$$

$$\mathbf{C} = \mathbf{d}_0 \mathbf{P}_A^T, \tag{6.57}$$

$$\boldsymbol{J}' = \mathbf{P}_{\text{eq}} \boldsymbol{J} - \mathbf{P}_{\text{eq}} \tilde{\mathbf{d}}_2 \mathbf{Q}_G^T \boldsymbol{G}_{\tilde{L}} - \mathbf{C}'^T \mathbf{H}_M \mathbf{d}_0 \mathbf{Q}_A^T \boldsymbol{A}_{L_F}, \tag{6.58}$$

where $\mathbf{H}_M \in \mathbb{R}^{\kappa_1 \times \kappa_1}$, $\mathbf{P}_{\text{eq}} \in \mathbb{R}^{\kappa_0^{\text{eq}} \times \tilde{\kappa}_3}$, $\mathbf{P}_A \in \mathbb{R}^{\kappa_0^A \times \kappa_0}$, and $\mathbf{Q}_A \in \mathbb{R}^{\kappa_0(L_F) \times \kappa_0}$.

In spatially two-dimensional Yee schemes and geometric methods, the length of the time-step Δt must obey the Courant condition (with $c = 1$) [72, 88]

$$\Delta t \leq \frac{1}{\sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2}}}, \tag{6.59}$$

where Δx and Δy are the grid sizes in the two spatial dimensions. This same result is valid also for our case; our 3-cells in three-dimensional spacetime are cuboids (rectangular hexahedra) of dimensions $\Delta x \times \Delta y \times \Delta t$. If the elements are spatially cubes, i.e., $\Delta x = \Delta y$, we must have

$$\Delta t \leq \frac{\Delta x}{\sqrt{2}}. \tag{6.60}$$

Thus our 3-cells cannot be cubes but cuboids with a temporal dimension of at least $\frac{1}{\sqrt{2}}$-times smaller than the spatial dimensions. In fact, we use the maximum value given by conditions (6.59) and (6.60) for a 3-cell along the temporal dimension.

Otherwise, the numerical computation and linear algebra resemble the two-dimensional spacetime case. In our simple prototype problem (the three-dimensional analog of the problem in figure 5.7), the system matrix in (6.55) is lower triangular, and the linear system can be solved by forward substitution. Our meshes are constructed from layers of nodes at a fixed time instant, and the nodes are enumerated continuously layer by layer from the initial surface, each layer in the same way. As in the two-dimensional case, with this mesh solving (6.55) is operation by operation identical to the leap-frog time-stepping method.

Examples

An electromagnetic wave problem in a three-dimensional spacetime—described by a geometry, materials, boundary conditions, and sources—can be solved by finding a solution to equation (6.55). As in two-dimensional spacetime, we look for a solution to the three-dimensional analogue to our prototype problem geometry in figure 5.7. Now the boundary of the domain contains six separate faces, one initial, one final, and four spatial boundaries. The Faraday field F (and thus A) vanishes on the edges of three spatial boundaries, and G vanishes on the dual edges of the fourth side—in the following figures, the G-condition is placed on the left boundary. G vanishes also on the initial boundary, and a nonvanishing initial A-condition serves as the source term.

We give three examples of different sources, materials, and geometries. The first one is the simplest vacuum case in our prototype geometry. The second example adds a material interface (between the vacuum and the dielectric medium) to the domain. The last example discusses the feasibility of lightcone meshes in a three-dimensional spacetime.

Example 6.7. The first three-dimensional spacetime example is analogous to the two-dimensional example 6.1. The potential A on the initial surface is forced into the form of a two-dimensional Gaussian function—centered first at the center of the spatial domain with a width of several mesh edges. The domain is a vacuum. The primal mesh comprises 3-cells that are equal-sized in the spatial dimensions but $\frac{1}{\sqrt{2}}$-times smaller in the time direction. Figure 6.12 shows the solved potential on several planes of simultaneity relative to the observer, in whose coordinates the mesh is represented—i.e., the observer, in whose coordinates the mesh do not change as a function of time. The figure shows the potential at four separate time instants. The upper left figure shows the initial situation. In the upper right figure, the potential has progressed equally in all spatial directions and a circular wavefront is formed. In the bottom left figure, the wavefront has progressed close to the boundary. Behind the main wavefront follows another noticeable wavefront, whose amplitude is smaller and sign opposite to the main wavefront. This is characteristic of spatially two-dimensional waves. Because the two-dimensional case models (for example) the perpendicular plane of a very long source, the field distribution on this single plane also originates from a source outside the plane. The main wavefront is formed from the source on the plane. Because the distance from the off-plane sources is longer, their contribution travels behind the main wavefront and has a smaller amplitude. Thus every wavefront has a tail behind it.

In the bottom right part of figure 6.12, the wavefronts have reflected

from the boundaries. The field pattern is not symmetric, because there is a homogenous G-condition on the left side and a homogenous F-condition on the other sides. Thus the wavefront reflects with a positive sign from the left side and with a negative sign from the other sides.

Figure 6.13 shows the potential and components of F at one time instant $t = 60/\sqrt{2}$, i.e., at the 60^{th} time-step. The temporal component of F is the electric field (as detected by the chosen observer); it is shown in the upper right figure. The spatial components of F are the two components of magnetic flux shown in the middle of figure 6.13. At the bottom of the figure, the potential of time $t = 60/\sqrt{2}$ is shown in two different cases. At bottom left, the source has moved rightwards and, at bottom right, leftwards. This shows clearly the difference between the reflection from the homogenous F-condition (on the left) and that from the homogenous G-condition (on the right).

Figure 6.14 represents the potential and the components of F on the xt-plane of $y = 50$. This plane passes through the source point. The upper left part shows the progress of the wavefront in a lightcone type pattern, as in the above two-dimensional spacetime examples. This differs from the two-dimensional case in that the fields thin out as the wave travels further from the source. This feature naturally originates from the sphere-wave nature of the wavefront—the energy from the source is divided into an ever growing cross-section as the wavefront advances. In addition, the tail of the wavefront is visible in these field patterns. The smallness of the y-component of $\mathcal{B}$ stems from the geometry—there is not much y-dependence of A on this plane.

If we use a mesh where the temporal dimension is smaller than $\frac{1}{\sqrt{2}}$ times the spatial dimension, the numerical error in the results increases. Thus the maximum value of the Courant condition (6.60) gives the optimum time-step size in our method.

Example 6.8. Figure 6.15 illustrates the effects of a material interface in the three-dimensional case. There is a dielectric medium of $\varepsilon_r = 2$ on the right of the domain (in the x-direction). Of the domain, 70% is vacuum and the rest 30% is filled with dielectric material. Otherwise, the situation is identical to problem 6.7.

The upper left of figure 6.15 shows the potential A at time $t = 60/\sqrt{2}$. A part of the circular wave is reflected, and another part is refracted from the material interface. In a dielectric medium, the traveling speed of the wave is $c/\sqrt{2}$; thus the radius of the circular wave becomes larger in the dielectric. The other parts of figure 6.15 show the components of the Maxwell field F, i.e., the electric field in the upper right and the components of the magnetic flux in the lower parts. They show a similar reflection-refraction division of

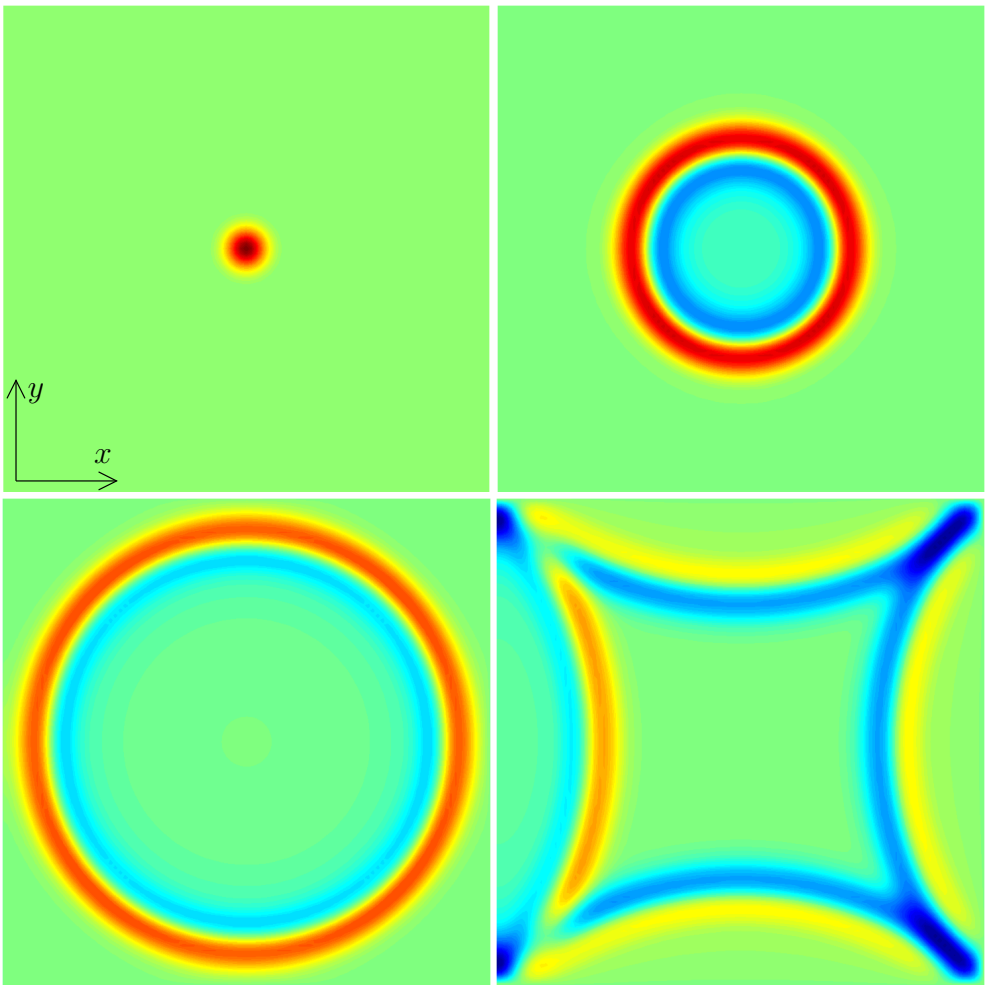


Figure 6.12: A three-dimensional example of the electromagnetic potential at various time instants; the potential's distributions are shown on purely spatial surfaces relative to the observer related to the mesh. The domain is a vacuum without source charge-currents; by the boundary condition, A vanishes on the top, right, and bottom boundary, and G vanishes on the initial and left sides. A nonvanishing initial A-condition is given as the source term: $\mathrm{t}_{\Gamma_{\mathrm{initial}}} A$ is a two-dimensional Gaussian function in the center of the spatial domain with a width of several edge sizes. The mesh has 100 3-cells in the two spatial directions and in the temporal direction. Above left: the distribution of the potential A on the initial boundary—horizontal is the x-direction, vertical the y-direction. Above right: A at time $30/\sqrt{2}$. Bottom left: A at time $60/\sqrt{2}$. Bottom right: A at time $100/\sqrt{2}$. The initial distribution is not on the same color scale as the rest.

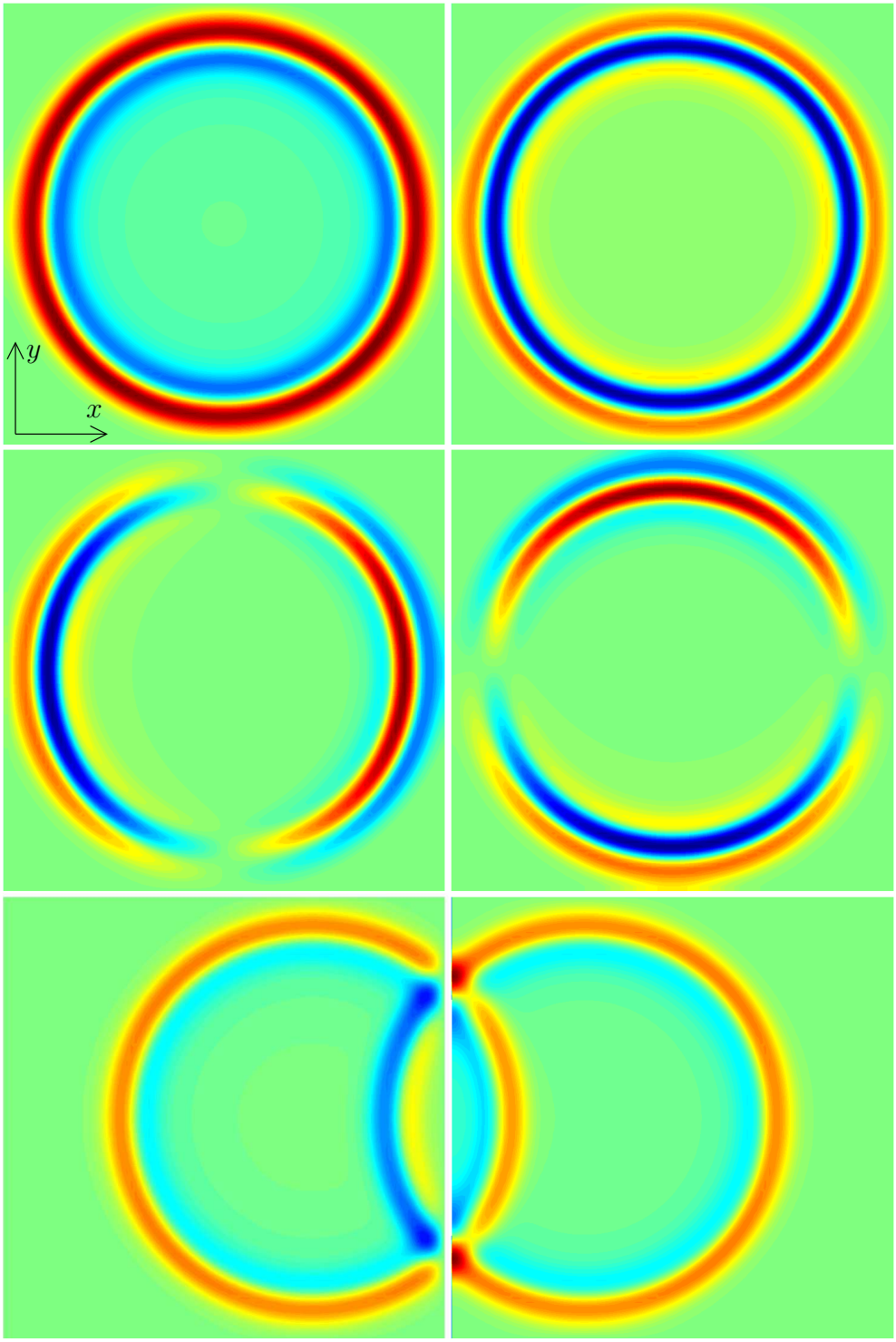


Figure 6.13: Figure 6.12 continued: fields at time $60/\sqrt{2}$. Above left: the potential A. Above right: the electric field $\mathcal{E}$. Middle left: the x-component of the magnetic flux $\mathcal{B}$. Middle right: $\mathcal{B}_y$. Bottom left: A with the source moved right. Bottom left: A with the source moved left.

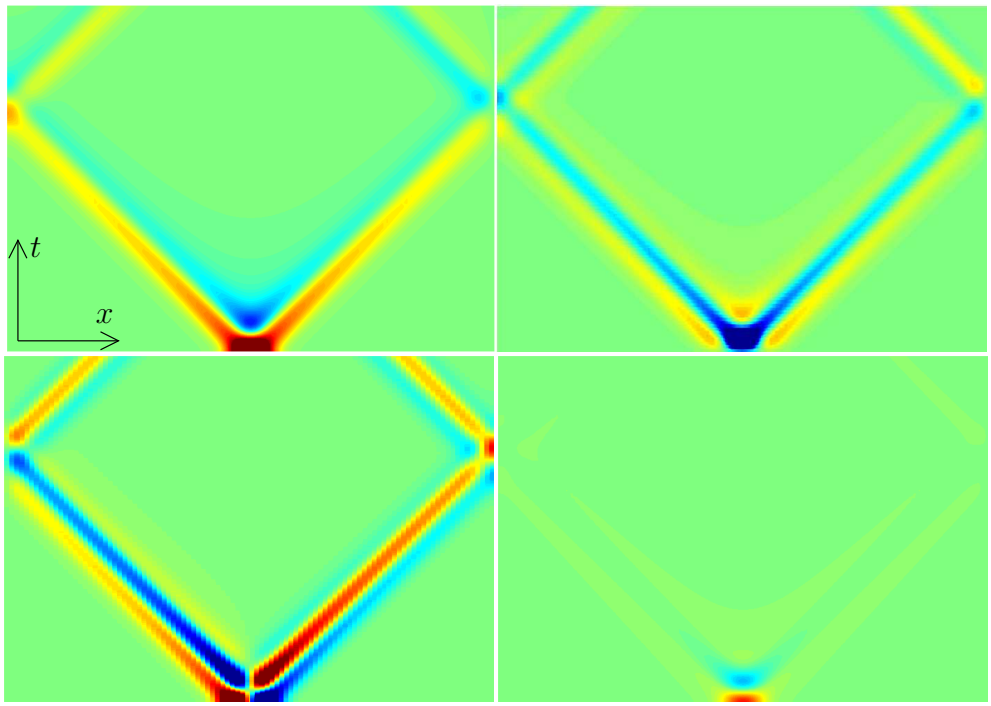


Figure 6.14: Figures 6.12 and 6.13 continued. Fields are shown on the xt-plane $y = 50$; i.e., on the plane of a fixed y in the middle of the domain. Above left: the potential A. Above right: the electric field $\mathcal{E}$. Bottom left: the x-component of the magnetic flux $\mathcal{B}$. Bottom right: the y-component of the magnetic flux $\mathcal{B}$. The color scale is identical in both components of $\mathcal{B}$.

the wave in the material interface.

Example 6.9. How do the lightcone meshes work in three-dimensional spacetime problems? It would be easy to do the same as in the two-dimensional case: take only the 3-cells that lie near the lightcone of the point source. However, three-dimensional spacetime introduces two annoying factors that obstruct this method. First, due to the $1/\sqrt{2}$ factor in the temporal dimension, no nodes lie exactly on the lightcone, only near it. This minor problem we can solve, but another difficulty ruins the whole idea of lightcone meshes. Namely, as we noticed above, there is no single lightcone-shaped wavefront in the spatially two-dimensional case. There is always a group of smaller wavefronts behind the main wave. Hence the inside of the point source's lightcone always holds a nonzero field, and we must therefore keep the inside fully meshed.

Let us demonstrate the above conclusion. For comparison, the upper left

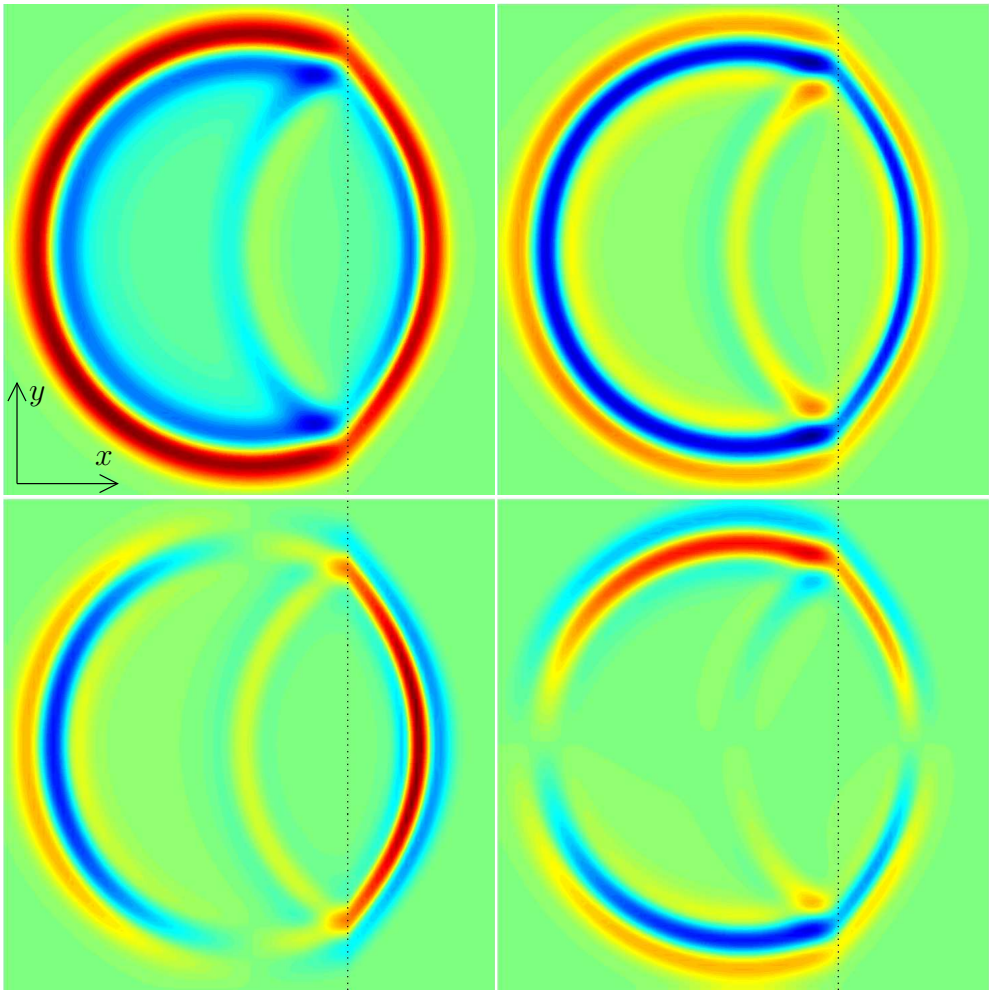


Figure 6.15: A three-dimensional example of electromagnetic fields in a domain with a material interface, shown at time $t = 60/\sqrt{2}$. On the right, 30% of the region is made from a dielectric medium with $\varepsilon_r = 2$, the rest of the domain is a vacuum. The material interface is shown with a dotted line. Otherwise, the situation is similar to the case in figure 6.13. Above left: the potential A. Above right: the electric field $\mathcal{E}$. Bottom left: the x-component of the magnetic flux $\mathcal{B}$. Bottom right: the y-component of the magnetic flux $\mathcal{B}$. The color scale is identical in both components of $\mathcal{B}$.

of figure 6.16 shows the full mesh solution of the potential A at time $t = 55/\sqrt{2}$ created by a point source on the initial surface. The main wavefront is traveling circularly outwards; in addition, the main wavefront circle retains a significant extent of A. In the upper right of figure 6.16, only the vicinity of the source's lightcone is meshed. The main wavefront resembles the full mesh pattern, but the tail of the wave is cut off. Thus the result is erroneous and must be discarded.

The bottom left of figure 6.16 represents the distribution of the potential A in a case where the inside of the lightcone is fully meshed but the outside mesh is removed. The bottom right shows A on the xt-plane that passes through the source point. These results are identical to the full mesh case. Thus we can use lightcone meshes in three-dimensional spacetime problems, if we use full meshes inside the lightcone of the point source. However, using these meshes is not as advantageous as in the two-dimensional case, because we must use a greater part of the original full mesh. Consequently, because this is not as efficient a method as in the two-dimensional case, we must turn to the four-dimensional case.

At bottom right in figure 6.16, we can observe how a lightcone-like wavefront from a point source does not fall into the nodes of the mesh but divides into a couple of neighboring nodes. Now the wavefront is not equally sharp at every time instant in every part of the circle wave. For this reason, the wavefront on the xy-plane parts in figure 6.16 does not seem to have an equal amplitude in every direction. This is due to the fact that only one node (in space and time) forms the source; thus the error cannot be fixed with a denser mesh, only by widening the source to multiple nodes in space or time.

6.3 Four-dimensional spacetime

Finally, let us focus on our main topic: numerical computation of electromagnetic wave problems in four-dimensional spacetime. In section 6.1.5, we developed a four-dimensional potential formulation, which differs from the above lower-dimensional cases mainly in that the electromagnetic potential A is now a 1-cochain instead of a 0-cochain, i.e., it is a vector potential instead of a scalar potential. Most of the formulation is identical with the lower-dimensional cases; the difference lies mainly in that we need a gauging method.

In contrast to scalar potential methods, where the result is made unique by proper boundary conditions, the vector potential requires some supplementary conditions. These extra conditions are called *gauge conditions*. Their analytic versions were discussed in section 5.3.1. In the first section

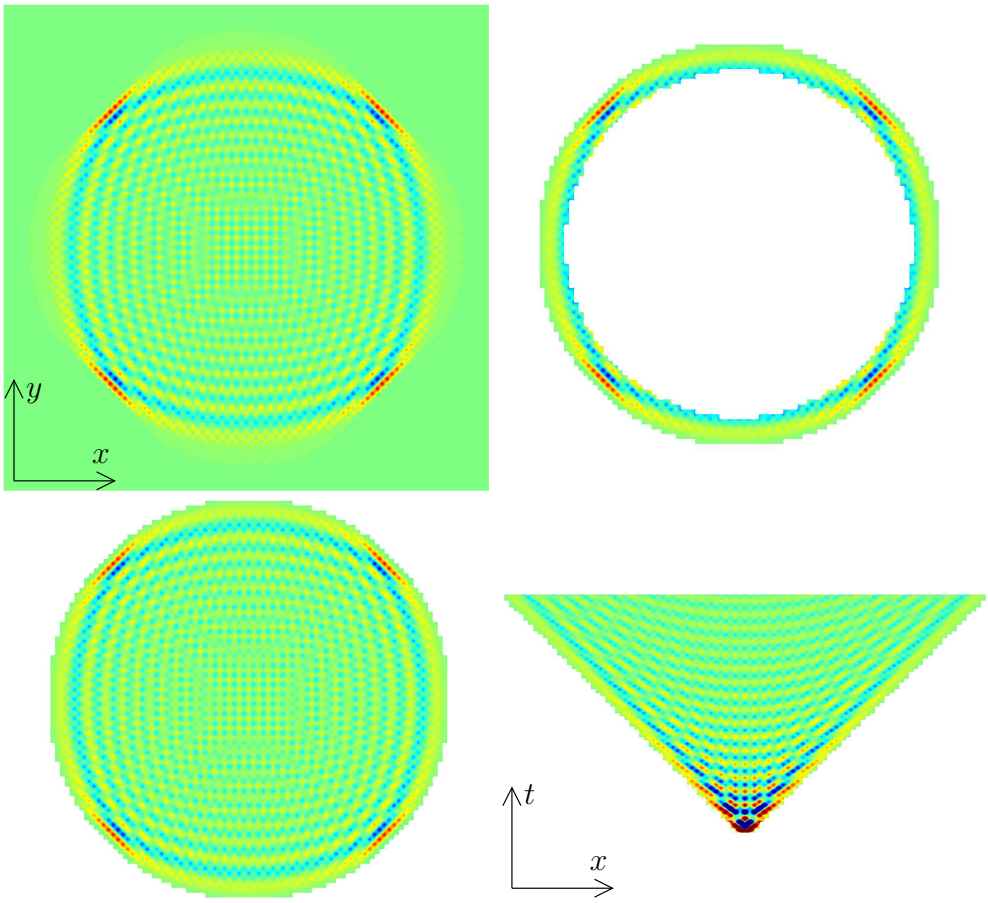


Figure 6.16: A three-dimensional spacetime example of lightcone meshes with a point source. This differs from the case in figure 6.12 in terms of the source: A vanishes on the initial boundary, except for one node in the center of the spatial domain. The source creates a circular wavefront that travels along the lightcone of the source. In the first three figures, the potential is shown at time $t = 55/\sqrt{2}$. Above left: for comparison, the problem is solved with a full mesh. In the other parts, the distribution of the potential A is solved in the lightcone meshes. Above right: only the vicinity of the lightcone of the source point is meshed. Bottom left: the inside of the lightcone is meshed throughout. Bottom right: the same result as in the bottom left part but now on the xt-plane through the source point.

below, we discuss a gauging method for finite problems, the so-called tree gauge [1].

However, besides the above, we have a second application of the gauging theory. The equations included in (6.36), or equivalently in (6.39), are linearly dependent; thus we must discard part of the equations to make them independent. This operation is related to gauging and, therefore, it is dealt with in the second section.

In the third section, we discuss computational aspects of the method with different choices of tree gauges. In the last section, we give a few examples of electromagnetic computation in four-dimensional spacetime.

6.3.1 Tree gauge

Let us consider a finite counterpart of the short exact sequence (5.45). Let K be the complex and L its subcomplex. Then

$$0 \to \mathcal{Z}^p(K,L) \xrightarrow{i} C^p(K,L) \xrightarrow{d} \mathcal{B}^{p+1}(K,L) \to 0 \tag{6.61}$$

is exact. Here $\mathcal{Z}^p(K,L)$ and $\mathcal{B}^{p+1}(K,L)$ are the spaces of relative cocycles and coboundaries modulo L in K, respectively. As a consequence of (6.61), we have $C^p(K,L)/i(\mathcal{Z}^p(K,L)) \cong \mathcal{B}^{p+1}(K,L)$ and $C^p(K,L) \cong \mathcal{Z}^p(K,L) \oplus C^p_{\mathrm{d}}(K,L)$, where d is an isomorphism from $C^p_{\mathrm{d}}(K,L)$ onto $\mathcal{B}^{p+1}(K,L)$. Thus the ungauged potential is a coset in the space $C^p(K,L)/i(\mathcal{Z}^p(K,L))$, and the gauge condition chooses a unique representative A from each coset $[A]$. In the case of electromagnetic potential, $p = 1$.

By the definition (6.9) of finite electromagnetic potential, $(F \,|\, \Sigma)$ is known if we know $(A \,|\partial\Sigma)$ for each boundary $\partial\Sigma \in \mathcal{B}_1(K,L)$ modulo L. Additionally, we get the same value for every $A \in [A]$ connected to a chosen $F \in C^2(K,L)$. Thus a member $[A]$ of the space $C^1(K,L)/i(\mathcal{Z}^1(K,L))$ can be seen as a unique homomorphism

$$[A] : \mathcal{B}_1(K,L) \to \mathbb{R} : \partial\Sigma \mapsto (F \,|\, \Sigma). \tag{6.62}$$

Thus the ungauged potential $[A]$ belongs to the dual space of $\mathcal{B}_1(K,L)$, and thus $C^1(K,L)/i(\mathcal{Z}^1(K,L)) \cong \mathcal{B}_1(K,L)$. The same applies more generally, i.e., $C^p(K,L)/i(\mathcal{Z}^p(K,L)) \cong \mathcal{B}_p(K,L)$.

As we have often mentioned before, we assume that our domain is topologically simple so that $\mathcal{B}_p(K) = \mathcal{Z}_p(K)$ $(p > 0)$; i.e., the space of boundaries equals the space of cycles. However, this does not automatically imply that also the relative cycles and boundaries are equal. For example, if L is disjoint, and if $C \in C_1(K)$ joins two separate parts of L, $C \in \mathcal{Z}_1(K,L)$ but is not necessarily a boundary of any 2-chain. However, in our case, L is connected

and, more generally, we assume that $\mathcal{B}_p(L) = \mathcal{Z}_p(L)$ $(p > 0)$. Consequently, by exact homology sequence, $\mathcal{B}_p(K, L) = \mathcal{Z}_p(K, L)$ $(p > 0)$ [30, 69]. This is the main point enabling the use of the tree gauge. Its basic idea is to find a subset of $K \setminus L$ that spans a space isomorphic to $\mathcal{Z}_p(K, L)$, and thus with $\mathcal{B}_p(K, L)$. Later, we find that the complement of a spanning tree is suitable for such a subset.

Definition 6.1. A set of p-simplices of a complex K such that it does not contain any cycle modulo L, except the null one, is called a p-dimensional *tree* (or a p-tree) modulo L.

Definition 6.2. A p-tree T_p is a *spanning tree* modulo L, if there is no strictly larger p-tree modulo L containing it. The complement of a spanning tree with respect to p-cells in $K \setminus L$ is called the *cotree* modulo L and is denoted by T_p^*. [6]

The cells of a p-cotree modulo L span $\mathcal{Z}_p(K, L)$, whereas a spanning tree modulo L spans the part of $C_p(K, L)$ isomorphic with $\mathcal{B}_{p-1}(K, L)$ by ∂, denoted by $C_p^{\partial}(K, L)$. Thus the spanning tree–cotree division is accordant with the decomposition $C_p(K, L) \cong \mathcal{Z}_p(K, L) \oplus C_p^{\partial}(K, L)$. [40]

By combining the above results, we get an isomorphism between the spaces of cotree chains and potential cosets, $C_p(T_p^*) \cong C^p(K, L)/i(\mathcal{Z}^p(K, L))$. Thus we can represent an arbitrary coset $[A'] \in C^p(K, L)/i(\mathcal{Z}^p(K, L))$ by elementary cochains associated with the p-cotree modulo L. Especially, we can represent the electromagnetic potential coset $[A] \in C^1(K, L)/i(\mathcal{Z}^1(K, L))$ with the degrees of freedom associated with the cells of an *edge-cotree* (i.e., a 1-cotree) modulo L.

By the definition of spanning tree, if we add an arbitrary p-cell into a p-dimensional spanning tree, we form a unique cycle modulo L from the p-cells in T_p and the new cell. Thus each cell in a p-cotree is uniquely associated with a p-cycle. With a 1-cotree, a co-edge closes a unique circuit with the elements of the spanning tree. Thus the potential A associated with a co-edge uniquely defines F through the 2-chain with whose boundary the co-edge is associated.

To sum up, in order to make the electromagnetic potential unique, we must build a 1-dimensional spanning tree in K and remove all the degrees of freedom associated with the spanning tree. The degrees of freedom of the gauged potential A in a cotree are enough to represent the electromagnetic field F in the domain. Additionally, the spanning tree is constructed relative to such a boundary where F or A are fixed by boundary conditions, i.e., at L_F in our case. Thus the gauged electromagnetic potential A is assumed to be zero on the p-tree modulo L_F.

6.3.2 Dual volume-tree

Our main equation (6.36), or equivalently (6.39), is about variables on the boundary of a dual volume. First, let us forget the effects of the domain's boundary and consider all the equations associated with the eight boundary 3-cells of a dual 4-cell. Because the boundary of a boundary is zero, the boundary of the eighth 3-cell equals the boundary taken from the sum of the other seven 3-cells (with a negative sign). That is, a 3-cell has the same boundary as the chain formed from the remaining seven 3-cells. Thus if we have already applied the Maxwell-Ampère law (6.36) to the seven dual 3-cells on the boundary of a dual 4-cell, those seven equations enforce the Maxwell-Ampère law also to be valid in the eighth 3-cell.[5] Thus if we enforce the Maxwell-Ampère law (6.36) in each of those boundary 3-cells, the instances of the law in the eight 3-cells are linearly dependent. And the same happens on the boundary of each 4-cell of $\tilde{K}$, and the choice of the eighth cell to each of them is arbitrary.

Let us put the above into a more exact form, yet continue with the boundaryless case. The instances of the Maxwell-Ampère law applied volume-wise are linearly dependent on every dual 3-boundary, i.e., on every 3-cycle. To keep the equations linearly independent, we must extract a subset of dual 3-cells that does not contain any cycle. The instances of the Maxwell-Ampère law in that cycle-free set would we linearly independent. As we observed in the previous section, a 3-dimensional spanning tree of the dual complex would be exactly an eligible cycle-free set.

How does the boundary of the domain affect the situation? As discussed in section 6.1.4, the equations associated with some boundary 3-cells were removed when boundary conditions were imposed. Therefore, the *dual volume-tree* must be built relative to these removed 3-cells. Let us denote by $\tilde{L}_{\text{eq}}$ the union of $(\tilde{K} \setminus \tilde{K}_{\text{eq}})$ with the 3-cells on $\tilde{L}_G$ outside Γ_{initial}; the equations associated with this part of $\tilde{K}$ were removed in section 6.1.4 from the Maxwell-Ampère laws to achieve the form (6.36). Thus we need a dual volume spanning tree modulo $\tilde{L}_{\text{eq}}$, denoted $\tilde{T}_3$. All the equations associated with the cotree $(\tilde{T}_3^*)$ are removed to get a full rank system matrix for equation (6.39).

Are these two trees, the primal edge-tree and the dual volume-tree, somehow related, or are the concepts separate? First, the volumes of a dual mesh are dual to primal edges. Consequently, a dual volume-tree is connected to a subset of primal edges. In fact, the complement of this subset forms a 1-dimensional spanning tree in K modulo L_{eq}, where L_{eq} is a primal complex on $(\Gamma_F \cap \Gamma_{\text{spatial}}) \cup \Gamma_{\text{final}}$. Thus the dual volume-tree is dual to a 1-cotree modulo L_{eq}. Thus in both cases we are effectively dealing with primal 1-

[5] By current continuity (5.11), we also know the source current of the eighth 3-cell.

cotrees, which are made, however, relative to different subcomplexes. Moreover, there is no imminent reason why the two trees should be similar in any way. Nevertheless, both trees must have the same number of edges, denoted by $\kappa_1(T_1^*) = \kappa_1(\tilde{T}_3)$.

6.3.3 Computational aspects

For computation, we must restrict the main equation (6.39) to apply only to the potential A in the cotree T_1^* and the equations associated with the tree $\tilde{T}_3$. Therefore, let us define a new projection matrix, which chooses those elements of $\boldsymbol{A}' \in \mathbb{R}^{\kappa_1^A}$ that are associated with primal edges in T_1^*; i.e., we define $\mathbf{P}_{T_1^*} \in \mathbb{R}^{\kappa_1(T_1^*)\times\kappa_1^A}$ so that the gauged potential array is $\boldsymbol{A} = \mathbf{P}_{T_1^*}\boldsymbol{A}'$. Similarly, let us define a second projection matrix $\mathbf{P}_{\tilde{T}_3} \in \mathbb{R}^{\kappa_1(\tilde{T}_3)\times\kappa_1^A}$ so that it chooses those equations from $\mathbf{C}'^T\mathbf{H}_M\mathbf{C}$ that are associated with the dual volumes in $\tilde{T}_3$.

After the above definitions, the gauged version of (6.39) is written in the form

$$\mathbf{P}_{\tilde{T}_3}\mathbf{C}'^T\mathbf{H}_M\mathbf{C}\mathbf{P}_{T_1^*}^T\boldsymbol{A} = \boldsymbol{J}'. \tag{6.63}$$

Now the four-dimensional electromagnetic wave problem can be solved by solving $\boldsymbol{A} \in \mathbb{R}^{\kappa_1(T_1^*)}$ in (6.63) from the known values of $\boldsymbol{J}'$, as defined in equation (6.42). What is left open is the effect of the structures of the two trees on the system matrix $\mathbf{P}_{\tilde{T}_3}\mathbf{C}'^T\mathbf{H}_M\mathbf{C}\mathbf{P}_{T_1^*}^T$—and how this affects the linear algebra in solving (6.63).

Let us consider the spanning trees in our prototype geometry of figure 5.7.[6] Because the geometry is so austere, we can choose very straightforwardly structured trees. For example, a three-dimensional tree on the initial hyperplane combined with all t-directional edges forms a (nonrelative) tree in our hypercuboidal domain. In the relative case, the situation is even simpler. Because L_F contains the initial boundary, the set of all the t-directional edges that do not lie on any spatial boundary with a F-condition forms a tree modulo L_F. Additionally, the duals of those edges form a dual volume-tree modulo $\tilde{L}_{\text{eq}}$.

We decide to use relative trees that contain only edges (and dual volumes) in one direction. The above case is one example, and a tree modulo L_F could be similarly constructed from the x-, y- and z-directional edges. We must be careful that we build a relative tree, i.e., that it does not form cycles modulo

[6]Unfortunately, all the numerical demonstrations—including the demonstration of the spanning trees—are made with a mesh with hyperbrick elements for above-mentioned reasons. However, nothing forbids us from using more general spanning trees in more general meshes.

L_F. In the following examples, all the spatial boundaries have a boundary condition for F; i.e., their edges belong to L_F. Therefore, in this case, we cannot include all the x-directional edges in the tree, but must leave, e.g., all the edges touching one yz-boundary out of the edge tree. Consequently, we find easily x-, y-, z-, and t-directional edge-trees modulo L_F.

Similarly, we get dual volume-trees from duals to edge cotrees modulo L_{eq} and call them x-, y-, z-, and t-directional dual volume-trees modulo L_{eq}. For example, the cotree of an x-directional dual volume-tree consists of dual volumes on yzt-hyperplanes. In our case, both L_{eq} and L_F contain all spatial boundary cells; in addition, L_{eq} contains the cells on the final boundary and L_F the ones on the initial boundary. Except for cells on the initial and final surfaces, an x-directional dual volume tree is cell by cell dual to the cotree of an x-directional primal edge-tree. Because both of these trees can be identified by a single direction, we can identify the edge-tree and the dual volume-tree in use by two directions. For example, in the following, the zt-tree means that the edge-tree is z-directional and the dual volume-tree is t-directional.

Figure 6.17 presents the sparsity pattern of the gauged system matrix $\mathbf{P}_{\tilde{T}_3}\mathbf{C}'^T\mathbf{H}_M\mathbf{C}\mathbf{P}_{T_1^*}^T$ in our prototype problem geometry in the case of a zt-, tz-, tt-, xz-, and zz-tree. The four first ones correspond to the cases of the first example below. As we can see, only in the tt-tree case the system matrix is lower diagonal. Accordingly, only in that case the forward substitution solving process of problem (6.63) resembles strictly the explicit leap-frog time-stepping method. However, with Gaussian elimination, the zt-, tz-, and xz-tree cases are solved with almost the same effort as the lower diagonal case. Loosely speaking, the solving process in the latter cases resembles some implicit but efficient time-stepping scheme, possibly a different scheme for each distinct spanning tree.

However, the zz-tree case is different. There Gaussian elimination causes an enormous fill-in; i.e., an introduction of a vast number of nonzero entries into the matrix that uses up a huge amount of memory. Therefore, it is numerically impossible to use edge and dual volume-trees, which are dually identical and consist entirely of spatial edges.

Remark 6.2. Interpretation by time-stepping could be used to bypass one of drawbacks of the method: the need to fix the number of time-steps before the mesh generation and solution process. In some practical problems, the time-stepping is terminated when some solution-based parameter (or CPU time) reaches a pre-determined limit and the number of time-steps is not known in advance. If our method is rewritten with time-stepping approach, the number of time-steps would not need to be fixed beforehand.

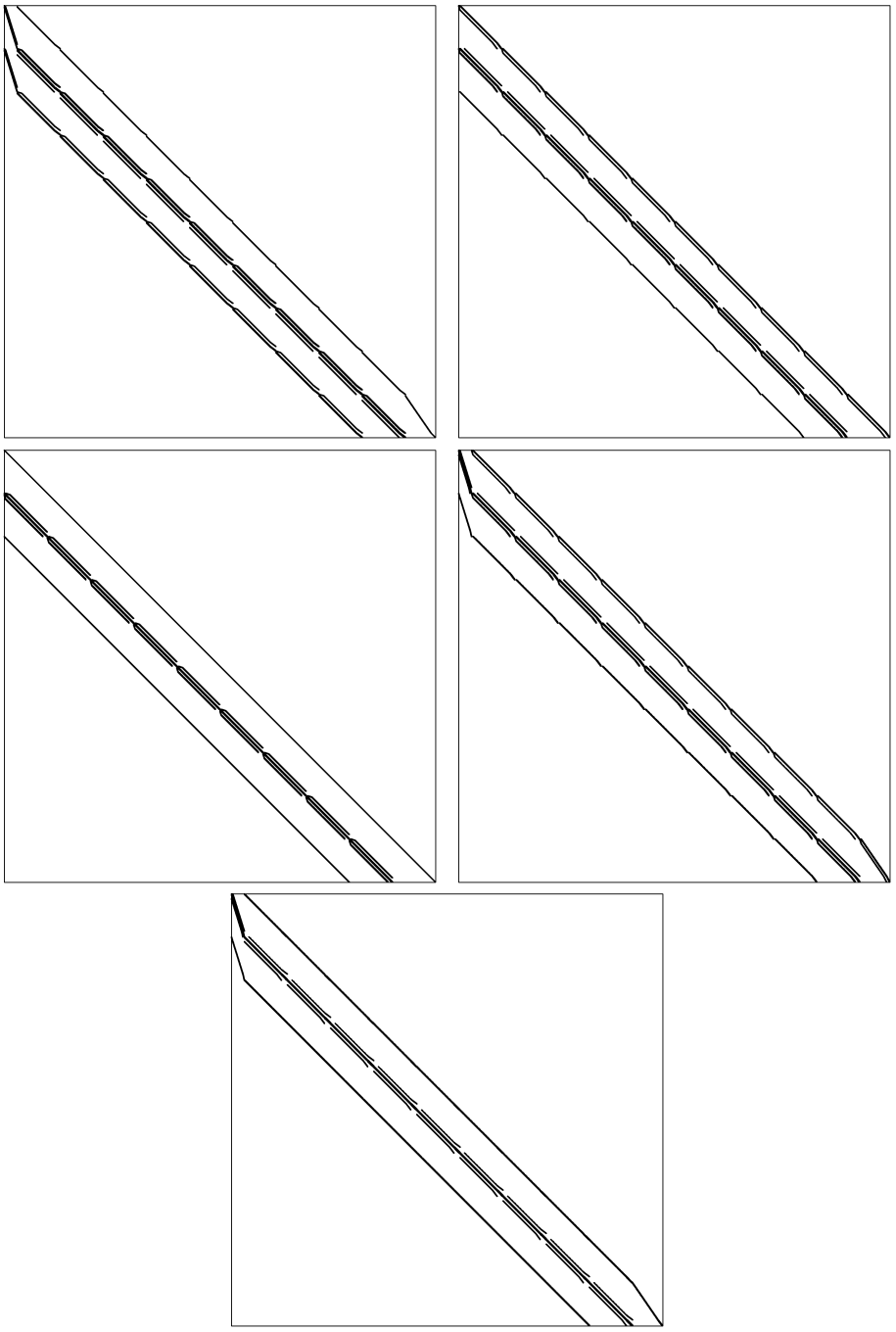


Figure 6.17: Sparsity pattern of the system matrix $\mathbf{P}_{\tilde{T}_3}\mathbf{C}'^T\mathbf{H}_M\mathbf{C}\mathbf{P}_{T_1^*}^T$ with different choices of edge-tree and dual volume-tree in a mesh with about 50000 edges. Above left: T_1 consists of z-directional edges, and $\tilde{T}_3$ is t-directional. Above right: t-directional T_1 and z-directional $\tilde{T}_3$. Middle left: a tt-tree. Middle right: an xz-tree. Bottom: a zz-tree.

In spatially three-dimensional geometric methods, the time-step Δt is restricted by the Courant condition (with $c = 1$) [72, 88]

$$\Delta t \leq \frac{1}{\sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2}}}, \tag{6.64}$$

where Δx, Δy, and Δz are grid sizes in spatial dimensions. We can use the same result; our 4-cells are hypercuboids of dimensions $\Delta x \times \Delta y \times \Delta z \times \Delta t$ with Δt restricted by (6.64). If the elements are spatially cubes, i.e., $\Delta x = \Delta y = \Delta z$, we must have

$$\Delta t \leq \frac{\Delta x}{\sqrt{3}}. \tag{6.65}$$

Thus our 4-cells could not be hypercubes, but rather hypercuboids with a temporal dimension of at least $\frac{1}{\sqrt{3}}$-times smaller than the spatial dimensions. In fact, we use the maximal size for the 4-cell given by (6.64) along the temporal dimension.

Remark 6.3. In section 5.2.2, we concluded that (in the electromagnetic wave problem) the Maxwell source equations are not necessary after the initial boundary, because they are already enforced by the Faraday and Ampère laws. That is, the source equations for $\mathcal{B}$ and $\mathcal{D}$ depend on other Maxwell laws. In the light of the above, we could interpret that a t-directional dual volume-tree is used canonically in numerical methods for three-dimensional electromagnetic wave problems. Nothing forbids us from using any other dual volume-tree in these methods.

6.3.4 Examples

We can solve an electromagnetic wave problem in four-dimensional spacetime by solving equation (6.63) in a given geometry. Let us look for a solution in our prototype problem geometry of figure 5.7. In all examples, the domain is a vacuum. We give a homogenous boundary condition for F on all spatial boundaries. In the first example, we demonstrate the effects of different choices of edge and dual volume-trees. The second example shows what happens if we pose an unphysical current source. The last example offers a more comprehensive solution to the problem in a rather large mesh.

Example 6.10. Figures 6.18, 6.19, 6.20, and 6.21 show the effects of different trees. In all cases, we use a nonvanishing initial A-condition as the source. The potential's x-component on the initial hypersurface is forced into the form of a three-dimensional Gaussian function centered in the center of the spatial domain. The other components of A on Γ_{initial} are zero.

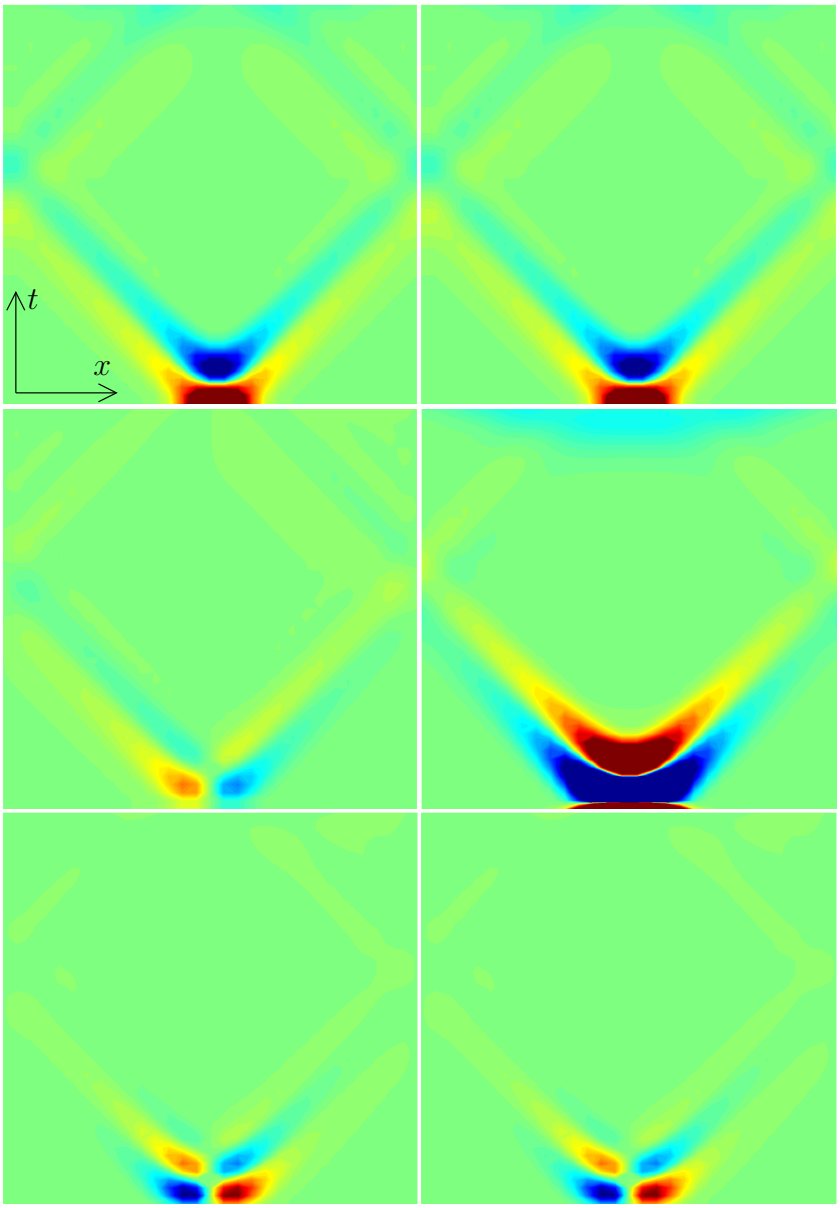


Figure 6.18: A four-dimensional example with a z-directional edge-tree and a t-directional dual volume-tree. The potential and the fields are shown on an xt-plane passing through the source. The x-component A_x of A is given as the initial condition source term: $\mathrm{t}_{\Gamma_{\mathrm{initial}}} A_x$ is a three-dimensional Gaussian function in the center of the spatial domain with a width of several edge sizes. The mesh has 30 3-cells in all spatial dimensions and 50 in the temporal direction. Above left: the x-component of the electromagnetic potential A. Above right: A_y. Middle left: A_t. Middle right: the x-component of the electric field $\mathcal{E}$. Bottom left: $\mathcal{E}_y$. Bottom right: $\mathcal{E}_z$.

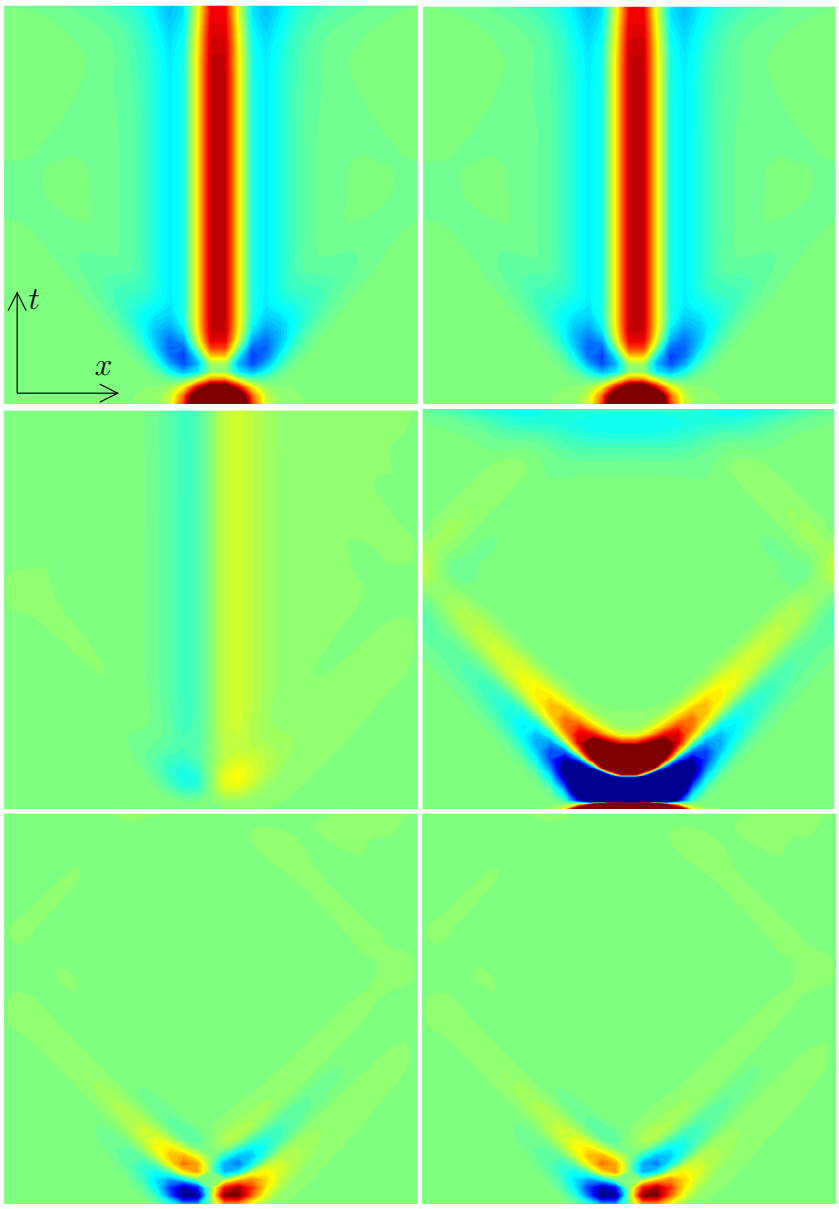


Figure 6.19: Figure 6.18 continued, now with a tz-tree. Above left: the x-component of the electromagnetic potential A. Above right: A_y. Middle left: A_z. Middle right: the x-component of the electric field $\mathcal{E}$. Bottom left: $\mathcal{E}_y$. Bottom right: $\mathcal{E}_z$.

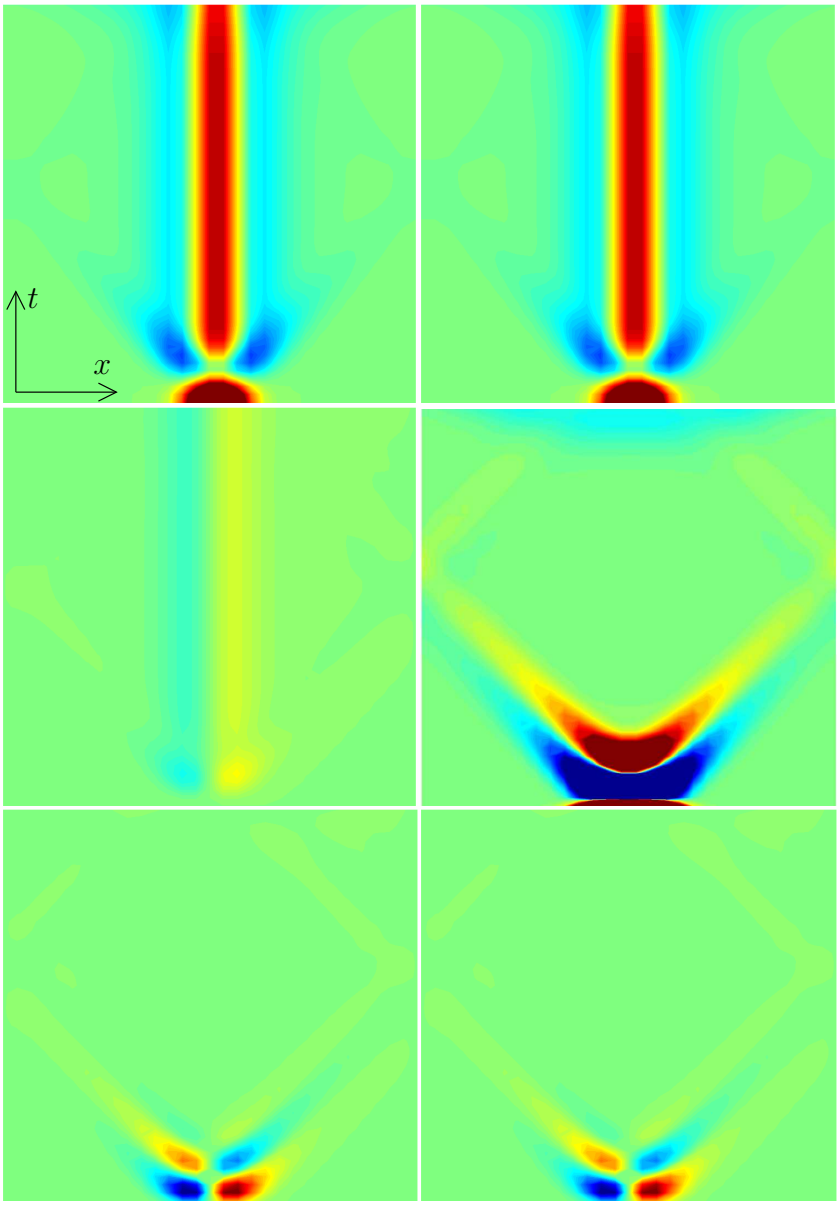


Figure 6.20: Figures 6.18 and 6.19 continued, now with a tt-tree. Above left: the x-component of the electromagnetic potential A. Above right: A_y. Middle left: A_z. Middle right: the x-component of the electric field $\mathcal{E}$. Bottom left: $\mathcal{E}_y$. Bottom right: $\mathcal{E}_z$.

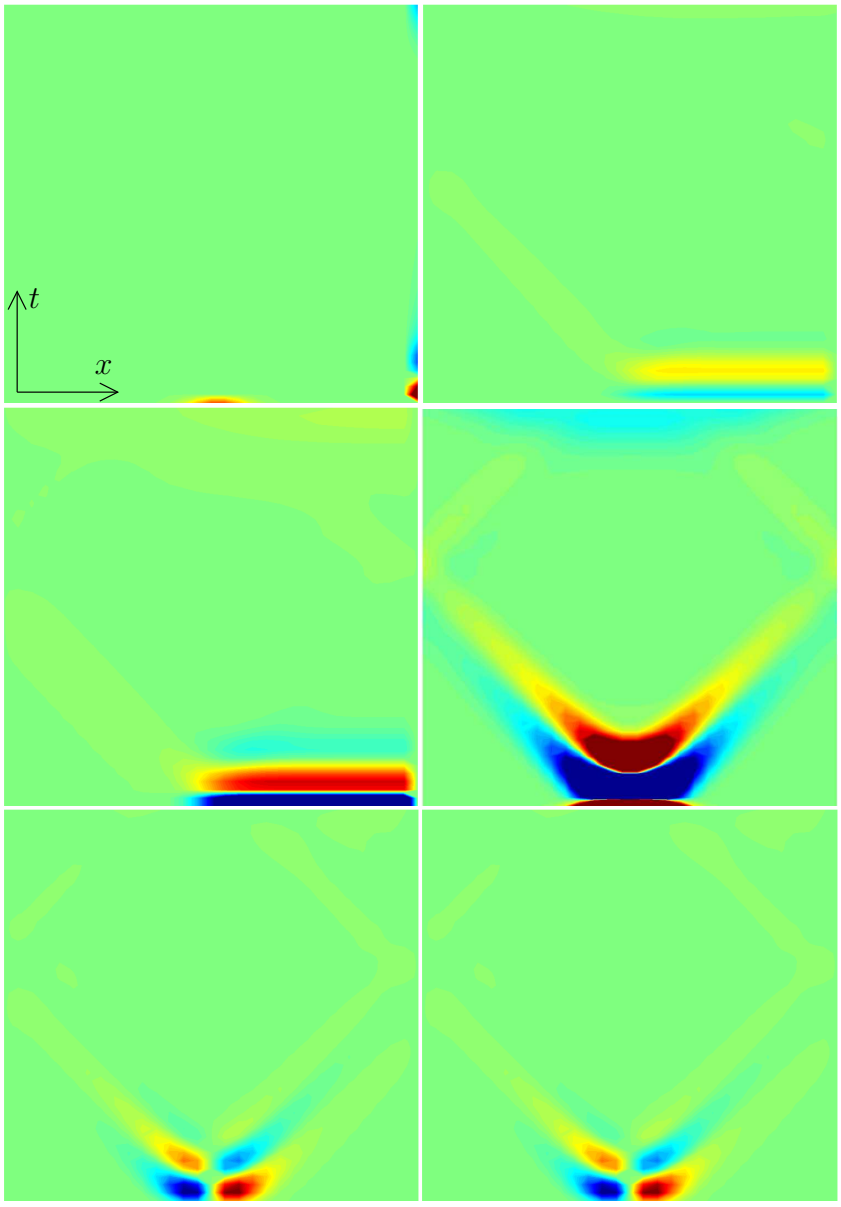


Figure 6.21: Figures 6.18, 6.19, and 6.20 continued, now with an xz-tree. Above left: the x-component of the electromagnetic potential A. Above right: A_z. Middle left: A_t. Middle right: the x-component of the electric field $\mathcal{E}$. Bottom left: $\mathcal{E}_y$. Bottom right: $\mathcal{E}_z$. A_y is identical to A_x.

Figures 6.18, 6.19, 6.20, and 6.21 show the components of the potential and the electric field $\mathcal{E}$ solved with a zt-, tz-, tt-, and xz-tree, respectively.

In the zt-tree case in figure 6.18, the spherical wavefront travels outwards from the source. The figure shows the wavefront on a xt-plane passing through the source. In fact, there are two successive wavefronts in all the components of A and $\mathcal{E}$, and the consecutive wavefronts have opposite signs. With this choice of spanning trees, the potential is nonzero only in the regions of the wavefront.

Figure 6.19 shows the fields in the tz-tree case. Now the wavefronts are almost invisible in the plots of A's components. The plots are dominated by the potential in the spatial source region at all time instants; i.e., the potential remains nonzero near the source also at later times. However, the components of the electric field are identical to the zt-tree case. This demonstrates how the same electromagnetic field F can be achieved with different potentials, which are solved with a different gauge condition. The situation of the tt-tree case in figure 6.20 is identical to that of the tz-tree. Accordingly, the t-directional edge-tree causes a t-directional beam in the values of A.

The case of an xz-tree is shown in figure 6.21. We have now an x-directional beam for the nonzero values of A. Because the x-directional edge gauge forces A_x into zero value at the tree edges, A_x is zero except on those x-directional edges that belong to the cotree. With our choice of tree, only the most rightward edges have nonzero values; i.e., they do not belong to the tree. Once again, the electric field is identical to the above cases.

Example 6.11. There is a catch in placing a charge-current source for a four-dimensional electromagnetic problem. We cannot place just any charge-current source, because the charge-current J must obey the current continuity equation (5.11). Its finite version can be written as follows:

$$(J|\partial\tilde{\Sigma}) = 0, \quad \forall\tilde{\Sigma} \in \tilde{C}_4(\tilde{K}). \tag{6.66}$$

Now, what would happen, if the source J did not satisfy equation (6.66)? Let us give an example of it and choose a tt-tree as in figure 6.20 but change the source term. Instead of A_x, we enforce J_x as a three-dimensional Gaussian function on the initial surface. In fact, the source is placed on the dual volumes that are dual to the x-directional edges on Γ_{initial}. Figure 6.22 shows the results with this source.

The potential has a t-directional beam in the spatial source region, as in the tt-tree case in figure 6.20, but now the amplitude of the potential grows linearly as a function of time. This is shown at bottom right in figure 6.22. Similarly, also the electric field has a time-directional beam of nonzero values

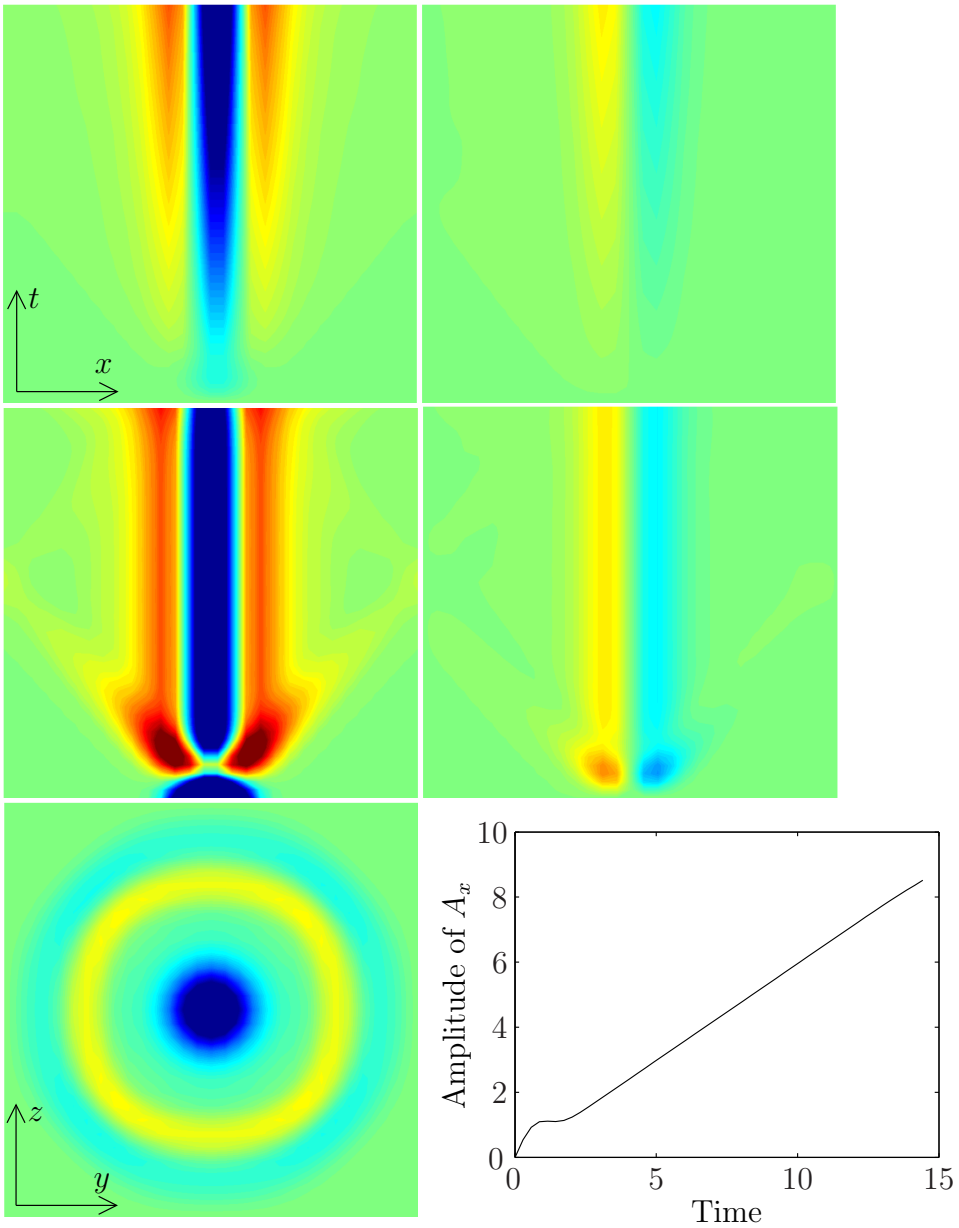


Figure 6.22: A four-dimensional example that is identical to the case in figure 6.20, except for the source term. The form of the source remains the same, but now we force a charge-current in the corresponding dual volumes. The x-component of J is given as the source term on the initial boundary: $\mathrm{t}_{\Gamma_{\text{initial}}} J_x$ is a three-dimensional Gaussian function in the center of the spatial domain with a width of several edge sizes. Above left: A_x. Above right: A_z. Middle left: $\mathcal{E}_x$. Middle right: $\mathcal{E}_y$. Bottom left: $\mathcal{E}_x$ on the yz-plane at time $t = 19/\sqrt{3}$. Bottom right: A_x as a function of time in the center of the spatial source region. Further, A_y is identical to A_x and $\mathcal{E}_z$ is identical to $\mathcal{E}_y$.

in the spatial source region. Thus this unphysical source generates an extra field in addition to the traveling sphere wavefront. In fact, this extra field equates with the electric field created by the static charge content, which the source current must leave in the source area—for the current continuity equation to be valid.

Example 6.12. Because the potentials with a zt-tree are most sphere-wave-like, we choose to use this tree in our final example. This example does not differ from the case in figure 6.18, except that its mesh is bigger and its source wider. Now, this mesh has 50 elements in spatial directions and 40 in the temporal direction.

Figure 6.23 shows the components of the potential A and electromagnetic field F at time instant $t = 30/\sqrt{3}$ on a yz-plane passing through the center of the source. The only nonzero component of A is A_x; likewise, all electric field components except $\mathcal{E}_x$ vanish. Because the potential has two sequential and oppositely signed wavefronts, the fields have three wavefronts with the middle one having the biggest amplitude.

Figure 6.24 shows the progress of the sphere wave as a function of time. The potential A_x is shown at several time instants on a yz-plane through the source. The sphere wave travels outwards from the source and begins to reflect from the boundary of a homogenous boundary condition for F. Figure 6.25 shows the progression on an xt-plane through the center of the source. As the sphere wave travels, the potential's amplitude thins out inverse to the radius from the source.

The above plots show the fields on yz- and xt-planes. Some components vanish on those planes, even though they gain nonzero values elsewhere. Consequently, figure 6.26 shows the fields on an xy-plane passing through the source.

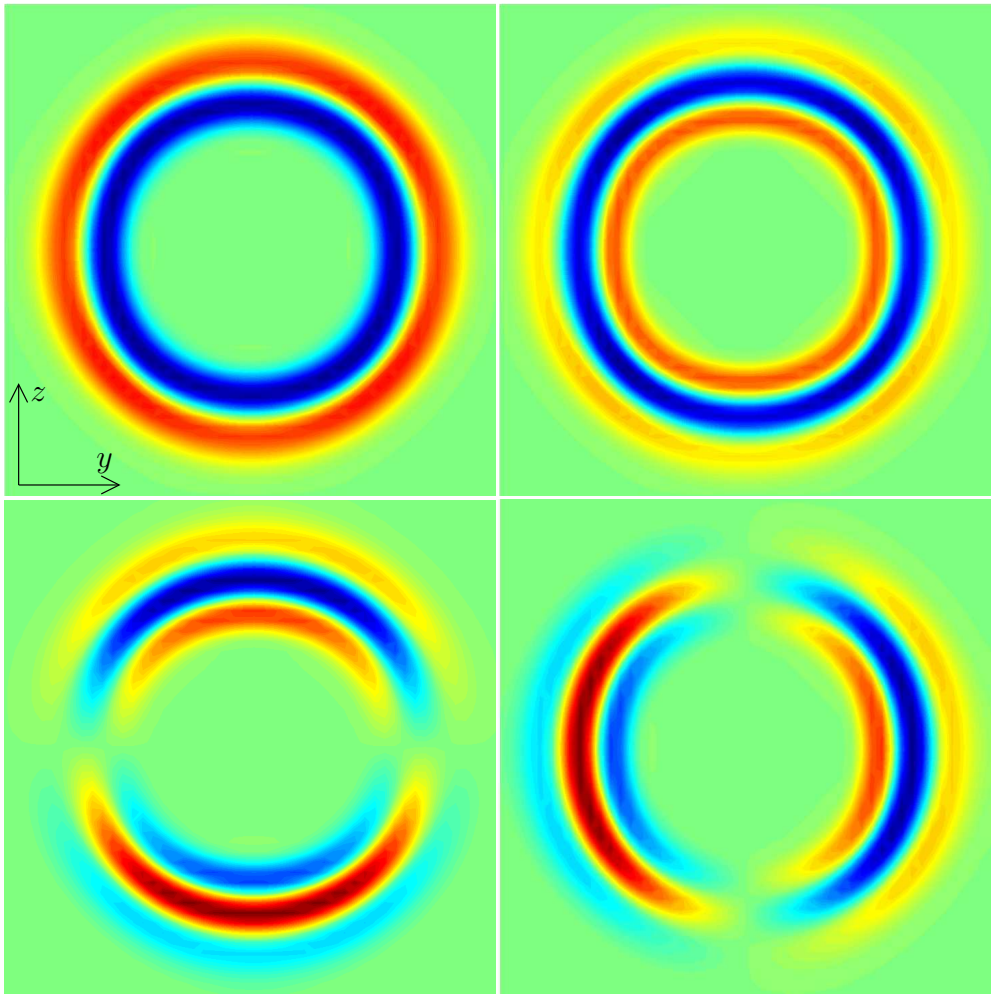


Figure 6.23: A four-dimensional example with a zt-tree; the potential and the fields are shown at the time instant $t = 30/\sqrt{3}$, on a yz-plane passing through the center of the source. The x-component A_x of A is given as the initial condition, as in example 6.10. The mesh has 50 3-cells in all the spatial dimensions and 40 in the temporal direction. Above left: the x-component of the electromagnetic potential A. Above right: the x-component of the electric field $\mathcal{E}$. Bottom left: the y-component of the magnetic flux $\mathcal{B}$. Bottom right: the z-component of the magnetic flux $\mathcal{B}$. All the other field components of A and F vanish.

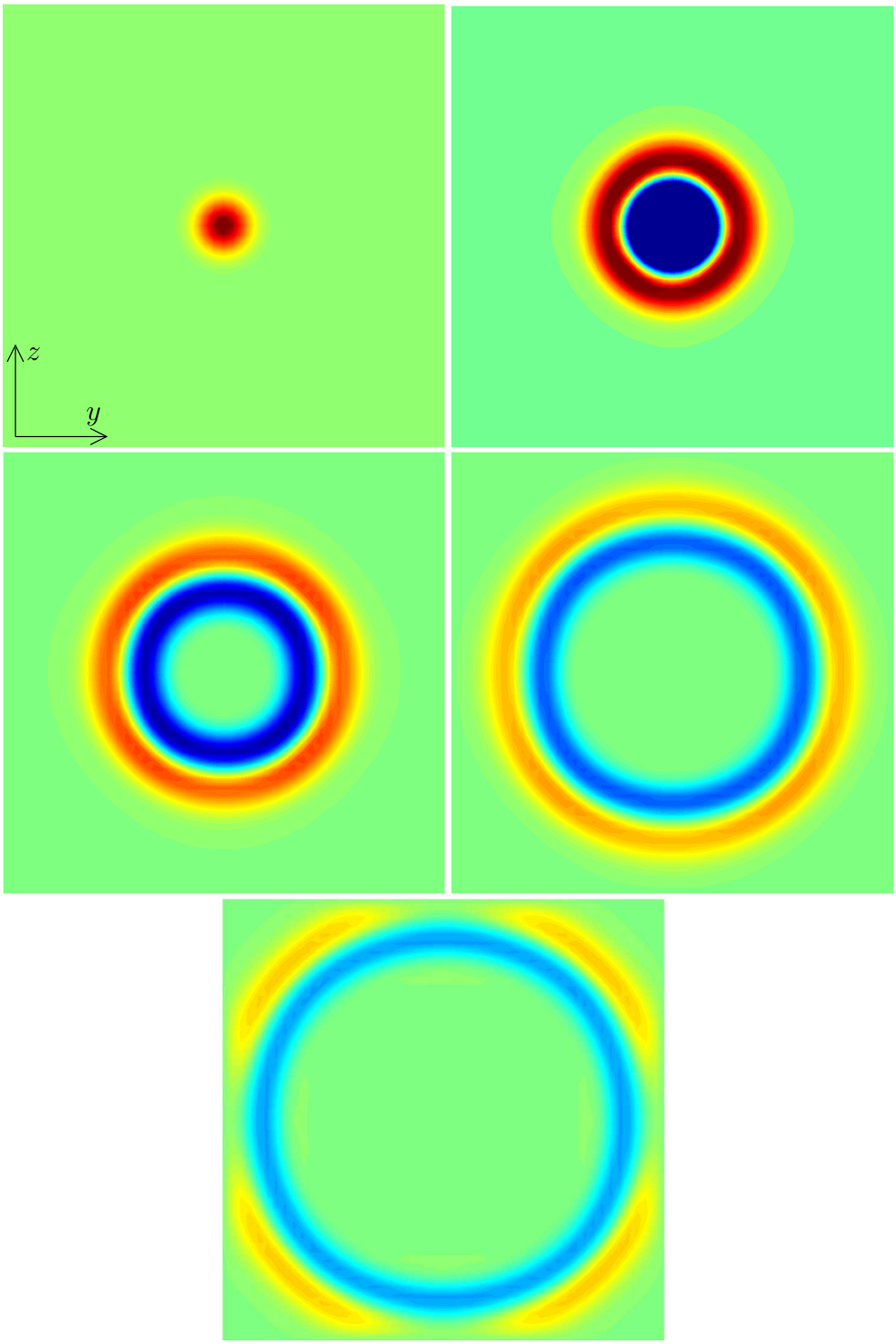


Figure 6.24: Figure 6.23 continued. The potential A_x is shown on the same yz-plane at times $t = 0$, $t = 10/\sqrt{3}$, $t = 20/\sqrt{3}$, $t = 30/\sqrt{3}$, and $t = 40/\sqrt{3}$. The case of $t = 0$ is not on the same color scale as the others.

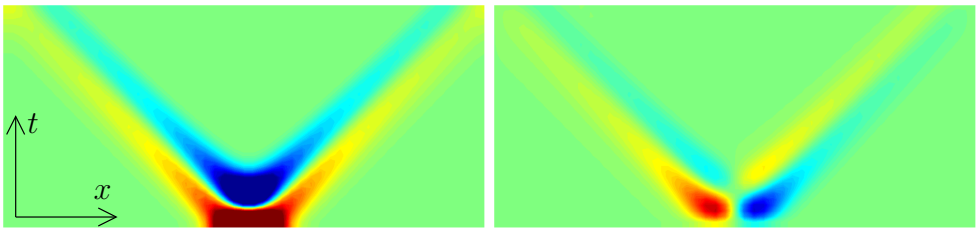


Figure 6.25: Figures 6.23 and 6.24 continued. Shown are A_x (left) and A_t (right) on the xt-plane, which passes through the source in the y- and z-directions.

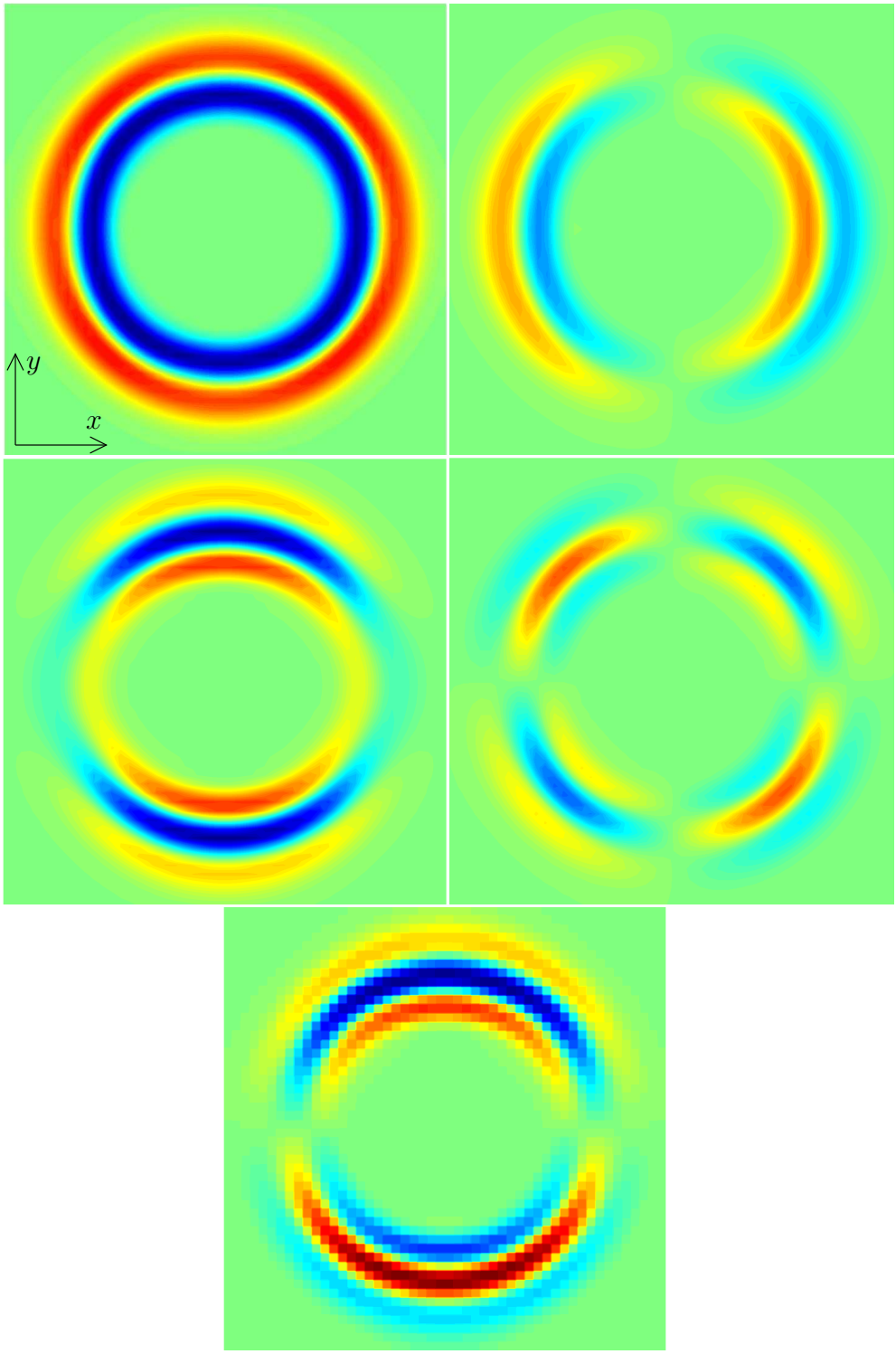


Figure 6.26: Figures 6.23, 6.24, and 6.25 continued. The fields are illustrated at $t = 30/\sqrt{3}$ on an xy-plane passing through the source. Above left: A_x. Above right: A_t. Middle left: $\mathcal{E}_x$. Middle right: $\mathcal{E}_y$. Bottom: $\mathcal{B}_z$. All the other components of A and F are fractional or zero.

Chapter 7

Conclusion

Let us summarize the results of this thesis. One of our main goals was to develop a method for solving our challenge problems introduced in section 2.3. In the first section, we discuss their solvability by our spacetime method.

In the second section, we discuss the main points of this thesis and briefly consider the significance of our findings—without forgetting the problematic cases.

7.1 Challenge problems

In section 2.3, we gave ourselves five challenge problems. The problems sought to motive the development of a geometric spacetime method. Though the thesis did not have enough scope to solve them all, we aimed to build a method that could be used to solve such problems. Let us discuss how this work succeeded with each problem.

Lightcone mesh

The lightcone mesh challenge problem was the most straightforward. The idea was to mesh only the spacetime region around the lightcone of a point source and thus make the method numerically more efficient. In example 6.6, the problem was solved in the two-dimensional spacetime case without any problems. The three-dimensional spacetime case proved to be much harder. The problem was only partially solved in example 6.9, and we were unable to avoid meshing the lightcone's interior; however, the exterior mesh could be cut out.

The four-dimensional case should not be any harder than the above two. Because four-dimensional sphere waves have no tail, we will not encounter the same problems as in the three-dimensional case. However, to deal with

gauging, we have to form two spanning trees in the four-dimensional lightcone mesh. Anyhow, this should not be a problem.

To sum up, in our method, lightcone meshes can be used in all dimensions, and their implementation is quite straightforward. However, with practical electromagnetic wave problems, lightcone meshes are not much help. Consequently, they only demonstrate the possibilities of our spacetime computation method.

Local time-stepping and the locally dense mesh

Our numerical examples were not particularly diverse, for we used only the disparity of Lorentzian orthogonality in example 6.5, which utilizes a pair of rhomboidal meshes. Lorentzian geometry gives us more freedom with the element shapes than we had used here.

As discussed in section 2.3, the forms of possible mesh elements in three-dimensional time-stepping methods do not allow locally dense meshes or local time-stepping without additional approximations. For example, see [80] for a subgridding method with a local time-stepping scheme, with which we needed both spatial and temporal interpolation on the boundary between meshes of different density. With our spacetime method, these interpolations could be avoided. However, our method requires a truly four-dimensional mesh, as opposed to a three-dimensional mesh extruded in the time direction. By exploiting Lorentzian orthogonality, we can build a pair of Lorentzian orthogonal meshes, in which the mesh density varies smoothly from point to point, both spatially and temporally. Figure 7.1 shows a possible two-dimensional primal mesh, which is locally condensed—especially in the time direction. It should be possible to construct an orthogonal dual mesh for this primal mesh, and thus obtain a pair of orthogonal meshes for our spacetime method. The time-step is shorter in the center than elsewhere, and the mesh size varies according to the time-step. The same idea should also generalize to four-dimensional meshes.

The problem lies in the use of truly four-dimensional meshes, which require vigorous modifications to current geometric time-stepping methods. However, the solution process with the above mesh can be interpreted as time-stepping. Thus we could use standard three-dimensional elements in most of the domain and truly four-dimensional elements only in the layer between standard elements. Unfortunately, for reasons mentioned in the Preface, we had to omit the real implementation of the method from this thesis.

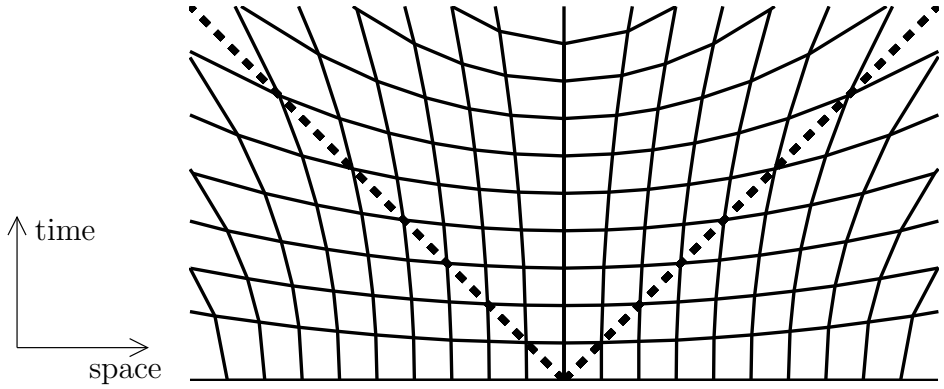


Figure 7.1: Primal mesh for local time-stepping in two-dimensional spacetime; the horizontal axis is the single spatial dimension, the vertical axis is time. In the center of the spatial domain, the time-step is shorter than elsewhere. The spatial mesh size varies according to the time-step so that the adjacent edges are Lorentzian orthogonal. The mesh is truly two-dimensional; i.e., the facets cannot be constructed by extrusion of a spatial edge along the time direction.

Remaining challenge problems

Other challenge problems were not further illuminated in this work. A moving geometry commonly requires a truly four-dimensional mesh. Such a problem can be solved with our spacetime method, but its interpretation as a three-dimensional time-stepping problem is not so straightforward. However, if we construct a standard three-dimensional mesh for the initial situation and distort it at the following time-instants, we could achieve a time-stepping method for moving geometries.

There is no cause to believe that our spacetime algorithm makes parallelization elementary. The solving process of our single system of linear equations resembles so closely the time-stepping scheme that a breakthrough in this field is not imminent with our new ideas. However, use of standard parallelization packages for linear algebra—instead of custom parallelization of the time-stepping scheme—could be advantageous in some cases.

With our spacetime method, managing gravitational effects in an electromagnetic wave problem is as easy as using curved manifolds in lower-dimensional problems. Thus the method qualifies directly for curved spacetime problems.

7.2 Summary

In this study, our aim was to develop a method for computing electromagnetic wave problems in spacetime. In the process, we developed and discovered many little-known mathematical and computational structures for modeling. In the end, those new findings turned out to be as important as the original objective.

Because this work involves two very distinct fields of the natural sciences, electromagnetic numerical computation and the theory of relativity, both had to be given a brief background (see chapters 2 and 3). However, since this thesis is mainly for electrical engineers, we covered relativity more extensively. The theory of relativity itself was not important for this work, except that it provided a physical background to our spacetime model.

Mathematics

We gave a mathematical model for spacetime, which was enough for relativity and for our needs. In the spacetime manifold model, we kept topology and the indefinite metric strictly separate. In addition, we demonstrated the curvature of spacetime with the geodesic, which generalizes straight lines to a curved manifold. Moreover, the observer-structure we introduced was quite unique, because it combines a multitude of physical observers, i.e., observers that the physicists use, into one concept (see section 4.2).

Our main modeling structure for electromagnetic fields was the cochain. Chain and cochain complexes are rarely used in semi-Riemannian manifolds. Because their metric-dependent features are normally related to the Riemannian manifold, we could not use them here, but instead made a great effort to separate topological and metrical structures. For example, chain norms were defined based on the canonical measure of the manifold, which was uniquely determined by the indefinite manifold metric. Because the topology inherited from chain norms was unsound for us, we struggled to develop proper chain topologies not connected with the metric. One of the truly unique accomplishments in this thesis was the use of uniform structures for completing the chain space without a metric (section 4.3.4).

In contrast, we used a quite standard differential form structure, because it did not necessitate many metric-dependent structures. A major shortcoming in this work was the connection between flat cochains and flat differential forms. Here we had to resort to hypothesizing that such a one-to-one connection exists (see section 4.6.2).

Electromagnetics

In electromagnetic theory, we emphasized the connection between basic measurements and modeling structures—directly in four-dimensional spacetime. Additionally, we linked the four-dimensional electromagnetic concepts to the well-known three-dimensional ones by using the observer-structure to separate observer-independent four-dimensional fields into observer-dependent three-dimensional fields. For constitutive laws, we employed differential forms and thus called for a connection between the forms and cochains. Further, four-dimensional boundary conditions were divided into spatial boundary conditions and initial conditions.

The main objective of this work, however, was to develop a formulation for computational electromagnetics in four-dimensional spacetime. Instead of a leap-frog-scheme, we used a method whereby the electromagnetic potential was solved through one big linear equation. The method resembles geometric methods, i.e., FIT and the cell method, except that the whole time-stepping scheme is combined into one large linear equation. The second unique feature of our method was the use of (four-dimensional vector) potentials for electromagnetic wave computation. In the two- and three-dimensional spacetime cases, this method differed from the equal-dimensional static problem in terms of the finite constitutive law and boundary conditions. The constitutive laws of the non-spacetime geometric methods had to be modified to accommodate them to the indefinite metric.

In four-dimensional spacetime, our algorithm is unmatched. First, only a few methods can perform electromagnetic computation in a four-dimensional mesh, and they do not use a geometric method, except the work of Stern et. al. [68]. Second, our four-dimensional geometric method embraces many details that exist only in four dimensions. For example, in sections 6.1.3 and 6.1.4, we saw that the boundary conditions dropped different equations and quantities in the solution phase than in non-spacetime methods. However, the most unmatched feature in our four-dimensional method is the spanning-tree gauging method. In contrast to lower-dimensional methods, our method requires two separate spanning trees, one for restricting the quantity space and one for restricting the equations (see sections 6.3.1 and 6.3.2). As we saw in section 6.3, the latter was connected with dropping the Maxwell source equations in the three-dimensional geometric methods. However, we concluded that this corresponds only to one specific choice of the second spanning tree in the four-dimensional mesh.

To complete the work, we demonstrated the method with numerical examples in two-, three-, and four-dimensional domains. The progression of the examples also illustrates the development of the method. Some basic

properties of spacetime computation are similar in all dimensions; those are the easiest to (find and) demonstrate in a two-dimensional case. The third dimension brought only a minor extra challenge to the method. The fourth dimension, however, added many unique features to the computation.

Multiple black spots occurred during this work, many of them were dealt with only briefly in a small remark in this thesis. Example 6.11 represents the biggest source of complications in this work. Problematic was also the suitability of chain and cochain norms for electromagnetic modeling, as discussed in section 4.3.4.

Final words

In conclusion, we developed and demonstrated a geometric method in spacetime for electromagnetic wave problems. Despite a few minor flaws in the mathematical model, the method proved viable. Additionally, the method enabled some improvements in current geometric methods.

Bibliography

[1] R. Albanese and G. Rubinacci. Integral formulation for 3D eddy-current computation using edge elements. *IEE Proc.*, 135, Pt. A(7):457–462, 1988.

[2] P. Alotto, A. De Cian, and G. Molinari. A time-domain 3-d full-Maxwell solver based on the cell method. *IEEE Trans. Magn.*, 42:799–802, 2006.

[3] T. Asselmeyer-Maluga and C. Brans. *Exotic Smoothness in Physics*. World Scientific, Singapore, 2007.

[4] J. Beem and P. Ehrlich. *Global Lorentzian Geometry*. Marcel Dekker, New York, 1981.

[5] E. Binz, J. Śniatycki, and H. Fischer. *The Geometry of Classical Fields*. North Holland, Amsterdam, 1988.

[6] A. Bossavit. *Computational Electromagnetism, Variational Formulations, Edge Elements, Complementarity*. Academic Press, Boston, 1998.

[7] A. Bossavit. Computational electromagnetism and geometry: Building a finite-dimensional "Maxwell's house". (1): Network equations. *J. Japan Soc. Appl. Electromagn. & Mech.*, 7(2):150–159, 1999.

[8] A. Bossavit. Computational electromagnetism and geometry. (5): The 'Galerkin hodge'. *J. Japan Soc. Appl. Electromagn. & Mech.*, 8(2):203–209, 2000.

[9] A. Bossavit. The Sommerville mesh in Yee-like schemes. In *Scientific Computing in Electrical Engineering*, Eindhoven, The Netherlands, 2002.

[10] A. Bossavit. Discretization of electromagnetic problems: The "generalized finite differences" approach. In W.H.A. Schilders and E.J.W. ter Maten, editors, *Numerical Methods in Electromagnetics*, volume 13 of *Handbook of Numerical Analysis*, pages 105–197. Elsevier, 2005.

[11] A. Bossavit and L. Kettunen. Yee-like schemes on a tetrahedral mesh with diagonal lumping. *Int. J. Numer. Modell.*, 12:129–142, 1999.

[12] A. Bossavit and J. C. Vérité. A mixed FEM-BIEM method to solve 3-d eddy current problem. *IEEE Trans. Magn.*, 18(2):431–435, 1982.

[13] S. Carroll. Lecture notes on general relativity. *arXiv.org*, arXiv:gr-qc/9712019v1, December 1997.

[14] M. Cinalli, F. Edelvik, R. Schuhmann, and T. Weiland. Consistent material operators for tetrahedral grids based on geometrical principles. *Int. J. Num. Modell.: Elect. Netw., Dev. and Fields*, 17(5):487–507, 2004.

[15] F. Collino, T. Fouquet, and P. Joly. A conservative space-time mesh refinement method for the 1-d wave equation. Part II: Analysis. *Numerische Mathematik*, 95:223–251, 2003.

[16] H. S. M. Coxeter. *Regular Polytopes*. Dover, 3rd edition, 1973.

[17] Z. Csendes and P. Silvester. Numerical solution of dielectric loaded waveguides: I-Finite-Element analysis. *IEEE Trans. on Microwave Theory and Techniques*, 18(12):1124 – 1131, 1970.

[18] K. Duggal and A. Bejancu. *Lightlike Submanifolds of Semi-Riemannian Manifolds and Applications*. Kluwer Academic Publishers, Dordrecht, 1996.

[19] J. Erickson, D. Guoy, M. Sullivan, and A. Üngör. Building spacetime meshes over arbitrary spatial domains. *Engineering with Computers*, 20(4):342–353, 2005.

[20] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81(8):1562–1567, Aug 1998.

[21] T. Euler. *Consistent Discretization of Maxwell's Equations on Polyhedral Grids.* Ph.D. thesis, TU Darmstadt, 2007.

[22] H. Flanders. *Differential Forms with Applications to the Physical Sciences*. Dover Publications, 1989.

[23] G. Galilei. *Dialogo sopra i due massimi sistemi del mondo*. Florenze, 1632.

[24] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer, Berlin-Heidelberg, 2004.

[25] G. Galloway. Spacetime geometry. Beijing International Mathematics Research Center, 2007 Summer School, 2007. [Online; accessed 16-May-2008].

[26] R. Geroch. *Mathematical Physics*. The University of Chicago Press, Chicago, 1985.

[27] W. Gordon. Zur lichtfortpflanzung nach der relativitätstheorie. *Ann. Phys. (Leipzig)*, 72:421–456, 1923.

[28] E. Gourgoulhon and J.L. Jaramillo. A 3+1 perspective on null hypersurfaces and isolated horizons. *Phys. Rep.*, 423:159–294, 2006.

[29] J. Harrison. Lectures on chainlet geometry - new topological methods in geometric measure theory. *arXiv.org*, arXiv:math-ph/0505063v1 [math.CA], May 2005.

[30] A. Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002.

[31] F.W. Hehl and Yu.N. Obukhov. *Foundations of Classical Electrodynamics: Charge, Flux, and Metric*. Birkhäuser, Boston, MA, 2003.

[32] J.S. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer-Verlag, New York, 2008.

[33] P. J. Hilton and S. Wylie. *Homology Theory, An Introduction to Algebraic Topology*. Cambridge University Press, Cambridge, 1965.

[34] R. Hiptmair. Discrete Hodge-operators: An algebraic perspective. *PIER, Geometric Methods for Computational Electromagnetics*, 32:247–269, 2001.

[35] K. Huang. *Quarks, leptons & gauge fields*. World Scientific, Singapore, 2nd edition, 1992.

[36] K. Itô, editor. *Encyclopedic Dictionary of Mathematics*, volume 1. MIT Press, 2nd edition, 1996.

[37] V. Ivancevic and T. Ivancevic. *Applied differential geometry: a modern introduction*. World Scientific, Singapore, 2007.

[38] K. Jänich. *Vector analysis*. Springer-Verlag, Berlin, 2001.

[39] H. B. Lawson Jr. Foliations. *Bull. Amer. Math. Soc.*, 80(3):369–418, 1974.

[40] J. Kangas. *Algebraic Decompositions in Analysis of Quasistatic Electromagnetic Problems*. Ph.D. thesis, Tampere University of Technology, 2008.

[41] J. Keränen, E. Koljonen, T. Tarhasaari, and L. Kettunen. Effect of cell type on convergence of wave propagation schemes. *IEEE Trans. Magn.*, 40(2):1452–1455, 2004.

[42] B. Krietenstein, R. Schuhmann, P. Thoma, and T. Weiland. The perfect boundary approximation technique facing the big challenge of high precision field computation. In *Proc. 19th LINAC Conf.*, pages 860–862, Chicago, IL, August 1998.

[43] D.N. Kupeli. On null submanifolds in spacetimes. *Geometriae Dedicata 23*, pages 33–51, 1987.

[44] S. Kurz. Relativistic electrodynamics - modeling and simulation. Lecture Series at ETH Zürich, Seminar for Applied Mathematics, 2009.

[45] J. Lee. *Riemannian manifolds: an introduction to curvature*. Springer-Verlag, New York, 1997.

[46] M. Marrone. Computational aspects of the cell method in electrodynamics. *PIER, Geometric Methods for Computational Electromagnetics*, 32:317–356, 2001.

[47] C. Mattiussi. An analysis of finite volume, finite element, and finite difference methods using some concepts from algebraic topology. *Journal of Comp. Physics*, 133:289–309, 1997.

[48] C. Mattiussi. The finite volume, finite element, and finite difference methods as numerical methods for physical field problems. *Advances in Imaging and Electron Physics*, 113:1–146, 2000.

[49] D. B. Melrose and R. C. McPhedran. *Electromagnetic Processes in Dispersive Media: A Treatment Based on the Dielectric Tensor*. Cambridge University Press, Cambridge, 1991.

[50] H. Minkowski. Raum und Zeit. *Physikalische Zeitschrift*, 10:104–111, 1909. Lecture at the 80th Meeting of German scientists and physicians, Cologne 21.9.1908.

[51] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman and Company, New York, 1973.

[52] J. C. Nédelec. Mixed finite elements in R^3. *Numer. Math.*, 35:315–341, 1980.

[53] B. O'Neill. *Semi-Riemannian geometry*. Academic Press, San Diego, CA, 1983.

[54] O. Podebrad, M. Clemens, and T. Weiland. New flexible subgridding scheme for the finite integration technique. *IEEE Trans. Magn.*, 39(3):1662–1665, 2003.

[55] P. Raumonen. *Mathematical Structure for Dimensional Reduction and Equivalence Classification of Electromagnetic Boundary Value Problems*. Ph.D. thesis, Tampere University of Technology, 2009.

[56] J. Reitz, F. Milford, and R. Christy. *Foundation of Electromagnetic theory*. Addison-Wesley, 4 edition, 1993.

[57] G. Richter. An explicit finite element method for the wave equation. *Applied Numerical Mathematics*, 16(1-2):65 – 80, 1994.

[58] A. Robb. *Optical geometry of motion, a new view of the theory of relativity*. Cambridge University Press, Cambridge, 1911.

[59] T. Rylander and A. Bondeson. Stable FEM-FDTD hybrid method for Maxwell's equations. *Comput. Phys. Comm.*, 125(1-3):75–82, 2000.

[60] T. Rylander and A. Bondeson. Stability of explicit-implicit hybrid time-stepping schemes for Maxwell's equations. *J. of Comput. Phys.*, 179(2):426 – 438, 2002.

[61] H. Schaefer. *Topological vector spaces*. MacMillan, New York, 1966.

[62] V. Schlegel. Theorie der homogen zusammengesetzten Raumgebilde. *Verhandlungen der Kaiserlichen Leopoldinisch-Carolinischen Deutschen Akademie der Naturforscher*, 4, 1883.

[63] R. Schuhmann and T. Weiland. A stable interpolation technique for FDTD on nonorthogonal grids. *Int. J. Num. Modell.: Elect. Netw., Dev. and Fields*, 11(6):299–306, 1998.

[64] G. Shephard. Convex polytopes with convex nets. *Math. Proc. Camb. Phil. Soc.*, 78:389–403, 1975.

[65] L. Silberstein. *The theory of relativity.* Macmillan and co., London, 1914.

[66] P. Silvester. Finite-element solution of homogeneous waveguide problems. *Alta Frequenza*, 38:313–317, 1969. Special issue, conference record of the URSI Symposium on Electromagnetic Waves, Stresa, Italy, June 1968.

[67] H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers, and E. Herlt. *Exact Solutions of Einstein's Field Equations.* Cambridge University Press, Cambridge, 2003.

[68] A. Stern, Y. Tong, M. Desbrun, and J. Marsden. Geometric computational electrodynamics with variational integrators and discrete differential forms. *arXiv.org*, arXiv:0707.4470v3 [math.NA], July 2007.

[69] S. Suuriniemi. *Homological Computations in Electromagnetic Modeling.* Ph.D. thesis, Tampere University of Technology, 2004.

[70] J. L. Synge. *Relativity: The Special Theory.* North-Holland, Amsterdam, 1958.

[71] A. Taflove. Application of the Finite-Difference Time-Domain Method to Sinusoidal Steady-State Electromagnetic-Penetration Problems. *IEEE Trans. on Electromagnetic Compatibility*, 22(3):191–202, 1980.

[72] A. Taflove. *Computational Electromagnetics: The Finite-Difference Time-Domain Method.* Artech House, Boston, 1995.

[73] T. Tarhasaari. *Mathematical Structure of Cochain in Computational Electromagnetics.* Ph.D. thesis, Tampere University of Technology, 2002.

[74] T. Tarhasaari and L. Kettunen. Wave propagation and cochain formulations. *IEEE Trans. Magn.*, 39(3):1195–1198, 2003.

[75] T. Tarhasaari, L. Kettunen, and A. Bossavit. Some realizations of a discrete Hodge operator: A reinterpretation of finite element techniques. *IEEE Trans. Magn.*, 35(3), 1998.

[76] C. Taubes. Gauge theory on asymptotically periodic 4-manifolds. *J. Differential Geom.*, 25:363–430, 1987.

[77] E. Taylor and J. Wheeler. *Spacetime Physics*. W.H. Freeman and Company, New York, 2nd edition, 1992.

[78] S. V. Thite. *Spacetime Meshing for Discontinuous Galerkin Methods*. PhD thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 2005.

[79] P. Thoma. *Zur numerischen Lösung der Maxwellschen Gleichungen im Zeitbereich*. Ph.D. thesis, TU Darmstadt, 1997.

[80] P. Thoma and T. Weiland. A consistent subgridding scheme for the finite difference time domain method. *Int. J. Num. Modell.: Elect. Netw., Dev. and Fields*, 9:359–374, 1996.

[81] E. Tonti. A direct discrete formulation of field laws: The cell method. *Comput. Model. Eng. Sci.*, 2(2):237–258, 2001.

[82] E. Tonti. Finite formulation of the electromagnetic field. *PIER, Geometric Methods for Computational Electromagnetics*, 32:1–44, 2001.

[83] E. Tonti. Finite formulation of electromagnetic field. *IEEE Trans. Magn.*, 38(2):333–336, March 2002.

[84] J.S. van Welij. Calculation of eddy currents in terms of H on hexahedra. *IEEE Trans. Magn.*, 21(6):2239–2241, November 1985.

[85] J. Vanderlinde. *Classical electromagnetic theory*. Springer, Dordrecht, The Netherlands, 2004.

[86] F. W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer-Verlag, Berlin, 1983.

[87] T. Weiland. A discretization method for the solution of Maxwell's equations for six-component fields. *Int. J. Electron. Commun. (AEÜ)*, 31(3):116–120, 1977.

[88] T. Weiland. Time domain electromagnetic field computation with finite difference methods. *Int. J. Num. Modell.*, 9:295–319, 1996.

[89] H. Weyl. *Space, Time, Matter*. Dover Publications, New York, 1952.

[90] J. Wheeler. *A Journey into Gravity and Spacetime*. Scientific American Library, San Francisco, 1990.

[91] H. Whitney. *Geometric Integration Theory*. Princeton University Press, 1957.

[92] K. S. Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans. Antennas Propagat.*, 14:302–307, May 1966.

[93] K. Yosida. *Functional analysis*. Springer-Verlag, Berlin, 1980.

[94] W. Yu and R. Mittra. A conformal finite difference time domain technique for modeling curved dielectric surfaces. *IEEE Microwave and Wireless Comp. Lett.*, 11:25–27, Jan 2001.

[95] I. Zagorodnov, R. Schuhmann, and T. Weiland. A uniformly stable conformal FDTD-method in Cartesian grids. *Int. J. Num. Modell.: Elect. Netw., Dev. and Fields*, 16:127–141, 2003.

[96] I. Zagorodnov, R. Schuhmann, and T. Weiland. Conformal FDTD-methods to avoid time step reduction with and without cell enlargement. *Journal of Comp. Physics*, 225(2):1493–1507, 2007.

Index