



**GHENT  
UNIVERSITY**

**ADVANCED NUMERICAL METHODS FOR INVERSE PROBLEMS  
IN EVOLUTIONARY PDES**

**MICHAL GALBA**

*SUPERVISOR: MARIÁN SLODIČKA*

Thesis submitted to Ghent University in candidature for the academic degree  
of Doctor of Philosophy in Mathematics

Ghent University  
Faculty of Engineering and Architecture  
Department of Mathematical Analysis  
Research Group for Numerical Analysis  
and Mathematical Modelling (*NaM<sup>2</sup>*)

Academic year  
2016-2017



*Mathematics knows no races  
or geographic boundaries;  
for mathematics, the cultural  
world is one country.*

DAVID HILBERT



# Acknowledgements

This dissertation is the result of my research activities that I have concluded within the past four years under the supervision of Professor Marián Slodička, to whom I would like to express my immense gratitude. He has always shared his knowledge with me and offered his helping hand even if the problems I struggled with were of non-mathematical nature. At the same time, I would like to thank my previous teachers, especially to Professor Jozef Kačur and Mgr. Dagmar Krajčiová, whose passion for numbers taught me to love the beauty of mathematics. My research has been supported by Belgian Science Policy project IAP P7/02. I intend to express my appreciation for the essential financial support.

I am also very grateful for the exceptional people I have met. Among them Jaroslav Chovan and Katarína Šišková, who have always been the best friends and colleagues, Karel Van Bockstal who thoroughly read my dissertation and friendly suggested some modifications, and Vladimír Vrábek who helped me to find a way when I was lost, especially during my first year at Ghent University.

I am obliged to my family for all that they have done for me. And last but not least, my sincere gratitude belongs to my beloved wife Janka, whose never-ending love and support made me strong enough to overcome all the obstacles I have faced.

Michal



# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Summary</b>	<b>xiii</b>
<b>Samenvatting</b>	<b>xv</b>
<b>1 MATHEMATICAL BACKGROUND</b>	<b>1</b>
1.1 Partial Differential Equations . . . . .	1
1.1.1 History of PDEs . . . . .	1
1.1.2 Formal definition of PDEs . . . . .	2
1.1.3 Initial and boundary conditions . . . . .	2
1.1.4 Methods for finding a solution to a PDE . . . . .	4
1.1.5 Strong and weak solutions . . . . .	5
1.1.6 Applications of PDEs in practice . . . . .	7
1.1.7 Well-posed and Ill-posed problems . . . . .	9
1.2 Inverse problems . . . . .	10
1.3 Rothe's method . . . . .	14
1.4 Finite element method . . . . .	15

---

1.4.1	Galerkin method . . . . .	15
1.4.2	Finite element method . . . . .	18
1.5	Maxwell's equations . . . . .	23
1.5.1	Description of quantities and notations . . . . .	24
1.5.2	Establishing equations for linear media . . . . .	29
1.5.3	Boundary conditions . . . . .	31
<b>Part I</b>		<b>35</b>
<b>2</b>	<b>RECOVERY OF A TIME-DEPENDENT SOURCE</b>	<b>35</b>
2.1	Problem formulation . . . . .	35
2.2	Time discretization . . . . .	38
2.3	A priori estimates . . . . .	39
2.4	Existence of a solution . . . . .	46
2.5	Uniqueness of a solution . . . . .	51
2.6	Error estimates . . . . .	53
2.7	Numerical Computations . . . . .	58
2.7.1	First Experiment . . . . .	59
2.7.2	Second Experiment . . . . .	61
<b>3</b>	<b>DETERMINATION OF A TIME-DEPENDENT CONVOLUTION KERNEL</b>	<b>65</b>
3.1	Problem formulation . . . . .	65
3.2	Time discretization . . . . .	70
3.3	A priori estimates . . . . .	72
3.4	Existence of a solution . . . . .	79
3.5	Uniqueness . . . . .	84

---



---

3.6	Numerical computations . . . . .	90
<b>Part II</b>		<b>97</b>
<b>4</b>	<b>MODEL ORDER REDUCTION</b>	<b>97</b>
4.1	Model Order Reduction . . . . .	99
4.2	Proper Orthogonal Decomposition . . . . .	100
4.2.1	POD basis construction . . . . .	100
4.2.2	Method of Snapshots . . . . .	103
4.3	Numerical Examples . . . . .	106
4.3.1	Identification of a space-dependent source from final time data . . . . .	107
4.3.2	Identification of a time-dependent source from a surface measurement . . . . .	112
<b>CONCLUSION AND FUTURE RESEARCH</b>		<b>121</b>
<b>Appendices</b>		<b>125</b>
<b>A</b>	<b>Differential operators and functions in n-dimensional spaces</b>	<b>125</b>
<b>B</b>	<b>Important inequalities</b>	<b>133</b>
<b>C</b>	<b>Functional analysis</b>	<b>137</b>

---



# List of Figures

2.1	Mesh in domain: The left hand side of the cube (colored black) is denoted $\gamma \subset \partial\Omega$ . Measurement $m(t)$ is performed only on this subset of the boundary. . . . .	59
2.2	Exact and numerical solutions for $\mathbf{E}$ . . . . .	60
2.3	Relative error $\frac{\mathbf{E} - \mathbf{E}_{app}}{ \mathbf{E} }$ depicted for $t = T$ . Its absolute value is denoted by color. . . . .	60
2.4	Reconstruction of the function $h(t)$ and its error . . . . .	61
2.5	Mesh in domain. The right-hand-side of the sphere (colored black) is denoted $\gamma \subset \partial\Omega$ . Measurement $m(t)$ is performed only on this subset of the boundary . . . . .	62
2.6	Exact and numerical solutions for $\mathbf{E}$ . . . . .	63
2.7	Relative error $\frac{\mathbf{E} - \mathbf{E}_{app}}{ \mathbf{E} }$ depicted for $t = T$ . Its absolute value is denoted by color. . . . .	63
2.8	Reconstruction of the function $h(t)$ and its error . . . . .	64
2.9	Reconstruction of the function $h(t)$ when the measurement $m(t)$ contain a certain level of noise . . . . .	64
3.1	Decay of error of $\mathbf{E}$ and $\alpha(t)$ depending on the size of time step $\tau$	92

---

3.2	Exact and Numerical solutions of $\mathbf{E}$ at time $t = T$ . . . . .	92
3.3	Relative error $\frac{\mathbf{E} - \mathbf{E}_{app}}{ \mathbf{E} }$ at $t = T$ . The absolute value of the error is indicated with color . . . . .	93
3.4	Reconstruction of function $\alpha(t) = t^3 + t + 1$ . For the sake of clarity only every $5^{th}$ time step is plotted. . . . .	93
3.5	Reconstruction of the function $\alpha(t)$ when the measurement $m(t)$ contain a certain level of noise . . . . .	94
4.1	Theory - Experiments - Simulations . . . . .	98
4.2	The concept of model order reduction . . . . .	99
4.3	Reconstruction of the function $h(t)$ by all three methods . . . . .	116
4.4	Reconstruction of the function $h(t)$ by the decoupling scheme . . . . .	117
4.5	Minimization technique requires doubling of the time interval . . . . .	119
4.6	Difference between the minimization technique combined with 1st and 2nd order finite elements . . . . .	120

---

# List of Tables

4.1	Comparison of a standard minimization technique with POD method	111
4.2	Comparison of a standard minimization technique, POD method, and decoupling scheme . . . . .	116



# Summary

This dissertation investigates advanced numerical methods that are to be used to solve problems that, when translated into the language of mathematics, involve partial differential equations (PDEs) and inverse problems (IPs).

Partial differential equations are very often the result of mathematical modelling, which is a process of describing the forces that govern the physical world in terms of numbers and mathematical symbols. The more precise and accurate description is required, the more complicated equations are gained. Since there is no global theory whose application would provide us with the solution of a general PDE, and it is very unlikely that such a theory will ever be discovered, the only way how to treat unsolved problems is to patiently and carefully keep seeking new methods that uncover step by step the solutions of a whole new class of problems.

Sometimes, and in practice very often, it can happen that a part of a model, e.g. a diffusivity coefficient, a time-dependent source or a convolution kernel, is unknown. In such a situation this parameter is to be identified based on the knowledge of some additional information. This extra information can be either provided by a measurement or derived from theory. This type of problem is usually labeled as an *inverse problem*.

The content of this dissertation is based on research activities which I concluded within the last four years, and the two articles [79, 80] that have been published in the respected journal *Computers & Mathematics with applications* with high impact factor 1.398; according to the Web of Science it occupies the 46th place in the list of 254 journals in the category of Applied mathematics.

This dissertation is organized as follows:

Chapter 1 proposes a general overview of mathematical tools, which are essential for coherent reading of the main parts of this work. In this chapter, the formal definition of PDEs is introduced as well as the concepts of initial and boundary conditions, strong and weak solutions, and inverse problems. The core part of this chapter is the Section 1.3 that explains Rothe's method as a tool for finding solutions to evolution problems. By means of Rothe's method the existence and uniqueness of the solution is proven too.

Chapter 2 is based on article [80] and in this chapter a time-dependent source in Maxwell's equations is recovered from a surface measurement. The main contribution of this part is, that the additional measurement is only taken along a 2D surface, which means that a non-invasive measurement is assumed. The existence of a solution together with its uniqueness is proven, error estimates are obtained, and the theoretical results are supported with numerical examples.

Chapter 3 deals with non-linear hyperbolic Maxwell's equations with missing time-dependent convolution kernel. The missing data are determined from a surface measurement, which is a new and a very extensive aspect. The uniqueness and existence of a solution is proven in this chapter. Furthermore, a numerical scheme for finding a solution based on backward Euler's method is proposed.

The last chapter, Chapter 4, introduces a new class of techniques called *Model order reduction*, which has been subject of extensive amount of studies within the past decades. We propose a discussion of the advantages and disadvantages of this method compared with techniques which are considered to be standard, e.g. Decoupling scheme or Minimization of a cost functional. This chapter is strongly linked to the two previous chapters, because it allows one to see the problems described in Chapters 2 and 3 from a new perspective.

---



# Samenvatting

Dit doctoraatsproefschrift onderzoekt een aantal geavanceerde numerieke methoden voor wiskundige problemen waarin partiële differentiaalvergelijkingen (PDV) voorkomen en die als inverse problemen (IP) geformuleerd kunnen worden.

Partiële differentiaalvergelijkingen zijn vaak het resultaat van wiskundige modellering. Dit is een proces waarbij natuurkundige fenomenen beschreven worden met behulp van getallen en wiskundige symbolen. Hoe nauwkeuriger en realistischer het vereiste model moet zijn, hoe ingewikkelder de bekomen vergelijkingen gaan zijn.

Er bestaat geen algemene theorie die het bestaan van een oplossing van een PDV garandeert. Bovendien is het heel onwaarschijnlijk dat zulke theorie ooit zal ontdekt worden. Een goede manier om de onopgeloste problemen aan te pakken is door aandachtig en stap voor stap te blijven zoeken naar nieuwe methoden voor bepaalde klassen van problemen.

In de praktijk komt het vaak voor dat een parameter van het model onbekend is, bv. een diffusiecoëfficiënt, een bronterm, een convolutiekern, etc. In deze situatie is het de bedoeling om die parameter te identificeren door middel van bijkomende informatie. Deze extra informatie kan bekomen worden door middel van een bijkomende meting, maar kan soms ook afgeleid worden uit de theorie. Dit soort problemen krijgen meestal de benaming *inverse problemen*.

Dit doctoraatsproefschrift is gebaseerd op vier jaar lang doctoraatsonderzoek dat resulteerde in twee artikels [79, 80]. Deze artikels zijn beide gepubliceerd in het gerespecteerde tijdschrift *Computers & Mathematics with applications* (impact

factor 1,398, 46e plaats in de lijst van 254 tijdschriften in de categorie Toegepaste wiskunde volgens de Web of Science).

Dit doctoraatsproefschrift bestaat uit vier hoofdstukken. Hoofdstuk 1 geeft een algemeen overzicht van de wiskundige instrumenten die essentieel zijn voor de analyse in het vervolg van het proefschrift. In dit hoofdstuk leggen we de basisbegrippen uit zoals de definitie van een PDV, van beginvoorwaarde en van randvoorwaarde, van sterke en zwakke oplossingen, en van inverse problemen. In de belangrijkste sectie, i.e. Sectie 1.3, wordt de Rothemethode beschreven. Deze methode is een belangrijk instrument om een oplossing van een tijdsafhankelijke PDV te vinden. Meer precies, kan men onder bepaalde veronderstellingen met behulp van deze methode bewijzen dat er een unieke oplossing bestaat.

Hoofdstuk 2 is gebaseerd op het artikel [80]. We beschouwen hier de quasi-statische Maxwellvergelijkingen in combinatie met een onbekende tijdsafhankelijke bron die bepaald wordt op basis van een meting langs een 2D-oppervlak. We tonen aan dat er een unieke oplossing bestaat en we berekenen ook foutenschattingen. De theoretische resultaten worden ondersteund met numerieke voorbeelden.

In Hoofdstuk 3 wordt een niet-lineaire hyperbolische Maxwell vergelijkingen beschouwd met een onbekende tijdsafhankelijke convolutiekern. De ontbrekende gegevens worden bepaald op basis van een meting over de rand van het domein. Opnieuw wordt het bestaan van een uniek oplossing bewezen. Bovendien wordt een numeriek schema gebaseerd op de achterwaartse Eulermethode voorgesteld.

In het laatste hoofdstuk, Hoofdstuk 4, onderzoeken we wiskundige technieken genoemd *model order reduction*, die de laatste jaren veel aandacht hebben getrokken in wetenschappelijk onderzoek. Wij bestuderen de verschillende aspecten van deze technieken, en hun voordelen en nadelen. We vergelijken deze technieken met andere standaardtechnieken zoals bv. ontkoppelde schemas en het minimaliseren van een functionaal. Op deze manier zijn we in staat om de problemen in Hoofdstuk 2 en 3 vanuit een nieuw perspectief te bekijken.

---

# Chapter 1

## Mathematical background

### 1.1 Partial Differential Equations

Considering the fact that most of our research effort, as well as this thesis itself, is devoted to Partial Differential Equations (in literature very often abbreviated as PDEs), we would like to briefly explain and illuminate this term first in order to avoid any misinterpretations.

#### 1.1.1 History of PDEs

Partial differentiation and partial integration are processes which had been known even before the term partial differential equations was established [13]. Leibniz employed partial process in his letter to l'Hospital in 1694 where he introduced the symbol  $\delta m$  to denote the partial derivative  $\partial m / \partial x$  and  $\vartheta m$  to denote  $\partial m / \partial y$ . Leibniz considered  $bdx + cdy$  where  $b$  and  $d$  involved  $x$  and  $y$ . His goal was to find  $m = 0$ , where  $m$  also involved  $x$  and  $y$ . Differentiating  $m = 0$  gives  $\delta m dx + \vartheta m dy = 0$ , which is in fact a total differential equation. However, it remains unclear whether he himself recognised this as a partial differential equation.

In the same year he published an article where he tried to find the envelope of the circles  $x^2 + y^2 + b^2 = 2bx + ab$ . From the steps he made in order to find the

required envelope (the parabola  $ax + a^2/4 = y^2$ ) it is clear that he used partial processes when he kept both  $x$  and  $y$  constant and took  $b$  as independent variable, even though he did not use any special name for this process.

Partial derivatives do not have only a geometric interpretation. One of the greatest mathematicians, Sir Isaac Newton, solved in 1736 the differential equation  $2\dot{x} - \dot{z} + \dot{y}x = 0$ , where the fluxion  $\dot{x}$  stated for  $dx/dt$ . Many writers interpreted this as partial differential equation.

### 1.1.2 Formal definition of PDEs

As the two historical examples show a partial differential equation is an equation that involves an unknown function which depends on several independent variables and its partial derivatives. In a case when the unknown function depends only on a single variable, the term ordinary differential equations is used.

**Definition 1.1.1.** Let  $(x_1, \dots, x_n) \in \Omega \in \mathbb{R}^n$  be the domain of a function  $u = u(x_1, \dots, x_n)$ . A partial differential equation for this function is an equation of a form:

$$f\left(x_1, \dots, x_n, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_n}, \frac{\partial^2 u}{\partial x_1^2}, \dots, \frac{\partial^2 u}{\partial x_n^2}, \dots\right) = 0.$$

The highest order of derivative that appears in such an equation also defines the order of the PDE. Moreover, if  $f$  from Definition 1.1.1 is linear, then the whole equation is called linear. Otherwise the term nonlinear PDE is used.

**Example 1.** For better illustration

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} - f(x, t) = 0$$

is a linear PDE of second order.

### 1.1.3 Initial and boundary conditions

Partial differential equations have in general infinitely many solutions. In practice there usually exists a need to find one particular solution. In order to find this

---

special one some additional conditions are taken into account. There exist two types of conditions that specify the solution:

- boundary conditions,
- initial conditions (for time-dependent equations).

### Initial conditions

In a case of evolution equations (that is a type of problems where the unknown quantity does not depend only on the space variable, but the solution evolves in time as well) an initial state at some time (mostly  $t = 0$ ) is given. A good example is the heat equation, which models the distribution of temperature in a given space and time;  $u = u(x, t)$  where  $x \in \Omega$  and  $t \in (0, T)$ .

**Example 2.** Define an initial value problem as

$$\begin{aligned}u_t - \Delta u &= f && \text{in } \Omega \times (0, T) \\u(x, 0) &= u_0(x) && \text{in } \Omega\end{aligned}\tag{1.1}$$

Here the second equality stands for the initial condition (IC). Having this condition linked to the equation, the solution at any other time is determined<sup>1</sup>. A partial differential equation together with an initial condition create initial value problem.

### Boundary conditions

The other important type of additional condition is called boundary condition (BC) for a very simple reason. In this case the behaviour of the solution near the boundary of a space domain is known (or given). If the space domain is denoted by  $\Omega$ , the boundary is often marked as  $\partial\Omega$ . Boundary conditions together with a prescribed PDE form a *Boundary value problem*. A solution to a boundary value problem is a solution to the differential equation which also satisfies the boundary conditions.

---

<sup>1</sup>On condition that  $\Omega =$  the whole space. Otherwise the boundary conditions are still needed to determine the solution

---

It is worth mentioning that not only the value of a solution at a boundary can be known. Knowing a derivative of the solution may also be sufficient to characterise the desired solution. Therefore we distinguish between:

- *Dirichlet boundary condition.* The value of a solution at the boundary is prescribed.

For the Example 2 it means  $u(x, t) = g(x, t)$  on  $\partial\Omega \times (0, T)$ .

- *Neumann boundary condition.* The derivative of a solution at the boundary is known.

For the heat equation from Example 2 it means  $\frac{\partial u}{\partial \nu} = g(x, t)$  on  $\partial\Omega \times (0, T)$ .

These two condition can be joined together, when a linear combination of the values and the values of the derivative of the solution is given. In such a situation we talk about *Robin boundary condition*.

None of the above-mentioned conditions needs to be prescribed for the whole boundary  $\partial\Omega$ . The boundary can be split into several parts and for each part a different type of boundary conditions can be considered. A problem formulated in this way is called a *Mixed boundary problem*.

Apart from that there exist another two types of boundary conditions: *non-local boundary conditions* and *nonlinear boundary conditions*.

### 1.1.4 Methods for finding a solution to a PDE

At first, only analytic methods for solving PDEs were at hand. This *classical approach*, which was mostly used in the past, before the computer era, benefits only from the arguments of mathematical analysis. The best-known methods, which are used even today are:

- Separation of variables,
  - Method of characteristics,
  - Integral transform,
  - Change of variables.
-

Solutions obtained via one of these methods are very elegant, because they all give an explicit solution. Having an explicit solution is the best result that can be achieved. Nevertheless, in most cases, when the problem arises from practical applications, it can not be obtained. A cure to this issue are often numerical methods, which became very popular in a recent history. *Numerical methods* do not aim to give exact solution to the problem, but rather a numerical approximation. Among the most well-known numerical methods are:

- Finite element method (Section 1.4),
- Finite difference method,
- Finite volume method.

The reason why numerical methods are predominant is that computers have become very accessible due to their decreasing price and because these methods can be applied to a very broad range of problems.

### 1.1.5 Strong and weak solutions

The question which has not been answered yet is what does it actually mean to get a solution to a partial differential problem. The best way how to explain what a solution is required to fulfill, is to give an example. For instance, let us assume a problem which needs to be solved is <sup>2</sup>

$$-\Delta u = f, \tag{1.2}$$

where  $u = u(x)$ ,  $f = f(x)$  and  $x \in \Omega \subset \mathbb{R}^n$ , accompanied with Dirichlet boundary condition

$$u = 0, \quad \text{on } \partial\Omega. \tag{1.3}$$

Then it is very reasonable to require that the classical solution  $u$  belong to space  $C^2(\Omega) \cap C(\bar{\Omega})$ .

---

<sup>2</sup>The symbol  $\Delta$  stands for the Laplace operator A.0.3, but any linear or nonlinear operator of second order can be assumed at this point.

---

**Remark 1.1.1.** Note that to find a classical solution means to seek for a function  $u$  that is two times continuously differentiable. That is to say, the problem (1.2) has to be fulfilled for every point in  $\Omega$ . This is in fact a very restrictive condition, which very often can not be achieved in practice. Therefore, the concepts of strong and weak solutions have been developed.

**Example 3.** The need for the introduction of a weak solution may follow from either insufficient regularity of coefficients or from the nonlinear nature of the equation. E.g. porous media equation

$$\partial_t u - \partial_{xx} u^m = 0$$

does not have any classical solution, however, the weak solution does exist.

Weak solutions help to reduce the regularity conditions. In order to get the weak formulation of problem (1.2), it is required to do the following:

1. Multiply the equation with a test function  $\phi$  and integrate over the domain  $\Omega$ ;
2. Apply one of the integration by parts formula (see appendix A) and use the boundary condition;
3. Determine the proper test space.

The first step for the problem (1.2) gives

$$-\int_{\Omega} \Delta u \phi d\Omega = \int_{\Omega} f \phi d\Omega, \quad (1.4)$$

for all  $\phi \in C_0^\infty(\Omega)$ . This regularity condition can be weakened to  $u \in H^2(\Omega)$  and  $\phi \in L^2(\Omega)$ . A *strong* solution of problem (1.2) is a function  $u \in H^2(\Omega)$ , which fulfills (1.4). Note that it is no longer required for the derivatives to exist pointwise.

The second step lowers the regularity of the solution even more, by transferring one derivative to the test function  $\phi$ . Green's theorem (see Appendix A) applied to (1.4) yields

$$\int_{\Omega} \nabla u \cdot \nabla \phi d\Omega = \int_{\Omega} f \phi d\Omega, \quad (1.5)$$


---



where sufficient conditions for this formula to hold true are  $u, \phi \in H_0^1(\Omega)$  and  $f \in L^2(\Omega)$ . Thus, the *weak solution* (also called *variational solution*) is a function  $u \in H_0^1(\Omega)$  which fulfills the above-stated identity (1.5). The existence and uniqueness of such a solution is guaranteed by Riesz's representation theorem or the Lax-Milgram lemma (see C.0.2 and C.0.3).

If the classical solution to problem (1.2) exists, then it also solves the weak formulation (1.5). On the other hand, if a weak solution is regular enough (in this model situation  $u \in C^2(\Omega) \cap C(\bar{\Omega})$ ) then it also solves the classical formulation of the problem (1.2).

**Remark 1.1.2.** *The weak formulation (1.5) is often written as*

$$a(u, \phi) = f(\phi),$$

where  $a(\cdot, \cdot)$  is a bilinear form and stands for the left-hand side of the weak formulation and  $f(\phi)$  is a functional representing the right-hand side of the weak formulation. This notation is also used in this thesis e.g. (1.6).

### 1.1.6 Applications of PDEs in practice

The key point of this part is to demonstrate that partial differential equations are not just a theoretical concept without any practical meaning. Even from the very brief historical introduction it is clear that they arose from a need to model phenomena that could be observed in real life. The following three examples prove that partial differential equations can be and are used in practice.

#### Wave equation

The simplest situation to give rise to the one-dimensional wave equation is the motion of a stretched string – especially the transverse vibration of a string such as the string of a musical instrument. Assume that a string is placed along the  $x$ -axis, is stretched and then fixed at the ends  $x = 0$  and  $x = L$ . The string is then deflected by applying a force and at some instant, usually called  $t = 0$ , released and allowed to vibrate. The quantity of interest is the deflection  $u$  of the string at any point  $0 \leq x \leq L$ , and at any time  $t > 0$ . That means  $u = u(x, t)$ .

---

**Remark 1.1.3.** *In order to get the desired mathematical model, several details have to be neglected, such as the damping forces of air resistance, the weight of the string, etc.*

An application of Newton's Law of motion brings us to the following PDE

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2},$$

where  $c^2 = \frac{T}{\rho}$ , with  $\rho$  being the mass per unit length of the string and  $T$  being the horizontal component of the tension in the string. To determine  $u(x, t)$  uniquely (for a given set of boundary conditions), we must also know the initial deflection of the string at the time  $t = 0$  at which it is released and the initial velocity of the string. This information is in mathematical terms represented as

- $u(x, 0) = f(x)$  (initial position)
- $\frac{\partial u}{\partial t}(x, 0) = g(x)$  (initial velocity)

The boundary conditions are in fact  $u(0, t) = u(L, t) = 0$  for  $t \geq 0$ .

Even this simple model shows the potential of PDEs being applied in practice. More complicated and thus more precise models can be derived if fewer facts are neglected.

### Heat conduction equations

Another famous example is the model of heat conduction where a thin bar or wire of constant diameter and of homogeneous material oriented along the  $x$ -axis is assumed. Presume that the wire is insulated and so thin that it only allows the heat to flow in the  $x$ -direction. Then the temperature  $u$  at any point depends only on the  $x$ -coordinate and the time  $t$ . Application of the principle of conservation of energy yields that  $u(x, t)$  is represented by the PDE

$$\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2}$$

for  $0 \leq x \leq L$  and  $t > 0$ , where  $k$  is a positive constant called the thermal diffusivity of the bar. An additional information about the temperature at initial time

---

$t = 0$  is needed. The initial condition is  $u(x, 0) = f(x)$ . Various types of boundary conditions at  $x = 0$  and  $x = L$  are possible. Either the temperature at the ends of the bar can be known (represented by Dirichlet conditions) or the insulation condition at both ends are given (represented by homogeneous Neumann conditions)

### Maxwell's equations

Maxwell's equations is a common name for a set of partial differential equations that covers the fundamental laws of optics, electrodynamics, electric circuits, and even more. As the main part of this thesis concentrates its effort in solving Maxwell's equations, we omit more information about them at this point and refer the reader to the Section 1.5 where they are explained in more details.

### 1.1.7 Well-posed and Ill-posed problems

It follows from the above paragraphs that a deep knowledge of the tools of mathematical analysis is necessary for finding a solution to a problem modelled as a PDE. Even though the model may arise from physical or natural phenomena, its mathematical solvability is not guaranteed. A proper analysis needs to be done before starting to solve the problem. This holds true especially nowadays, when various automated numerical solvers are available. It could easily happen that a result gained via a computer program would be worth no more than a guess. The reason for that is following: the uniqueness of the solution is not always guaranteed, or the solution can be sensitive to a small perturbation in data. For that particular reason Hadamard<sup>3</sup> formulated criteria which if fulfilled, guarantee that the result is solid.

**Definition 1.1.2.** *Hadamard's criteria [35] for a problem to be called well-posed are:*

1. *The solution to the problem exists;*
2. *The solution is unique;*

---

<sup>3</sup>Jacques Hadamard (1865-1963)

---

3. *The behavior of solution changes continuously with the data.*

If some problem does not fulfill these criteria, it is said to be an *ill-posed* problem. We propose now a couple of ill-posed problems:

**Example 4.** *Arithmetic:*

- *well-posed problem: Multiplication by a small number  $A$ :  $Aq = f$ ;*
- *ill-posed problem: Division by a small number ( $|A| \ll 1$ ):  $A^{-1}f = q$ ;*

**Example 5.** *Calculus:*

- *well-posed problem: integration;*
- *ill-posed problem: differentiation;*

**Example 6.** *Elliptic PDEs:*

- *well-posed problem:  $\Delta u(x) = 0, x \in \Omega$  with boundary condition given for the whole boundary  $\Gamma = \partial\Omega$ ;*
- *ill-posed problem:  $\Delta u(x) = 0, x \in \Omega$  with boundary condition given only for a part of the boundary  $\Gamma_1 \subset \partial\Omega$ ;*

It is worth mentioning that inverse problems, which are discussed in the following section, are very often ill-posed and therefore they require an extremely sensitive approach.

## 1.2 Inverse problems

To this point partial differential equations were only depicted as a tool for modelling physical phenomena. If some universal law is translated into mathematical language then this model can be used to predict the behaviour of the quantity of interest. This is often called *Forward* or *Direct* problem. *Inverse problems* (IPs) as the name suggests do the opposite – they induce the reason that led to the result

---

from the observed data. Roughly speaking inverse problems are those where backward thinking is engaged: getting the data from the model is the direct problem and getting the model from the data is the the inverse problem.

From a mathematical point of view, though, it is unclear which problem is direct and which one is inverse. To distinguish whether or not to talk about inverse problems a very vague definition by J.B. Keller [58] is often cited: *"We call two problems inverses of one another if the formulation of each involves all or part of the solution of the other. Often for historical reasons, one of the two problems has been studied extensively for some time, while the other has never been studied and thus is not so well understood. In such cases, the former one is called the direct problem, while the later one is the inverse problem."*

In connection to PDEs, an inverse problem means for example recovering an unknown parameter used in a model from a measured data. That is in fact the main reason why inverse problems are studied extensively, because they provide valuable information about a physical parameter which is impossible to observe directly.

**Remark 1.2.1.** *Inverse problems do not require PDEs to be part of them. The area which is covered by the term inverse problems is much broader. However, this thesis is dedicated to PDEs and thus the other fields will not be explained in this short overview nor in subsequent chapters.*

Many applications of mathematics in science or engineering lead to inverse problems. Therefore, this research area has developed widely in the recent history. The list of branches of applied science where IPs appear is exhaustively large. For better illustration only a handful of them are named:

- biomedical engineering;
  - image processing;
  - geoscience, vulcanology;
  - natural language processing ;
  - non-destructive material evaluation;
-

- electricity and magnetism.

A huge need to solve inverse problems has arisen among scientists, especially mathematicians, who have been developing sensitive methods for handling these delicate problems.

Current literature identifies three main types of inverse problems:

1. *parameter identification*, where some specific material parameter present in the equation is unknown and the aim is to reconstruct it. For example a source term, a convolution kernel, a diffusion coefficient, etc;
2. *boundary value inverse problems*, where a measurement on the boundary (or on a part of the boundary) can not be taken directly and has to be determined;
3. *evolutionary inverse problems*, where the initial conditions are unknown and have to be reestablished.

In this thesis, parameter identification problems are discussed. Chapter 2 deals with time-dependent source term identification. In Chapter 3 the time-dependent convolution kernel is unknown.

In Section 1.1.7 is stated that inverse problems often do not meet the criteria of well-posedness (see Definition 1.1.2). Being ill-posed means that a problem is very challenging, if not impossible, to solve. We use a very primitive example to illustrate this.

**Example 7.** *Assume a set of problems is given:*

- *Direct problem: What is the result of summing of the integers 2 and 3?*
- *Inverse problem: Which calculational problem gives the result 5?*

The direct problem is straightforward, however, more than one solution to the inverse problem exist. In fact there are infinitely many solutions to this particular problem. Even the restriction of only using the sum of integers is not enough to guarantee uniqueness of a solution. The violation of the third condition also

---

appears regularly. Small *data perturbation* can significantly change the solution, which causes severe numerical problems, because the data are rarely known exactly due to noise in measurements and computational errors. Differentiation was used in the previous section as an example of an ill-posed problem. The following explanation clarifies why.

**Example 8.** Consider a differentiable function  $u \in C^1[0, 1]$  and the associated sequence

$$u_n^\delta(x) := u(x) + \delta \sin \frac{nx}{\delta}, \quad x \in [0, 1],$$

where  $n \in \mathbb{N}$  and  $\delta \in (0, 1)$ . The function  $u$  represents the exact data whereas  $u_n^\delta$  stands for the perturbed data. Then

$$(u_n^\delta)'(x) = u'(x) + n \cos \frac{nx}{\delta}.$$

The maximum norm yields

$$\|u - u_n^\delta\| = \delta,$$

$$\|u' - (u_n^\delta)'\| = n.$$

From the above-stated, it is obvious that an arbitrarily small error  $\delta$  in the data results in an arbitrarily large error  $n$  in the result, which means that differentiation does not depend continuously on the data.

Numerical schemes designed to cope with ill-posedness of inverse problems are named *regularization methods*. Under the term regularization, a process of approximation of an ill-posed problem by a family of neighbouring well-posed problems is understood. Regularization typically involves the inclusion of additional assumptions, the smoothness of the solution for example. Among the regularization methods *Tikhonov regularization* [84] is one of the most popular. The standard process for solving parameter identification problems consists of two parts. First the cost functional, which captures the error between the exact and parametrized solutions, is constructed. Then the minimum of this functional is sought. Even though this method is considered to be standard, one of the goals of this thesis is to propose a different approach, which is explained in later chapters.

---

### 1.3 Rothe's method

The goal of this section is to briefly present Rothe's method as a tool for finding solutions to evolution problems. Understanding of the fundamentals of this method is essential for the reader in order to comprehend the core parts of the topics explained in the following chapters.

Rothe's method bears the name of Erich Rothe<sup>4</sup>, who first published this method in [70] in 1930. The elementary principle of this method applied to evolution problems is to discretize the time domain in order to obtain a (finite) set of elliptic problems, which are solved by means of standard techniques for elliptic problems. This method is therefore sometimes referred to as the method of lines. Once the elliptic equations for each time instance are solved an approximation of a solution to the initial evolution problem is constructed. Subsequently, the convergence of the approximate solution towards the exact solution is proved using arguments of functional analysis.

The scheme consists of several steps.

1. A well-defined variational formulation must be defined, meaning that all terms are finite;
2. The Backward Euler method is used for time discretization, meaning all time derivatives are replaced by differences in order to get a set of elliptic problems;
3. The existence and uniqueness of the solutions of obtained elliptic problems is proved.
4. Using the fact that the solution is known at the initial time thanks to the initial condition, solutions in all other time instances are subsequently evaluated;
5. Rothe's functions (both piece-wise constant and piece-wise linear) are constructed.
6. To reach the goal of Rothe's method, the following steps are made:

---

<sup>4</sup>Erich Hans Rothe (1895-1988) - German mathematician

---



- (a) A priori estimates are made,
- (b) Uniqueness is proved,
- (c) Convergence is proved,
- (d) Error estimates are established (optional).

The undisputed advantage of this approach is twofold. Not only the uniqueness and existence of a solution are proved, but the constructive algorithm is built-in in this method as well. More information about this method can be found in [69] or in books of Jozef Kačur [56, 57], who extended the usability of this method to a broad area of applications.

## 1.4 Finite element method

The finite element method (FEM) is one of the numerical methods used for finding an approximate solution to a partial differential equation, especially to elliptic PDEs. The FEM is in fact a special variation of Galerkin method. For that reason the concept of Galerkin method is introduced first and the explanation of the FEM principle follows. Readers familiar with this topic can skip this section.

### 1.4.1 Galerkin method

The Galerkin method or more precisely Galerkin methods form a class of methods for transforming a continuous operator problem, a PDE for example, to a discrete problem. This method bears a name of Galerkin<sup>5</sup> despite the fact that it was Walther Ritz<sup>6</sup> who discovered this method and to whom Galerkin refers. Therefore, this method can be found in literature under several names: Galerkin, Ritz-Galerkin [33], Petrov-Galerkin or Bubnov-Galerkin method.

Assume  $V$  is a Hilbert space. Let  $a : V \times V \rightarrow \mathbb{R}$  be a bilinear form and  $f \in V^*$ , where  $V^*$  is a dual space. Suppose that  $a$  and  $f$  be the left-hand side and the right-hand side of a weak formulation of a PDE, respectively. The objective is

---

<sup>5</sup>Boris Galerkin (1871-1945), Soviet mathematician and engineer

<sup>6</sup>Walther Ritz (1878-1909), Swiss theoretical physicist

---

to find  $u$  such that

$$a(u, v) = f(v), \quad \forall v \in V. \quad (1.6)$$

Further, suppose that the conditions of Lax-Milgram Lemma are fulfilled in order to guarantee uniqueness of a solution. Due to the fact, that  $V$  is in general an infinite-dimensional space, it is impossible to find the solution  $u$  numerically. The idea of Galerkin method is to substitute the infinite-dimensional space  $V$  by a set of carefully constructed finite dimensional subspaces  $\{V_n\}_{n=1}^{\infty} \subset V$  with the property  $V_n \subset V_{n+1}$ , so that  $V_n$  fills the space  $V$  in the limit. The objective is then to solve (1.6) in every subspace gradually to get a sequence  $\{u_n\}_{n=1}^{\infty}$ , which under additional assumptions converges towards  $u$ , the solution of 1.6.

**Definition 1.4.1.** *Discrete problem. Find  $u_n \in V_n$  such that*

$$a(u_n, v) = f(v), \quad \forall v \in V_n. \quad (1.7)$$

One of the properties of Hilbert spaces is that every subspace inherits all properties of the original space. The properties of  $a(\cdot, \cdot)$  such as boundedness, V-ellipticity and bilinearity are also preserved when restricted to  $V_n \times V_n$ . Thus the Lax-Milgram Lemma guarantees the existence and uniqueness of the solution  $u_n$  to equation (1.7). Thanks to the fact that  $V_n$  is a finite dimensional Hilbert space a finite set  $\{\phi_i\}_{i=1}^n$  forming the basis of  $V_n$  exists. The importance of a finite basis lies in fact that  $u_n$  then can be rewritten as a linear combination of basis functions

$$u_n = \sum_{i=1}^n c_i \phi_i. \quad (1.8)$$

Substituting (1.8) into (1.7) yields

$$a\left(\sum_{i=1}^n c_i \phi_i, v\right) = \sum_{i=1}^n c_i a(\phi_i, v) = f(v), \quad \forall v \in V_n. \quad (1.9)$$

In (1.9)  $\phi_j, j = 1, \dots, n$  can replace  $v$  to get

$$\sum_{i=1}^n c_i a(\phi_i, \phi_j) = f(\phi_j), \quad j = 1, \dots, n \quad (1.10)$$

From equation (1.10) it follows that  $\mathbf{c} = (c_1, \dots, c_n)^T$  solves the algebraic system

$$\mathbf{A}\mathbf{c} = \mathbf{f}, \quad (1.11)$$

where  $\mathbf{f} = (f(\varphi_1), \dots, f(\varphi_n))^T$  and

$$A = \begin{pmatrix} a(\varphi_1, \varphi_1) & \dots & a(\varphi_n, \varphi_1) \\ \vdots & \ddots & \vdots \\ a(\varphi_n, \varphi_1) & \dots & a(\varphi_n, \varphi_n) \end{pmatrix}.$$

The  $V$ -ellipticity of  $a$  results in the positive definiteness of the matrix  $A$ . Then, the matrix is regular and a unique solution to problem (1.11) exists. The initial problem is in this manner transformed into an algebraic system which can be solved via any standard algebraic solver.

Moreover the error  $e_n = u - u_n$  has the following properties of orthogonality.

**Lemma 1.4.1.** *Suppose  $u \in V$  is the exact solution of continuous problem (1.6) and  $u_n \in V_n$  the exact solution of discrete problem (1.7). Then the error  $e_n = u - u_n$  satisfies*

$$a(u - u_n, v) = 0, \quad \forall v \in V_n.$$

This property can be interpreted as follows. The symmetric bilinear form induces an inner product and the norm  $\|v\|_e = \sqrt{a(v, v)}$ ,  $\forall v \in V$ . That means that the previous lemma implies that  $e_n$  is orthogonal to  $V_n$ , which must be understood in no other way than  $u_n$  is an orthogonal projection of  $u$  onto  $V_n$ .

A better relation between the error  $e_n$  and the Galerkin subspace  $V_n$  is established in Cea's lemma.

**Lemma 1.4.2.** *Cea's lemma: Let  $V$  be a Hilbert space,  $a : V \times V \rightarrow \mathbb{R}$  a bilinear bounded  $V$ -elliptic form and  $f \in V^*$ . Let  $u \in V$  be the solution of the problem (1.6). Further, let  $V_n$  be a subspace of  $V$  and  $u_n \in V_n$  be the solution of (1.7). Let  $C$  and  $C_e$  be the continuity and  $V$ -ellipticity constants, respectively. Then*

$$\|u - u_n\|_V \leq \frac{C}{C_e} \inf_{v \in V_n} \|u - v\|_V.$$

The important message of Cea's lemma is that the error  $e_n$  does not depend on the choice of basis of  $V_n$  but on the subspace  $V_n$  only. Therefore, the error remains unaffected when a different basis is chosen. Reasonable choice of the subspace  $V_n$  is thus crucial in order to get sufficiently good approximation of the solution.

---

Another important consequence of Cea's lemma is the convergence of the approximate solutions  $\{u_n\}$  of problem (1.7) towards the exact solution  $u$  of problem (1.6).

**Theorem 1.4.1.** *Let  $V$  be a Hilbert space and  $V_n$  a sequence of finite dimensional subspaces  $V_n \subset V$  for which*

$$\inf_{v \in V_n} \|u - v\|_V \rightarrow 0 \text{ as } n \rightarrow \infty, \forall u \in V,$$

where  $u \in V$  is the solution of problem (1.6). Let  $u_n \in V_n$  be the solution of the Galerkin approximation (1.7). Furthermore, let  $a : V \times V \rightarrow \mathbb{R}$  be a bilinear bounded  $V$ -elliptic form and  $f \in V^*$ . Then, the Galerkin method for problem (1.6) converges:

$$\lim_{n \rightarrow \infty} \|u - u_n\|_V = 0.$$

## 1.4.2 Finite element method

As stated above, the finite element method itself is a special case of Galerkin method. The explanation of what makes it so special and unique follows.

The FEM is a restriction of the Galerkin method to Hilbert spaces  $V$  consisting only of functions defined in a domain  $\Omega \subset \mathbb{R}^n$  and Galerkin subspaces  $V_n \subset V$  consisting only of piece-wise polynomial functions.

The Ciarlet definition of a *finite element* was first introduced in 1975 [16]. His definition reads as:

**Definition 1.4.2.** *A finite element in  $\mathbb{R}^n$  is a triple  $(K, P_K, \Sigma_K)$  where*

1.  $K$  is a closed bounded set in  $\mathbb{R}^n$  with nonempty interior and piecewise smooth boundary (the element domain),
  2.  $P_K$  is a finite dimensional space of functions over the set  $K$  (the space of basis functions),
  3.  $\Sigma_K$  is a set of linearly independent functionals  $\phi_j, j = 1, \dots, N$  (the degrees of freedom).
-

The set  $\Sigma_K$  is  $P_K$ -*unisolvent*, meaning that  $\Sigma_K$  can be taken as a basis of a space  $P'_K$ , the dual space to  $P_K$ . The functions  $p_i \in P_K$  form the *basis of the finite element functions* and have a property

$$\phi_j(p_i) = \delta_{ij} \quad \text{for } i, j = 1, \dots, N.$$

Suppose a sufficiently smooth function  $v : K \rightarrow \mathbb{R}$  is given, so that the degrees of freedom  $\phi_j(v)$  for  $j = 1, \dots, N$  are well-defined. Then the  $P_K$ -*interpolant* is unambiguously defined as

$$\pi v = \sum_{j=1}^N \phi_j(v) p_j.$$

The set  $K$  is usually an interval in 1D, a triangle in 2D and a tetrahedron in 3D, usually a polyhedron in  $\mathbb{R}^n$ <sup>7</sup>. The space  $P_K$  is often a *polynomial space*  $P_k(K)$ , consisting of polynomials with degree less or equal to  $k$ .

**Example 9.** *First-order Lagrange elements. Lagrange elements are the most used family of finite elements. Their degrees of freedom are point values. Let  $\hat{K} \subset \mathbb{R}^n$  be the unit  $n$ -simplex,*

$$\hat{K} = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \geq 0 \quad \forall i = 1, \dots, n \text{ and } \sum_{i=1}^n x_i \leq 1\}$$

which is in fact the convex hull of the set  $\{\mathbf{e}^j\}_{j=0}^n$ , where  $e_i^j = \delta_{ij}$ . For the sake of clarity, only the simplest member of this family is defined, the first-order Lagrange finite element  $(\hat{K}, \hat{P}, \hat{\Sigma})$ . The space  $\hat{P}$  is then defined as the space of all linear functions  $P_1(\hat{K})$  and the corresponding set of degrees of freedom  $\hat{\Sigma}$  is defined symbolically as

$$\hat{\Sigma} = \{p(\mathbf{e}^j), 0 \leq j \leq n\}.$$

**Remark 1.4.1.** *In electromagnetism, a key role is played by the so called Edge elements [63], which are often used for solving Maxwell's equations.*

**Example 10.** *First-order edge elements. Let  $\hat{K} \subset \mathbb{R}^3$  be the unit tetrahedron (the unit 3-simplex).  $\hat{P}$ , the associated finite element space, consisting of homogeneous*

---

<sup>7</sup>In general other types of domains are allowed and are often used in practice. For example  $K$  is an  $n$ -simplex or  $n$ -rectangle

---

linear vector polynomials  $\mathbf{p}(\mathbf{x})$  such that  $\mathbf{p} \cdot \mathbf{x} = \mathbf{0}$ . The degrees of freedom are in this case the line integrals along the edges of the tetrahedron

$$\hat{\Sigma} = \left\{ \int_{\hat{e}} \mathbf{p} \cdot \boldsymbol{\tau} ds, \text{ for each edge } \hat{e} \text{ of } \hat{K} \right\},$$

with  $\boldsymbol{\tau}$  being a unit vector in the direction of  $\hat{e}$ . For more information the reader is referred to [4].

The next step before using the FEM in practice is to set the “global” notions. Let  $\Omega$  be a bounded domain, with the boundary  $\partial\Omega$  such that  $\bar{\Omega} = \bigcup_{K \in T_h} K$ , where the set  $T_h$  is called a mesh or *triangulation* of the domain. The crucial part is to consider only *affine families of finite elements* meaning there exists a unique *invertible affine mapping*  $F_K$  to the *reference element*  $(\hat{K}, \hat{P}, \hat{\Sigma})$  which simplifies the description of the family  $(K, P_K, \Sigma_K)$ . Rather than prescribing such a family for every triple  $(K, P_K, \Sigma_K)$ ,  $K \in T_h$  separately, it is sufficient and much more convenient to have one *reference finite element*  $(\hat{K}, \hat{P}, \hat{\Sigma})$  and the affine mappings  $F_K$ . Having this mapping is not only of practical but of theoretical importance too. The union of all degrees of freedom  $\Sigma_K$  forms a set of global degrees of freedom  $\Sigma_h$ . The associated finite element space  $X_h$  then consists of all functions  $v$ , where  $v|_K \in P_K$ ,  $K \in T_h$ , fulfilling some continuity conditions on the vertices (edges) of the adjacent elements.

**Remark 1.4.2.** *The function  $v$  is not properly defined, because it does not require to be uniquely defined along the faces common to adjacent finite elements.*

The triple  $(K, P_K, \Sigma_K)$ ,  $K \in T_h$  is called to be of *class  $C^0$*  if the space  $X_h$  is in addition a subset of the space of continuous functions defined in  $\bar{\Omega}$ . Providing  $v : \bar{\Omega} \rightarrow \mathbb{R}$  is sufficiently smooth, the *global  $X_h$ -interpolant* of  $v$  is defined as

$$\pi_h v = \sum_{j=1}^M \phi_{j,h}(v) w_j, \tag{1.12}$$

where  $\phi_{j,h}$  are the global degrees of freedom and  $w_j$  are associated global basis functions.

---

**Example 11.** *1<sup>st</sup> order Lagrange FEM in 1D. In a case when  $\Omega = (a, b) \subset \mathbb{R}$ , the whole domain is divided into a finite system of  $n$  disjoint open subintervals  $\Omega_i = (x_{i-1}, x_i)$  of length  $h_i = x_i - x_{i-1}$  for  $i = 1, \dots, n$ , i.e.*

$$a = x_0 \leq x_1 \leq \dots \leq x_{n-1} \leq x_n = b.$$

The interval  $\Omega_i$  stands for  $K$  from the definition of the finite element. Let  $P_1(\Omega_i)$  be the set of all polynomials of degree 1 defined on  $\Omega_i$ . Moreover let  $h = \max_{1 \leq i \leq n} h_i$  denote the largest length (mesh diameter) of all subintervals. The space

$$V_h^1 := \{\varphi \in C(\bar{\Omega}) : \varphi|_{\Omega_i} \in P_1(\Omega_i), i = 1, \dots, n\}$$

is then called the 1<sup>st</sup> order Lagrange FEM space. Note that the space  $P_k(\Omega_i)$  in this scenario plays the role of  $P_K$  from the definition of finite element and  $V_h^1$  is the associated FEM space. The polynomials of degree 1 ( $p_1(x) = a_0 + a_1x$ ) can be determined from their values in 2 nodes. The degrees of freedom from the definition of FEM are the point values in the nodes. The global basis functions  $w_j \in V_h^1$  are defined as

$$w_j(x) = \begin{cases} \frac{x-x_{j-1}}{h_j} & x \in \bar{\Omega}_j; \\ 1 - \frac{x-x_j}{h_{j+1}} & x \in \bar{\Omega}_{j+1}; \\ 0 & \text{otherwise.} \end{cases}$$

The solution of a classical PDE problem in terms of the finite element method resembles very much the Galerkin solution. Consider (1.6). To approximate the solution  $u$ , the finite element space  $V_h \subset X_h$  needs to be defined first. Each finite element space  $V_h$  is then associated with the discrete solution  $u_h$  satisfying

$$a(u_h, v_h) = f(v_h), \quad \forall v_h \in V_h,$$

which is in fact equivalent to the linear system

$$A\mathbf{c} = \mathbf{f}, \quad \mathbf{c} \in \mathbb{R}^M,$$

where

$$A_{ij} = a(\phi_{i,h}, \phi_{j,h}), \quad f_j = f(\phi_{j,h}),$$

and

$$u_h = \sum_{i=1}^M c_i w_i.$$

The degrees of freedom are chosen in a way which ensures that the global basis functions are nonzero only on the adjacent elements. The result is that the matrix  $A$  is sparse. Sparse matrices are in general easier to handle and allow a broader set of advanced linear-system solvers.

This short introduction to the finite element method ends with the following convergence results [16].

**Definition 1.4.3. Conforming FEM.** A finite element method is called conforming if  $V_h \subset V$ .

**Definition 1.4.4. Regular triangulations.** A family of triangulations  $T_h$  of the domain  $\Omega$  is called regular if

- There exists a constant  $\sigma > 0$  such that

$$h_K/\varrho_k \leq \sigma \quad \text{for any simplex } K \in T_h,$$

where  $h_K$  is the diameter of  $K$  and  $\varrho_k$  the supremum of the diameters of the spheres inscribed into  $K$ ;

- The discretization parameter  $h$  approaches zero.

**Theorem 1.4.2.** Let  $T_h$  be a regular family of triangulations of  $\Omega$ . Assume that all the finite elements  $(K, P_K, \Sigma_K)$ ,  $K \in \bigcup_h T_h$  are affine-equivalent to a single reference finite element  $(\hat{K}, \hat{P}, \hat{\Sigma})$  and they are of class  $C^0$ . Assume moreover that there exist integers  $k \geq 0$  and  $l \geq 0$ , such that the following inclusions are satisfied:

$$P_k(\hat{K}) \subset \hat{P} \subset H^l(\hat{K}),$$

$$H^{k+1}(\hat{K}) \hookrightarrow C^s(\hat{K}),$$

where  $s$  is the maximal order of partial derivatives occurring in the definitions of the set  $\hat{\Sigma}$ . Then, there exists a constant  $C$  independent of  $h$  such that, for any function  $v \in H^{k+1}(\Omega)$ ,

$$\|v - \pi_h v\|_{H^m(\Omega)} \leq Ch^{k+l-m} |v|_{H^{k+1}(\Omega)}, \quad 0 \leq m \leq \min\{1, l\}, \quad (1.13)$$

where  $\pi_h v$  is the  $X_h$ -interpolant of the function  $v$ .

---



### FEM solvers

There exist many packages that can be used to solve PDEs using the finite element method. Among the most-known are: AFENA, Agros2D, DUNE, Elmer, FEniCS Project, UG, FreeFEM, GetDP, PLTMG, COMSOL, and more. Some of them are proprietary commercial software, but many of them fall under the GPL license.

To simulate the problems described in later chapters of this thesis the FEniCS Project software was used. The FEniCS Project is a collection of free, open source software components with the goal to enable automated solution of differential equations. The project was initiated in 2003 as a research collaboration between the University of Chicago and Chalmers University of Technology as an umbrella project for a collection of interoperable components among which are:

- UFL (Unified Form Language),
- FIAT (Finite element Automatic Tabulator),
- FFC (FEniCS Form Compiler),
- UFC (Unified Form-assembly Code),
- Instant,
- DOLFIN - a C++/Python library providing data structures and algorithms for finite element meshes and automated finite element assembly.

For more information the reader is referred directly to the FEniCS Project homepage [www.fenicsproject.org](http://www.fenicsproject.org).

## 1.5 Maxwell's equations

Due to the fact that in the core part of this thesis problems that arose from Maxwell's equations are solved, it is proper to state in this section their theoretical and practical meaning. More rigorous and detailed descriptions of Maxwell's equations can be found in [8, 15, 36, 60, 62, 63].

---

Maxwell's equations is a term used for a set of 4 complicated equations that describe the world of electromagnetism. Maxwell's<sup>8</sup> Publication *Treatise on Electricity and Magnetism* is considered to be the foundation of the modern theory of electromagnetism. The very brief and vague description of the meaning of Maxwell's equations is that they describe the universal law which governs the world of electromagnetism – how electric and magnetic fields propagate and interact, and how are they influenced by objects.

### 1.5.1 Description of quantities and notations

The electromagnetic field is on the macroscopic scale described by four vector functions of the space variable  $\mathbf{x} \in \mathbb{R}^3$  and the time variable  $t \in \mathbb{R}$  denoted by  $\mathbf{E}, \mathbf{D}, \mathbf{H}, \mathbf{B}$ , where the fundamental vector fields are  $\mathbf{E}$  – the electric field and  $\mathbf{H}$  – the magnetic field. The vector fields  $\mathbf{D}$  and  $\mathbf{B}$ , which are usually eliminated from the description of electromagnetic fields via suitable constitutive relations, are called the electromagnetic displacement and the magnetic induction, respectively.

An electromagnetic field is formed when sources of static electric charges and the directed flow of electric charges (which is called current) are distributed in time and space. The distribution of charges is given by a scalar charge density function  $\rho$ , whereas the vector function  $\mathbf{J}$  describes the current density. Maxwell's equations for a region of space  $\Omega \subset \mathbb{R}^3$  have the form:

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0, \quad (1.14)$$

$$\nabla \cdot \mathbf{D} = \rho, \quad (1.15)$$

$$\frac{\partial \mathbf{D}}{\partial t} - \nabla \times \mathbf{H} = -\mathbf{J}, \quad (1.16)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (1.17)$$

where equation (1.14) is called Faraday's law, the divergence condition (1.15) is also known as Gauss's law, (1.16) states for Ampère's law modified by Maxwell and finally the equation (1.17) indicates the solenoidal<sup>9</sup> nature of  $\mathbf{B}$ .

---

<sup>8</sup>James Clerk Maxwell (1831–1879) was an Newton/Einstein-level genius.

<sup>9</sup>The term solenoidal originates in Greek and in present context stand for fixed volume.

---

**Notations** In later chapters, where problems arising from the classical Maxwell's equations are solved, we use the symbol  $\Omega$  for denoting a bounded domain  $\Omega \subset \mathbb{R}^3$ , where the domain, either smooth ( $\Omega \in C^{1,1}$ ) or convex, is occupied by an electromagnetic material. The boundary of the domain  $\Omega$  is denoted by  $\partial\Omega = \Gamma$  and the symbol  $\boldsymbol{\nu}$  stands for the outer normal vector associated with the boundary  $\Gamma$ . The symbols  $\nabla \times, \nabla \cdot, \nabla$  denote the operators curl (rotor), divergence and gradient, respectively, as defined in Appendix A. Moreover, the notation  $\partial_t$  or  $\frac{\partial}{\partial t}$  is used for time derivative. Further, we denote by  $(\cdot, \cdot)$  the standard inner product of  $L^2(\Omega)$  and by  $\|\cdot\|$  its induced norm. When working at the boundary  $\Gamma$  a similar notation is used, namely  $(\cdot, \cdot)_\Gamma, L^2(\Gamma)$  and  $\|\cdot\|_\Gamma$ . By  $C([0, T], X)$  the set of abstract continuous functions  $w : [0, T] \rightarrow X$  equipped with the usual norm  $\max_{t \in [0, T]} \|\cdot\|_X$  is denoted and  $L^p((0, T), X)$  is furnished with the norm  $\left( \int_0^T \|\cdot\|_X^p dt \right)^{\frac{1}{p}}$  with  $p > 1$ , cf. [29]. The symbol  $X^*$  stands for the dual space to  $X$ . A standard notation is used for  $C, \varepsilon$  and  $C_\varepsilon$ , which denote generic positive constants depending only on a priori known quantities, where  $\varepsilon$  is small and  $C_\varepsilon = C(\varepsilon^{-1})$  is large.

### Faraday's law

Faraday's<sup>10</sup> law of induction, represented by equation (1.14), is a basic law of electromagnetism predicting the interaction of the magnetic field and electric circuits and the production of electromotive force. It is the fundamental operating principle of transformers, inductors, electrical motors, generators and solenoids. The equation (1.14) wraps the observations Faraday made during his experiments and can be interpreted as follows:

- Electric current gives rise to magnetic fields. Magnetic fields around a circuit give rise to electric current;
- Magnetic field changing in time induces electric field which circulates around it;
- A circulating electric field in time induces Magnetic field changing in time.

---

<sup>10</sup>Michael Faraday (1791 - 1867) was an English scientist

*Derivation of Faraday's law.* What Faraday observed was that the electromotive force is equal to the opposite of the rate of change of magnetic induction, formally written as:

$$\oint \mathbf{E} \cdot d\mathbf{R} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S}.$$

Applying of Stokes theorem yields

$$\int_S \nabla \times \mathbf{E} \cdot d\mathbf{S} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S},$$

which if all members of the equation are moved to the left hand side gives

$$\int_S \left( \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} \right) \cdot d\mathbf{S} = \mathbf{0}.$$

Since this holds true for every domain  $S$  with closed boundary, the final equation holds

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = \mathbf{0}.$$

### Gauss's law

Gauss's law (1.15) describes the behaviour of electric field when electric charges are present. This law is usually expressed in terms of the *electric flux density* and the *electric charge density*. It was first formulated by Lagrange<sup>11</sup> in 1773 and then by Gauss<sup>12</sup> in 1813. The interpretation of Gauss's law is following:

- Electric  $\mathbf{E}$  and electromagnetic displacement  $\mathbf{D}$  field lines start and stop on electric charge;
- $\mathbf{E}$  and  $\mathbf{D}$  field lines beam from positive charges towards negative charges;
- The divergence of electromagnetic displacement over any domain (volume) is equal to the net amount of charge inside the domain.

---

<sup>11</sup>Joseph-Louis Lagrange (1736 - 1813) was an Italian Enlightenment era mathematician and astronomer

<sup>12</sup>Johann Carl Friedrich Gauss (1777 - 1855) was German mathematician

---

*Derivation of Gauss's law.* Gauss's law can be derived from Coulomb's law<sup>13</sup>. A point charge  $q$  produces electric field as follows

$$\mathbf{E} = \frac{kq}{r^3} \mathbf{r}, \quad (1.18)$$

where  $k = \frac{1}{4\pi\epsilon_0}$ , and  $\mathbf{r}$  is the position in spherical coordinates. The way from Coulomb's law to Gauss's law leads through the calculation of the flux of electric field over a closed boundary surface  $S$  of some domain:

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \frac{q}{\epsilon_0}.$$

For several charges within the volume  $V$  enclosed by the surface  $S$ , the superposition principle  $\oint_S \mathbf{E} = \sum q_i/\epsilon_0$  is used and the point charge is replaced by the continuous electric charge density  $\rho$  to get

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} \int_V \rho \, dV.$$

Application of Gauss' theorem yields

$$\int_V \nabla \cdot \mathbf{E} \, dV = \frac{1}{\epsilon_0} \int_V \rho \, dV \Rightarrow \int_V \left( \nabla \cdot \mathbf{E} - \frac{1}{\epsilon_0} \rho \right) \, dV = 0.$$

The final step is to realize that this holds true for every volume  $V$ . Hence,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}.$$

**Remark 1.5.1.** *Microscopically, all charges are the same. There exists a difference between free and bound charge, though. Therefore the results of Gauss's law are sometimes interpreted in terms of  $\mathbf{E}$  and sometimes in terms of  $\mathbf{D}$ . In vacuum  $\mathbf{D} = \epsilon_0 \mathbf{E}$  holds.*

### Ampère's law

Ampère's<sup>14</sup> law, which is given by equation (1.16), explains the relation between the magnetic field around a closed loop to the electric current passing through the loop. The meaning of this law is the following:

<sup>13</sup>Coulomb's inverse-square law is a law of physics which describes the force interacting between static electrically charged particles.

<sup>14</sup>André-Marie Ampère (1775 - 1836) was a French physicist and mathematician.

- A flowing electric current  $\mathbf{J}$  induces a magnetic field that encircles the current;
- A change of electric flux density  $\mathbf{D}$  in time induces a magnetic field that flows around the  $\mathbf{D}$  field;

*Derivation of Ampère's law.* The observations Ampère made during his experiments are similar to those of Faraday. Ampère discovered that the sum of magnetic fields at each point around a current flowing in a wire is equal to the total amount of current encircled by the magnetic field. This in mathematical terms means

$$\oint \mathbf{H} \cdot d\mathbf{L} = I.$$

If a new constant value for the magnetic field is added and named  $H$ , then

$$\oint \mathbf{H} \cdot d\mathbf{L} = 2\pi r H = I \quad \implies \quad H = \frac{I}{2\pi r}$$

implies that the magnetic field is everywhere because  $r$  is arbitrary, and that the magnitude decreases the more the farther the observer stands from the wire. In order to get Ampère's law as stated in equation (1.16), Stoke's theorem has to be applied and the total current  $I$  needs to be replaced by the surface integral of the current density  $\mathbf{J}$ .

$$\int_S (\nabla \times \mathbf{H}) \cdot d\mathbf{S} = \oint \mathbf{H} \cdot d\mathbf{L} = I = \int_S \mathbf{J} \cdot d\mathbf{S}.$$

From there

$$\nabla \times \mathbf{H} = \mathbf{J}.$$

Then, Maxwell's modification of the original Ampère's law comes onto the scene. Maxwell introduced a new term  $\frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}_d$  and named it *displacement current density*. If this term is added to the previous equation, one gets the form of Ampère's law as formulated in (1.16). Maxwell's contribution not only made Faraday's and Ampère's law more symmetrical, but most of all, it led to a better understanding of the propagation of electromagnetic waves.

---

### Gauss's law for magnetism

The meaning of equation (1.17) is that

- The magnetostatic field does not have a source;
- Magnetic monopoles do not exist;
- The divergence of the fields  $\mathbf{B}$  and  $\mathbf{H}$  is always zero through any volume.

#### 1.5.2 Establishing equations for linear media

The relation between the fields  $\mathbf{E}$  and  $\mathbf{H}$  and the fields  $\mathbf{D}$  and  $\mathbf{B}$ , respectively, is straightforward for vacuum or a free space. In that situation

$$\mathbf{D} = \epsilon_0 \mathbf{E} \text{ and } \mathbf{B} = \mu_0 \mathbf{H}$$

where  $\epsilon_0$  is the *electric permittivity*<sup>15</sup> and  $\mu_0$  stands for *magnetic permeability*<sup>16</sup>. However, in general the relations are of the type

$$\mathbf{D} = \mathbf{D}(\mathbf{E}, \mathbf{H}) \text{ and } \mathbf{B} = \mathbf{B}(\mathbf{E}, \mathbf{H}),$$

and the exact form of the relationship depends on the situation under consideration. The dependencies can be linear (in linear materials) or nonlinear (in superconductors, nonlinear optics, ...). In order to establish the relation, the physical background needs to be studied.

**Remark 1.5.2.** *Even though both fields  $\mathbf{H}$  and  $\mathbf{B}$  are often called the same way – electric field, it must be obvious now, that they are not the same. In a nutshell, the magnetic field  $\mathbf{H}$  describes the region, where a magnetic force can be felt, whereas the magnetic flux  $\mathbf{B}$  describes how much magnetic field passes through an area. Similar arguments hold for  $\mathbf{E}$  and  $\mathbf{D}$ .*

---

<sup>15</sup>Permittivity measures how an electric field affects and is affected by a medium.

<sup>16</sup>Permeability measures the ability of a material to support the formation of a magnetic field within itself.

---

### Ohm's law

Another important constitutive relation, which links the equations for electricity and magnetism is *Ohm's law*<sup>17</sup>, which puts current, voltage and resistance into a well-known relation:

$$I = \frac{V}{R}.$$

However, in physics the term Ohm's law is often used to describe various generalizations of the law that was originally formulated by Ohm. The simplest example is

$$\mathbf{J} = \sigma \mathbf{E},^{18}$$

where  $\sigma$  is a material-dependent parameter called *conductivity* and is in fact the inverse of the resistance,  $\mathbf{J}$  is the flowing electric current (analogous to the current  $I$ ) and  $\mathbf{E}$  (an analogue for the voltage  $V$ ) stands for the electric field as described above.

**Example 12.** *In order to derive a numerical scheme, all constitutive relations available are combined together. An illustrative example is proposed:*

- Substitute  $\mathbf{B} = \mu \mathbf{H}$  in (1.14) and  $\mathbf{D} = \epsilon \mathbf{E}$  in (1.16) to get

$$\mu \partial_t \mathbf{H} + \nabla \times \mathbf{E} = \mathbf{0} \quad (1.19)$$

$$\partial_t(\epsilon \mathbf{E}) - \nabla \times \mathbf{H} = -\mathbf{J}. \quad (1.20)$$

- Multiply (1.19) by  $\mu$  and then apply  $\nabla \times$ . Differentiate (1.20) in time:

$$\nabla \times (\partial_t \mathbf{H}) + \nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{E} \right) = \mathbf{0}, \quad (1.21)$$

$$\partial_{tt}(\epsilon \mathbf{E}) - \nabla \times (\partial_t \mathbf{H}) = -\partial_t \mathbf{J}. \quad (1.22)$$

- Substitute (1.22) in (1.21) to get

$$\partial_{tt}(\epsilon \mathbf{E}) + \nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{E} \right) = -\partial_t \mathbf{J}, \quad (1.23)$$

which is an equation ready to be solved. It is worth noting that the governing equation can easily be expressed in terms of  $\mathbf{H}$ ,  $\mathbf{B}$  or  $\mathbf{D}$ . The choice of  $\mathbf{E}$  is arbitrary.

<sup>17</sup>Georg Simon Ohm (1789 - 1854) was a German physicist and mathematician.

<sup>18</sup>In region where  $\sigma \neq 0$  (also called conductor) the electric field produces an electric current.



### 1.5.3 Boundary conditions

The equations (1.14) and (1.16) do not describe the electric and magnetic fields completely, because at the interface between two materials (e.g. copper and air) either  $\mu$  or  $\epsilon$  are discontinuous. From Lemma 5.3. in [63] it follows that  $\nu \nabla \times \mathbf{E}$ , meaning that the tangential component of electric field, has to be continuous across the interface  $S$ . Otherwise the  $\nabla \times \mathbf{E}$  in equation (1.14) is not well defined. Let  $\mathbf{E}_1$  and  $\mathbf{E}_2$  denote the limiting value of electric field approaching the interface  $S$  from region 1 and 2, respectively. Then the condition

$$\nu \times (\mathbf{E}_1 - \mathbf{E}_2) = \mathbf{0} \quad \text{on } S \quad (1.24)$$

must be fulfilled. In a special case when, say, the material in region 2 is a perfect conductor, the condition is slightly different. For superconductors  $\sigma$  tends to become infinitely large. From Ohm's law  $\mathbf{J} = \sigma \mathbf{E}_2$ , it follows that if  $\mathbf{J}$  is to be bounded, then  $\mathbf{E}_2$  has to approach zero in the limit. Once  $\mathbf{E}_2$  vanishes from 1.24, the index 1 can be omitted and the boundary condition becomes of the form

$$\nu \times \mathbf{E} = \mathbf{0}.$$

A different type of boundary conditions is obtained if a well-defined  $\mu \mathbf{H}$  is to be guaranteed. In that case the normal component of  $\mu \mathbf{H}$  needs to be continuous across  $S$  so that

$$\nu \cdot (\mu_1 \mathbf{H}_1 - \mu_2 \mathbf{H}_2) = 0 \quad \text{on } S. \quad (1.25)$$



# Part I



## Chapter 2

# Recovery of a time-dependent source from a surface measurement in Maxwell's equations

### 2.1 Problem formulation

The content of this chapter is based on the article [80], which has been published in the journal *Computers & Mathematics with applications*.

This chapter solves a problem arising from Maxwell's equations (see Section 1.5), where the time-dependent source on the right-hand side of the equation is unknown and is to be determined from a given integral over-determination. The main contribution of this chapter is that the measurement is only taken along a 2D surface. Let us first derive the governing equation that is to be solved.

The classical Maxwell's equations (1.14), (1.15), (1.16), (1.17) are considered. In eddy current problems the change of electric displacement in time is negligible,

hence the term  $\partial_t \mathbf{D}$  in (1.16) is omitted. Further, Ohm's law in a standard form is assumed

$$\mathbf{J} = \sigma \mathbf{E},$$

and a linear homogeneous magnetic material is considered, i.e.

$$\mathbf{B} = \mu \mathbf{H}$$

with a positive constant permeability  $\mu$ . Elimination of  $\mathbf{H}$  leads to

$$\sigma \mu \partial_t \mathbf{E} + \nabla \times \nabla \times \mathbf{E} = -\mu \partial_t \mathbf{J}_{app}, \quad (2.1)$$

where  $\mathbf{J}_{app}$  stands for the source term, which is assumed to be separable, i.e.

$$-\mu \partial_t \mathbf{J}_{app} = h(t) \mathbf{f}(\mathbf{x}).$$

Here the space-dependent function  $\mathbf{f}(\mathbf{x})$  is given but the time-dependent function  $h(t)$  is unknown. The function  $\mathbf{f}$  is requested to obey

$$\mathbf{f} \times \boldsymbol{\nu} = \mathbf{0} \quad \text{on } \Gamma. \quad (2.2)$$

The perfect conductivity on the boundary is modelled via

$$\mathbf{E} \times \boldsymbol{\nu} = \mathbf{0} \quad \text{on } \Gamma \quad (2.3)$$

and the initial datum is prescribed as

$$\mathbf{E}(0, \mathbf{x}) = \mathbf{E}_0(\mathbf{x}). \quad (2.4)$$

For ease of explanation,  $\mu$  and  $\sigma$  can be set to 1.

### Direct problem

If all previous assumptions are combined, the direct problem can be defined as

$$\partial_t \mathbf{E} + \nabla \times \nabla \times \mathbf{E} = h(t) \mathbf{f}(\mathbf{x}) \quad \text{in } \Omega \times [0, T], \quad (2.5)$$

$$\mathbf{E} \times \boldsymbol{\nu} = \mathbf{0} \quad \text{on } \Gamma \times [0, T], \quad (2.6)$$

$$\mathbf{f} \times \boldsymbol{\nu} = \mathbf{0} \quad \text{on } \Gamma \times [0, T], \quad (2.7)$$

$$\mathbf{E}(0, \mathbf{x}) = \mathbf{E}_0(\mathbf{x}) \quad \text{in } \Omega. \quad (2.8)$$


---

**Remark 2.1.1.** *The source term — the right-hand side of the equation (2.5) can be easily augmented by adding a known vector function  $\mathbf{F}(t, \mathbf{x})$  which obeys (2.7) for all  $t$ . This modification will not cause any trouble in the proof technique proposed later in this chapter.*

Since the assumption of  $h(t)$  being unknown was made, it is no longer possible to approach (2.5) as a direct problem.

### Inverse problem

The inverse source problem (ISP) is to find a couple  $\{\mathbf{E}(t, \mathbf{x}), h(t)\}$ . The missing data function  $h(t)$  will be recovered by means of the following measurement along a 2D-surface  $\gamma$

$$\int_{\gamma} \mathbf{E} \cdot \boldsymbol{\nu} \, d\gamma = m(t), \quad \text{with } \gamma \subset \bar{\Omega} \quad (\text{normal component measurement}). \quad (2.9)$$

Such an integral over-determination is frequently used in various inverse problems (IPs) for evolutionary equations, cf. [50, 68, 76] and the references therein. This chapter assumes that the surface over which the measurement is made,  $\gamma$ , is a part of a boundary of a sub-domain  $G \subset \Omega$ , i.e.  $\gamma \subset \partial G^1$ .

**Remark 2.1.2.** *In the case of a divergence free electrical field  $\mathbf{E}$  (which is not assumed in this chapter), this PDE can be seen as a vectorial heat equation due to the well known relation*

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \Delta \mathbf{E}.$$

*Such a case subjected to various BCs has already been studied by many authors, e.g. [39, 40, 41, 68, 77, 78].*

The recovery of an unknown source belongs to the hot topics in IPs. If the unknown source depends on the space variable, one needs an additional space measurement (e.g. solution at the final time), cf. [14, 24, 26, 38, 49, 54, 67, 68, 71, 82]. This means that both kinds of ISPs need totally different additional data.

---

<sup>1</sup>Note that the case (2.9) allows to take the measurement surface  $\gamma$  to be part of boundary  $\Gamma$  as well.

---

Inspection of the existing literature shows that the additional integral measurement is usually taken over the whole domain (or over a sub-domain). Getting such a measurement can often be impossible from the physical point of view or too costly from the practical point of view. Therefore, in this chapter an integration over a 2D-surface  $\gamma$  in (2.9) is considered, which is one of the highlights of this part of our thesis. Only the normal component of the electrical field is measured. A numerical scheme, for approximating the problem, based on the Rothe's method [56] is proposed together with the proofs of convergence and uniqueness, which are obtained by arguments of functional analysis.

## 2.2 Time discretization

The ISP can be seen as a system consisting of the governing PDE (2.5) and the measurement (2.9) accompanied by the boundary conditions (2.6), (2.7) and the initial datum (2.8). The unknown  $h(t)$  can be eliminated from the equation if the measurement operator is applied to the PDE as follows

$$h(t) = \frac{m'(t) + \int_{\gamma} \nabla \times \nabla \times \mathbf{E} \cdot \boldsymbol{\nu} \, d\gamma}{\int_{\gamma} \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma} \quad \text{if } \int_{\gamma} \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma \neq 0. \quad (2.10)$$

Time discretization of an evolutionary problem means that the problem itself is approximated by a sequence of elliptic problems that are solved successively for every time layer. Rothe's method can be also used for the determination of the unknown time-dependent source  $h$ . For ease of explanation, an equidistant time-partitioning of the time frame  $[0, T]$  is considered, with a time step of the size  $\tau = T/n$ , for any  $n \in \mathbb{N}$ . We use the notation  $t_i = i\tau$  and for any function  $z$  we write

$$z_i = z(t_i), \quad \delta z_i = \frac{z_i - z_{i-1}}{\tau}.$$

Consider a system with unknowns  $\{\mathbf{e}_i, h_i\}$  for  $i = 1, \dots, n$ . The discretized ISP is given by (DPi) and (DMPi)

$$\begin{aligned} \delta \mathbf{e}_i + \nabla \times \nabla \times \mathbf{e}_i &= h_i \mathbf{f} && \text{in } \Omega \\ \mathbf{e}_i \times \boldsymbol{\nu} &= \mathbf{0} && \text{on } \Gamma \\ \mathbf{e}_0 &= \mathbf{E}_0, \end{aligned} \quad (\text{DPi})$$


---



$$h_i = \frac{m'_i + \int_{\gamma} \nabla \times \nabla \times \mathbf{e}_{i-1} \cdot \boldsymbol{\nu} \, d\gamma}{\int_{\gamma} \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma}. \quad (\text{DMPi})$$

Note that in (DMPi) the index  $i$  in term  $\nabla \times \nabla \times \mathbf{e}$  is shifted by one. This adjustment allows the creation of a linear decoupled scheme; for a given  $i$ , which corresponds to the  $i$ -th time layer, first the discrete measured problem (DMPi) is solved, then the discrete problem (DPi) and finally the index  $i$  is increased by one.

## 2.3 A priori estimates

Let us recall the standard notation of functional spaces for vectorial fields (cf. [3, 63]), which are used in this and further sections:

$$\begin{aligned} X &= \mathbf{H}(\mathbf{curl}; \Omega) \cap \mathbf{H}(\mathit{div}; \Omega) \\ X_N &= X \cap \{\boldsymbol{\varphi}; \boldsymbol{\varphi} \times \boldsymbol{\nu} = \mathbf{0} \text{ on } \Gamma\} = \mathbf{H}_0(\mathbf{curl}; \Omega) \cap \mathbf{H}(\mathit{div}; \Omega) \end{aligned}$$

The domain  $\Omega$  is supposed to be  $C^{1,1}$  or convex. In both cases  $X_N \subset \mathbf{H}^1(\Omega)$  holds. This follows from [3, Thm. 2.12] if  $\Omega \in C^{1,1}$ , and from [3, Thm. 2.17] if  $\Omega$  is convex.

**Lemma 2.3.1.** *Let  $\Omega \in C^{1,1}$  or  $\Omega$  be convex. Suppose  $\mathbf{f} \in \mathbf{H}_0(\mathbf{curl}; \Omega)$ ,  $\mathbf{E}_0 \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  and  $\nabla \times \nabla \times \mathbf{E}_0 \in X_N$ ,  $m \in C^1([0, T])$  and  $\int_{\gamma} \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma \neq 0$ . Then for any  $i = 1, \dots, n$  there exists a unique couple  $\{\mathbf{e}_i, h_i\}$ , which solves (DPi) and (DMPi). Moreover  $h_i \in \mathbb{R}$ ,  $\mathbf{e}_i \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  and  $\nabla \times \nabla \times \mathbf{e}_i \in X_N$ .*

*Proof.* For a given  $\mathbf{e}_{i-1} \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  and  $\nabla \times \nabla \times \mathbf{e}_{i-1} \in X_N$ , the value of  $h_i$

---

from (DMPi) can be calculated. We successively deduce that

$$\begin{aligned}
|h_i|^2 &= \left| \frac{m'_i + \int_\gamma \nabla \times \nabla \times \mathbf{e}_{i-1} \cdot \boldsymbol{\nu} \, d\gamma}{\int_\gamma \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma} \right|^2 \\
&\leq C \left( 1 + \int_\gamma |\nabla \times \nabla \times \mathbf{e}_{i-1} \cdot \boldsymbol{\nu}|^2 \, d\gamma \right) \\
&\leq C \left( 1 + \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{L}^2(\gamma)}^2 \right) \\
&\stackrel{\gamma \subset \partial G}{\leq} C \left( 1 + \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{L}^2(\partial G)}^2 \right) \\
&\stackrel{\mathbf{H}^1(G) \subset \mathbf{L}^2(\partial G)}{\leq} C \left( 1 + \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{H}^1(G)}^2 \right) \\
&\stackrel{G \subset \Omega}{\leq} C \left( 1 + \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{H}^1(\Omega)}^2 \right) \\
&\stackrel{[3]}{\leq} C \left( 1 + \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{X_N}^2 \right) \\
&\leq C_i.
\end{aligned} \tag{2.11}$$

The PDE from (DPi) can be rewritten as

$$\frac{1}{\tau} \mathbf{e}_i + \nabla \times \nabla \times \mathbf{e}_i = h_i \mathbf{f} + \frac{1}{\tau} \mathbf{e}_{i-1}.$$

The Lax-Milgram lemma guarantees that  $\mathbf{e}_i \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  is a unique solution of the elliptic equation on the  $i$ -th time layer. It can be further seen that

$$\nabla \times \nabla \times \mathbf{e}_i = h_i \mathbf{f} - \delta \mathbf{e}_i \in \mathbf{L}^2(\Omega),$$

and

$$\nabla \times \nabla \times \nabla \times \mathbf{e}_i = h_i \nabla \times \mathbf{f} - \delta \nabla \times \mathbf{e}_i \in \mathbf{L}^2(\Omega).$$

According to  $\mathbf{f} \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  for  $\boldsymbol{\nu}$  associated with  $\Gamma$

$$\nabla \times \nabla \times \mathbf{e}_i \times \boldsymbol{\nu} = h_i \mathbf{f} \times \boldsymbol{\nu} - \delta \mathbf{e}_i \times \boldsymbol{\nu} = \mathbf{0},$$

must hold. Thus  $\nabla \times \nabla \times \mathbf{e}_i \in X_N$ . □

**Lemma 2.3.2.** *Let the assumptions of Lemma 2.3.1 be fulfilled. Then there exists a positive constant  $C$  such that*

$$(i) \max_{1 \leq j \leq n} \|\mathbf{e}_j\|^2 + \sum_{i=1}^n \|\mathbf{e}_i - \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^n \|\nabla \times \mathbf{e}_i\|^2 \tau \leq C \left( 1 + \sum_{i=1}^n h_i^2 \tau \right)$$

$$(ii) \max_{1 \leq j \leq n} \|\nabla \times \mathbf{e}_j\|^2 + \sum_{i=1}^n \|\nabla \times \mathbf{e}_i - \nabla \times \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^n \|\delta \mathbf{e}_i\|^2 \tau \leq C \left( 1 + \sum_{i=1}^n h_i^2 \tau \right).$$

*Proof.* (i) Scalar multiplication of (DPi) by  $\mathbf{e}_i \tau$ , application of Green's theorem, and summation for  $i = 1, \dots, j$  give

$$\sum_{i=1}^j (\delta \mathbf{e}_i, \mathbf{e}_i) \tau + \sum_{i=1}^j \|\nabla \times \mathbf{e}_i\|^2 \tau = \sum_{i=1}^j h_i(\mathbf{f}, \mathbf{e}_i) \tau.$$

Using Abel's summation (B.0.13) yields

$$\begin{aligned} \frac{1}{2} \|\mathbf{e}_j\|^2 + \frac{1}{2} \sum_{i=1}^j \|\mathbf{e}_i - \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^j \|\nabla \times \mathbf{e}_i\|^2 \tau &= \frac{1}{2} \|\mathbf{e}_0\|^2 + \sum_{i=1}^j h_i(\mathbf{f}, \mathbf{e}_i) \tau \\ &\leq \frac{1}{2} \left( \|\mathbf{e}_0\|^2 + \|\mathbf{f}\|^2 \sum_{i=1}^j h_i^2 \tau + \sum_{i=1}^j \|\mathbf{e}_i\|^2 \tau \right) \\ &\leq C \left( 1 + \sum_{i=1}^n h_i^2 \tau \right) + \sum_{i=1}^j \|\mathbf{e}_i\|^2 \tau. \end{aligned}$$

The next steps in proof of this part of Lemma 2.3.2 are applications of the discrete Grönwall lemma (cf. [5]) and taking the maximum over  $1 \leq j \leq n$ .

(ii) If (DPi) is multiplied by  $\delta \mathbf{e}_i \tau$  and Green's theorem is involved and afterwards the result is summed up for  $i = 1, \dots, j$ , the following equality is obtained

$$\sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau + \sum_{i=1}^j (\nabla \times \mathbf{e}_i, \delta \nabla \times \mathbf{e}_i) \tau = \sum_{i=1}^j h_i(\mathbf{f}, \delta \mathbf{e}_i) \tau.$$

Further, Abel's summation, Cauchy's and Young's inequalities are applied to get

$$\begin{aligned}
& \frac{1}{2} \|\nabla \times \mathbf{e}_j\|^2 + \frac{1}{2} \sum_{i=1}^j \|\nabla \times \mathbf{e}_i - \nabla \times \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau = \\
& = \frac{1}{2} \|\nabla \times \mathbf{e}_0\|^2 + \sum_{i=1}^j h_i(\mathbf{f}, \delta \mathbf{e}_i) \tau \\
& \leq \frac{1}{2} \|\nabla \times \mathbf{e}_0\|^2 + C_\varepsilon \|\mathbf{f}\|^2 \sum_{i=1}^j h_i^2 \tau + \varepsilon \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau \\
& \leq C_\varepsilon \left( 1 + \sum_{i=1}^n h_i^2 \tau \right) + \varepsilon \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau.
\end{aligned}$$

Fixing a sufficiently small  $0 < \varepsilon < 1$  closes the proof of this part. □

**Lemma 2.3.3.** *Let the assumptions of Lemma 2.3.1 be fulfilled. Then there exists a positive constant  $C$  such that*

$$\begin{aligned}
& (i) \max_{1 \leq j \leq n} \|\nabla \times \nabla \times \mathbf{e}_j\|^2 + \sum_{i=1}^n \|\nabla \times \nabla \times \mathbf{e}_i - \nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 + \\
& + \sum_{i=1}^n \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau \leq C
\end{aligned}$$

$$(ii) \sum_{i=1}^n h_i^2 \tau \leq C.$$

(iii) Suppose  $\nabla \cdot \mathbf{E}_0, \nabla \cdot \mathbf{f} \in L^2(\Omega)$ . Then  $\max_{1 \leq j \leq n} \|\nabla \cdot \mathbf{e}_j\| \leq C$ .

(iv) Assume  $\nabla \times \nabla \times \mathbf{f} \in \mathbf{L}^2(\Omega)$ . Then

$$\begin{aligned}
& \max_{1 \leq j \leq n} \|\nabla \times \nabla \times \nabla \times \mathbf{e}_j\|^2 + \\
& + \sum_{i=1}^n \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i - \nabla \times \nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 + \\
& + \sum_{i=1}^n \|\delta \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau \leq C.
\end{aligned}$$


---

*Proof.* (i) First, the  $\nabla \times$  operator is applied to (DPi). Then a scalar product of the relation with  $\nabla \times \nabla \times \nabla \times \mathbf{e}_i \tau$  is taken. Afterwards Green's theorem is used and the result is summed up for  $i = 1, \dots, j$  to obtain

$$\sum_{i=1}^j (\delta \nabla \times \nabla \times \mathbf{e}_i, \nabla \times \nabla \times \mathbf{e}_i) \tau + \sum_{i=1}^j \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau \quad (2.12)$$

$$= \sum_{i=1}^j h_i (\nabla \times \mathbf{f}, \nabla \times \nabla \times \nabla \times \mathbf{e}_i) \tau. \quad (2.13)$$

The LHS of (2.12) can be rewritten as follows

$$\begin{aligned} & \frac{1}{2} \|\nabla \times \nabla \times \mathbf{e}_j\|^2 - \frac{1}{2} \|\nabla \times \nabla \times \mathbf{e}_0\|^2 + \\ & + \frac{1}{2} \sum_{i=1}^j \|\nabla \times \nabla \times \mathbf{e}_i - \nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^j \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau. \end{aligned}$$

We recall Nečas inequality [65] or [6, (7.116)]

$$\|z\|_{\mathbf{L}^2(\partial G)}^2 \leq \varepsilon \|\nabla z\|_{\mathbf{L}^2(G)}^2 + C_\varepsilon \|z\|_{\mathbf{L}^2(G)}^2, \quad \forall z \in \mathbf{H}^1(G), 0 < \varepsilon < \varepsilon_0. \quad (2.14)$$

In order to estimate the RHS of (2.13) first Cauchy's and Young's inequalities are engaged to obtain

$$\begin{aligned} \sum_{i=1}^j h_i (\nabla \times \mathbf{f}, \nabla \times \nabla \times \nabla \times \mathbf{e}_i) \tau & \leq \sum_{i=1}^j |h_i| \|\nabla \times \mathbf{f}\| \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i\| \tau \\ & \leq \varepsilon \sum_{i=1}^j \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau + C_\varepsilon \sum_{i=1}^j h_i^2 \tau. \end{aligned}$$

The part of the proof, where the term containing  $h_i$  is estimated, proceeds in a

---

similar way as in Lemma 2.3.1. The result of applying (2.14) is

$$\begin{aligned}
\sum_{i=1}^j h_i^2 \tau &= \sum_{i=1}^j \left| \frac{m'_i + \int_{\gamma} \nabla \times \nabla \times \mathbf{e}_{i-1} \cdot \boldsymbol{\nu} \, d\gamma}{\int_{\gamma} \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma} \right|^2 \tau \\
&\leq C \sum_{i=1}^j \left( 1 + \int_{\gamma} |\nabla \times \nabla \times \mathbf{e}_{i-1} \cdot \boldsymbol{\nu}|^2 \, d\gamma \right) \tau \\
&\leq C \sum_{i=1}^j \left( 1 + \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{L}^2(\gamma)}^2 \right) \tau \\
&\stackrel{\gamma \subset \partial G}{\leq} C \sum_{i=1}^j \left( 1 + \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{L}^2(\partial G)}^2 \right) \tau \\
&\stackrel{(2.14)}{\leq} \sum_{i=1}^j \left( C + \eta \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{H}^1(G)}^2 + C_{\eta} \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{L}^2(G)}^2 \right) \tau \\
&\stackrel{G \subset \Omega}{\leq} \sum_{i=1}^j \left( C + \eta \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{\mathbf{H}^1(\Omega)}^2 + C_{\eta} \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 \right) \tau \\
&\stackrel{[3]}{\leq} \sum_{i=1}^j \left( C + \eta \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|_{X_N}^2 + C_{\eta} \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 \right) \tau \\
&\leq \sum_{i=1}^j \left( C + \eta \|\nabla \times \nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 + C_{\eta} \|\nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 \right) \tau.
\end{aligned}$$

Combination of the previous results yields

$$\begin{aligned}
\|\nabla \times \nabla \times \mathbf{e}_j\|^2 &+ \sum_{i=1}^j \|\nabla \times \nabla \times \mathbf{e}_i - \nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 \\
&+ (1 - \varepsilon - \eta C_{\varepsilon}) \sum_{i=1}^j \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau \\
&\leq C_{\varepsilon} + C_{\varepsilon} C_{\eta} \sum_{i=1}^j \|\nabla \times \nabla \times \mathbf{e}_i\|^2 \tau.
\end{aligned}$$

The necessary steps to be made at this point are fixing a sufficiently small  $\varepsilon$ , and

---

then small  $\eta$  to get

$$\begin{aligned} & \|\nabla \times \nabla \times \mathbf{e}_j\|^2 + \sum_{i=1}^j \|\nabla \times \nabla \times \mathbf{e}_i - \nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 + \\ & \sum_{i=1}^j \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau \leq C + C \sum_{i=1}^j \|\nabla \times \nabla \times \mathbf{e}_i\|^2 \tau. \end{aligned}$$

The rest of the proof follows from Grönwall's argument.

(ii) This part can be obtained readily by combination of (2.11) and the the proof of part (i) stated above.

(iii) Applying the divergence operator to the PDE from (DPi) results in

$$\delta \nabla \cdot \mathbf{e}_i = h_i \nabla \cdot \mathbf{f}.$$

Summation for  $i = 1, \dots, j$  gives

$$\|\nabla \cdot \mathbf{e}_j\| = \left\| \nabla \cdot \mathbf{e}_0 + \nabla \cdot \mathbf{f} \sum_{i=1}^j h_i \tau \right\| \leq \|\nabla \cdot \mathbf{e}_0\| + \|\nabla \cdot \mathbf{f}\| \sum_{i=1}^j |h_i| \tau \leq C$$

by Cauchy's inequality and (ii).

(iv) Apply the  $\nabla \times$  operator to (DPi) and take a scalar product of the relation with  $\nabla \times \nabla \times \nabla \times \delta \mathbf{e}_i \tau$ . Then make use of Green's theorem and sum the result up for  $i = 1, \dots, j$  to get

$$\begin{aligned} & \sum_{i=1}^j \|\delta \nabla \times \nabla \times \mathbf{e}_i\|^2 \tau + \sum_{i=1}^j (\nabla \times \nabla \times \nabla \times \mathbf{e}_i, \delta \nabla \times \nabla \times \nabla \times \mathbf{e}_i) \tau \\ & = \sum_{i=1}^j h_i (\nabla \times \mathbf{f}, \nabla \times \nabla \times \nabla \times \delta \mathbf{e}_i) \tau. \end{aligned} \tag{2.15}$$

The LHS can be rewritten as follows

$$\begin{aligned} & \frac{1}{2} \|\nabla \times \nabla \times \nabla \times \mathbf{e}_j\|^2 - \frac{1}{2} \|\nabla \times \nabla \times \nabla \times \mathbf{e}_0\|^2 \\ & + \frac{1}{2} \sum_{i=1}^j \|\nabla \times \nabla \times \nabla \times \mathbf{e}_i - \nabla \times \nabla \times \nabla \times \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^j \|\nabla \times \nabla \times \delta \mathbf{e}_i\|^2 \tau. \end{aligned}$$


---

For the RHS of (2.15), it can be deduced that

$$\begin{aligned}
& \sum_{i=1}^j h_i (\nabla \times \mathbf{f}, \nabla \times \nabla \times \nabla \times \delta \mathbf{e}_i) \tau \\
&= \sum_{i=1}^j h_i (\nabla \times \nabla \times \mathbf{f}, \nabla \times \nabla \times \delta \mathbf{e}_i) \tau \\
&\leq \sum_{i=1}^j |h_i| \|\nabla \times \nabla \times \mathbf{f}\| \|\nabla \times \nabla \times \delta \mathbf{e}_i\| \tau \\
&\leq C \sum_{i=1}^j |h_i| \|\nabla \times \nabla \times \delta \mathbf{e}_i\| \tau \\
&\leq C_\varepsilon \sum_{i=1}^j h_i^2 \tau + \varepsilon \sum_{i=1}^j \|\nabla \times \nabla \times \delta \mathbf{e}_i\|^2 \tau \\
&\stackrel{(ii)}{\leq} C_\varepsilon + \varepsilon \sum_{i=1}^j \|\nabla \times \nabla \times \delta \mathbf{e}_i\|^2 \tau.
\end{aligned}$$

Freezing a sufficiently small positive  $\varepsilon$ , we conclude the proof.  $\square$

## 2.4 Existence of a solution

Before we start the proof of existence of a solution, let us first introduce the following piecewise linear and piecewise constant functions in time, respectively.

$$\mathbf{E}_n : [0, T] \rightarrow \mathbf{L}^2(\Omega) : t \mapsto \begin{cases} \mathbf{E}_0 & t = 0 \\ \mathbf{e}_{i-1} + (t - t_{i-1})\delta \mathbf{e}_i & t \in (t_{i-1}, t_i] \end{cases}, 0 \leq i \leq n,$$

$$\overline{\mathbf{E}}_n : [0, T] \rightarrow \mathbf{L}^2(\Omega) : t \mapsto \begin{cases} \mathbf{E}_0 & t = 0 \\ \mathbf{e}_i & t \in (t_{i-1}, t_i] \end{cases}, 0 \leq i \leq n.$$

Similarly, the functions  $\overline{h}_n$ ,  $\overline{m}_n$  and  $\overline{m}'_n$  are defined. These prolongations are also called Rothe's (piece-wise linear and continuous, or piece-wise constant) functions. With the help of the above-defined functions, it is now possible to rewrite

---



(DPi) and (DMPi) for the whole time frame as

$$\begin{aligned} \partial_t \mathbf{E}_n(t) + \nabla \times \nabla \times \overline{\mathbf{E}_n}(t) &= \overline{h}_n(t) \mathbf{f} & \text{in } \Omega \\ \overline{\mathbf{E}_n}(t) \times \boldsymbol{\nu} &= \mathbf{0} & \text{on } \Gamma \\ \mathbf{E}_n(0) &= \mathbf{E}_0, \end{aligned} \quad (2.16)$$

$$\overline{h}_n(t) = \frac{\overline{m}'_n(t) + \int_\gamma \nabla \times \nabla \times \overline{\mathbf{E}_n}(t - \tau) \cdot \boldsymbol{\nu} \, d\gamma}{\int_\gamma \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma}. \quad (2.17)$$

At this point, everything is prepared for the proof of existence of a weak solution to (2.5), (2.6), (2.8) and the measurement (2.9) itself.

**Theorem 2.4.1.** *Let  $\Omega \in C^{1,1}$  or  $\Omega$  be convex. Suppose  $\mathbf{f} \in X_N$ ,  $\mathbf{E}_0 \in X_N$  and  $\nabla \times \nabla \times \mathbf{E}_0 \in X_N$ ,  $m \in C^1([0, T])$  and  $\int_\gamma \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma \neq 0$ ,  $\nabla \times \nabla \times \mathbf{f} \in \mathbf{L}^2(\Omega)$ . Then there exists a weak solution  $\{\mathbf{E}, h\}$ , which solves (2.5), (2.3), (2.4) and (2.10). Moreover  $h \in L^2([0, T])$ ,  $\mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty([0, T], X_N)$  with  $\partial_t \mathbf{E} \in L^2([0, T], \mathbf{L}^2(\Omega))$  and  $\nabla \times \nabla \times \mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty([0, T], X_N)$  with  $\partial_t \nabla \times \nabla \times \mathbf{E} \in L^2([0, T], \mathbf{L}^2(\Omega))$ .*

*Proof.* The reflexivity of  $L^2([0, T])$  together with  $\int_0^T \overline{h}_n^2(t) \, dt \leq C$  gives for a subsequence (denoted by the same symbol again)

$$\overline{h}_n \rightharpoonup h \quad \text{in } L^2([0, T]).$$

According to [3, Thm. 2.8] the following compact embedding holds

$$X_N \Subset \mathbf{L}^2(\Omega).$$

A priori estimates from Lemmas 2.3.2 and 2.3.3 imply that

$$\int_0^T \|\partial_t \mathbf{E}_n(t)\|^2 \, dt \leq C, \quad \|\mathbf{E}_n(t)\|_{X_N} \leq C, \quad \forall t \in [0, T].$$

Having the two above-stated estimates, we are allowed to apply [56, Lemma 1.3.13] to get the existence of a vector field  $\mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty([0, T], X_N)$  with  $\partial_t \mathbf{E} \in L^2([0, T], \mathbf{L}^2(\Omega))$  ( $\mathbf{E}$  is differentiable a.e. in  $[0, T]$ ) and a subsequence of  $\mathbf{E}_n$  (which we denote by the same symbol again) such that

$$\begin{cases} \mathbf{E}_n \rightarrow \mathbf{E}, & \text{in } C([0, T], \mathbf{L}^2(\Omega)) & (2.18a) \\ \mathbf{E}_n(t) \rightharpoonup \mathbf{E}(t), & \text{in } X_N, \quad \forall t \in [0, T] & (2.18b) \\ \overline{\mathbf{E}_n}(t) \rightharpoonup \mathbf{E}(t), & \text{in } X_N, \quad \forall t \in [0, T] & (2.18c) \\ \partial_t \mathbf{E}_n \rightharpoonup \partial_t \mathbf{E}, & \text{in } L^2([0, T], \mathbf{L}^2(\Omega)). & (2.18d) \end{cases}$$

In the light of these considerations, we can integrate (2.16) in time, pass to the limit for  $\tau \rightarrow 0$  in the corresponding variational formulation and differentiate with respect to the time variable to arrive at

$$(\partial_t \mathbf{E}(t), \boldsymbol{\varphi}) + (\nabla \times \mathbf{E}(t), \nabla \times \boldsymbol{\varphi}) = h(t) (\mathbf{f}, \boldsymbol{\varphi}), \quad (2.19)$$

which is valid for all  $\boldsymbol{\varphi} \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  and  $\forall t \in [0, T]$ . It can be seen that (2.19) is the variational formulation of (2.5). Thus  $\{\mathbf{E}, h\}$  is a weak solution to (2.5).

The next step of the proof is to pass to the limit for  $\tau \rightarrow 0$  in (2.17). The most difficult task concerns the term  $\int_{\gamma} \nabla \times \nabla \times \overline{\mathbf{E}_n}(t - \tau) \cdot \boldsymbol{\nu} \, d\gamma$ . To handle this, better convergence of approximates is needed. A priori estimates from Lemmas 2.3.2 and 2.3.3 yield

$$\int_0^T \|\partial_t \nabla \times \nabla \times \mathbf{E}_n(t)\|^2 \, dt \leq C, \quad \|\nabla \times \nabla \times \mathbf{E}_n(t)\|_{X_N} \leq C, \quad \forall t \in [0, T].$$

Similarly as in the previous part of this proof we employ [56, Lemma 1.3.13] to get  $\nabla \times \nabla \times \mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty([0, T], X_N)$  with  $\partial_t \nabla \times \nabla \times \mathbf{E} \in L^2([0, T], \mathbf{L}^2(\Omega))$  ( $\nabla \times \nabla \times \mathbf{E}$  is differentiable a.e. in  $[0, T]$ ) and the following convergences in the sense of subsequences

$$\begin{cases} \nabla \times \nabla \times \mathbf{E}_n \rightarrow \nabla \times \nabla \times \mathbf{E}, & \text{in } C([0, T], \mathbf{L}^2(\Omega)) & (2.20a) \\ \nabla \times \nabla \times \mathbf{E}_n(t) \rightarrow \nabla \times \nabla \times \mathbf{E}(t), & \text{in } X_N, \quad \forall t \in [0, T] & (2.20b) \\ \nabla \times \nabla \times \overline{\mathbf{E}_n}(t) \rightarrow \nabla \times \nabla \times \mathbf{E}(t), & \text{in } X_N, \quad \forall t \in [0, T] & (2.20c) \\ \partial_t \nabla \times \nabla \times \mathbf{E}_n \rightarrow \partial_t \nabla \times \nabla \times \mathbf{E}, & \text{in } L^2([0, T], \mathbf{L}^2(\Omega)) & (2.20d) \end{cases}$$

It holds

$$\begin{aligned}
& \left| \int_{\gamma} \nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)] \cdot \boldsymbol{\nu} \, d\gamma \right|^2 \\
& \leq C \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)]\|_{L^2(\gamma)}^2 \\
& \leq \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)]\|_{\mathbf{L}^2(\partial G)}^2 \\
& \stackrel{(2.14)}{\leq} \varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)]\|_{\mathbf{H}^1(G)}^2 \\
& \quad + C_\varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)]\|_{\mathbf{L}^2(G)}^2 \\
& \stackrel{G \subset \Omega}{\leq} \varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)]\|_{\mathbf{H}^1(\Omega)}^2 \\
& \quad + C_\varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)]\|^2 \\
& \stackrel{[3]}{\leq} \varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)]\|_{X_N}^2 \\
& \quad + C_\varepsilon \int_{t-\tau}^t \|\partial_t \nabla \times \nabla \times \mathbf{E}_n(s)\|^2 \, ds \\
& \leq \varepsilon + C_\varepsilon \tau.
\end{aligned}$$

For the limit case, when  $n \rightarrow \infty$ , and then for  $\varepsilon \rightarrow 0$ , we find that

$$\lim_{n \rightarrow \infty} \left| \int_{\gamma} \nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)] \cdot \boldsymbol{\nu} \, d\gamma \right|^2 = 0.$$


---

In a similar way, we deduce that

$$\begin{aligned}
& \left| \int_{\gamma} \nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)] \cdot \boldsymbol{\nu} \, d\gamma \right|^2 \\
& \leq C \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|_{L^2(\gamma)}^2 \\
& \leq \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|_{\mathbf{L}^2(\partial G)}^2 \\
& \stackrel{(2.14)}{\leq} \varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|_{\mathbf{H}^1(G)}^2 \\
& \quad + C_\varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|_{\mathbf{L}^2(G)}^2 \\
& \stackrel{G \subset \Omega}{\leq} \varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|_{\mathbf{H}^1(\Omega)}^2 \\
& \quad + C_\varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|^2 \\
& \stackrel{[3]}{\leq} \varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|_{X_N}^2 \\
& \quad + C_\varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|^2 \\
& \leq \varepsilon + C_\varepsilon \|\nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)]\|^2.
\end{aligned}$$

Passing to the limit for  $n \rightarrow \infty$  and using (2.20a) we get

$$\lim_{n \rightarrow \infty} \left| \int_{\gamma} \nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)] \cdot \boldsymbol{\nu} \, d\gamma \right|^2 \leq \varepsilon.$$

Now, we let go  $\varepsilon \rightarrow 0$  to obtain

$$\lim_{n \rightarrow \infty} \left| \int_{\gamma} \nabla \times \nabla \times [\mathbf{E}_n(t) - \mathbf{E}(t)] \cdot \boldsymbol{\nu} \, d\gamma \right|^2 = 0.$$

The measurement obeys  $m \in C^1([0, T])$ , thus  $\overline{m}'_n(t) \rightarrow m'(t)$  for any  $t \in [0, T]$ .

Collecting the considerations above we see that we may pass to the limit for  $n \rightarrow \infty$  in (2.17) to arrive at (2.10).

□

## 2.5 Uniqueness of a solution

In this section, the uniqueness of a solution to the ISP (2.5), (2.6), (2.8) and (2.10) is proven.

**Theorem 2.5.1.** *Let the assumptions of Theorem 2.4.1 be satisfied. Then there exists at most one weak solution to the ISP (2.5), (2.6), (2.8) and (2.10).*

*Proof.* Due to the linearity of the problem it is sufficient to show that if  $m = 0$  and  $\mathbf{E}(0) = \mathbf{0}$  then  $\mathbf{E}(t) = \mathbf{0}$  and  $h(t) = 0$  for any  $t \in [0, T]$ . We will follow a similar way as we used for the time discretization. Therefore only the main points are pointed out.

**1st step** Take a scalar product of (2.5) with  $\mathbf{E}$ , integrate over  $\Omega$ , make use of Green's theorem and integrate in time to get in a standard way, similarly as in Lemma 2.3.2(i)

$$\|\mathbf{E}(t)\|^2 + \int_0^t \|\nabla \times \mathbf{E}(s)\|^2 ds \leq C \int_0^t h^2(s) ds, \quad \forall t \in [0, T].$$

**2nd step** Take a scalar product of (2.5) with  $\partial_t \mathbf{E}$ , integrate over  $\Omega$ , apply Green's theorem and integrate in time to get as in Lemma 2.3.2(ii)

$$\|\nabla \times \mathbf{E}(t)\|^2 + \int_0^t \|\partial_t \mathbf{E}(s)\|^2 ds \leq C \int_0^t h^2(s) ds, \quad \forall t \in [0, T].$$

**3rd step** Apply the  $\nabla \times$  operator to (2.5) and take a scalar product with  $\nabla \times \nabla \times \mathbf{E}$ . Then integrate over  $\Omega$ , involve Green's theorem and integrate in time to get similarly to Lemma 2.3.3(i) for all  $t \in [0, T]$

$$\|\nabla \times \nabla \times \mathbf{E}(t)\|^2 + \int_0^t \|\nabla \times \nabla \times \mathbf{E}(s)\|^2 ds \leq C \int_0^t h^2(s) ds.$$


---

**4th step** Start from (2.10) with  $m = 0$ . Then it can be deduced that

$$\begin{aligned}
\int_0^t h^2(s) \, ds &= \int_0^t \left| \frac{\int_\gamma \nabla \times \nabla \times \mathbf{E}(s) \cdot \boldsymbol{\nu} \, d\gamma}{\int_\gamma \mathbf{f}(\mathbf{x}) \cdot \boldsymbol{\nu} \, d\gamma} \right|^2 \, ds \\
&\leq C \int_0^t \left( \int_\gamma |\nabla \times \nabla \times \mathbf{E}(s) \cdot \boldsymbol{\nu}|^2 \, d\gamma \right) \, ds \\
&\leq C \int_0^t \left( \|\nabla \times \nabla \times \mathbf{E}(s)\|_{\mathbf{L}^2(\gamma)}^2 \right) \, ds \\
&\stackrel{\gamma \subset \partial G}{\leq} C \int_0^t \left( \|\nabla \times \nabla \times \mathbf{E}(s)\|_{\mathbf{L}^2(\partial G)}^2 \right) \, ds \\
&\stackrel{(2.14)}{\leq} \int_0^t \left( \eta \|\nabla \times \nabla \times \mathbf{E}(s)\|_{\mathbf{H}^1(G)}^2 + C_\eta \|\nabla \times \nabla \times \mathbf{E}(s)\|_{\mathbf{L}^2(G)}^2 \right) \, ds \\
&\stackrel{G \subset \Omega}{\leq} \int_0^t \left( \eta \|\nabla \times \nabla \times \mathbf{E}(s)\|_{\mathbf{H}^1(\Omega)}^2 + C_\eta \|\nabla \times \nabla \times \mathbf{E}(s)\|^2 \right) \, ds \\
&\stackrel{[3]}{\leq} \int_0^t \left( \eta \|\nabla \times \nabla \times \mathbf{E}(s)\|_{X_N}^2 + C_\eta \|\nabla \times \nabla \times \mathbf{E}(s)\|^2 \right) \, ds \\
&\leq \int_0^t \left( \eta \|\nabla \times \nabla \times \nabla \times \mathbf{E}(s)\|^2 + C_\eta \|\nabla \times \nabla \times \mathbf{E}(s)\|^2 \right) \, ds.
\end{aligned} \tag{2.21}$$

Combining this relation with the third step yields

$$\begin{aligned}
\|\nabla \times \nabla \times \mathbf{E}(t)\|^2 + (1 - \eta) \int_0^t \|\nabla \times \nabla \times \nabla \times \mathbf{E}(s)\|^2 \, ds \\
\leq C_\eta \int_0^t \|\nabla \times \nabla \times \mathbf{E}(s)\|^2 \, ds.
\end{aligned}$$

Fixing a small positive  $\eta$ , Grönwall's argument returns

$$\max_{t \in [0, T]} \|\nabla \times \nabla \times \mathbf{E}(t)\|^2 + \int_0^T \|\nabla \times \nabla \times \nabla \times \mathbf{E}(s)\|^2 \, ds = 0.$$

This and (2.21) imply that  $\int_0^T h^2(s) \, ds = 0$ . Taking into account the 1st and 2nd steps, we also see that

$$\max_{t \in [0, T]} \|\mathbf{E}(t)\|^2 + \max_{t \in [0, T]} \|\nabla \times \mathbf{E}(t)\|^2 = 0,$$

which concludes the proof of uniqueness.  $\square$

Theorem 2.4.1 shows convergence of iterates for a subsequence. Theorem 2.5.1 ensures the uniqueness of a solution. Combination of both results implies convergence for the whole sequences of approximations  $\mathbf{E}_n, h_n$ .

## 2.6 Error estimates

Finally, in this section, the error estimates between the semi-discretized and the exact solution are proposed.

**Theorem 2.6.1.** *Suppose the conditions of Theorem 2.4.1 are fulfilled. Then*

$$\begin{aligned} \max_{t \in [0, T]} \|\mathbf{E}(t) - \mathbf{E}_n(t)\|^2 &\leq C\tau, \\ \int_0^T \|\nabla \times (\mathbf{E}(t) - \bar{\mathbf{E}}_n(t))\|^2 dt &\leq C\tau, \\ \max_{t \in [0, T]} \|\nabla \times \nabla \times (\mathbf{E}(t) - \mathbf{E}_n(t))\|^2 &\leq C\tau, \\ \int_0^T \|\nabla \times \nabla \times \nabla \times (\mathbf{E}(t) - \bar{\mathbf{E}}_n(t))\|^2 dt &\leq C\tau, \\ \int_0^T |h(t) - \bar{h}_n(t)|^2 dt &\leq C\tau. \end{aligned}$$

*Proof.* The proof is split into several parts. First the terms

$$\begin{aligned} &\int_0^t |\bar{h}_n(s) - h(s)|^2 ds, \\ \|\mathbf{E}(t) - \mathbf{E}_n(t)\|^2 &+ \int_0^t \|\nabla \times (\mathbf{E}(s) - \mathbf{E}_n(s))\|^2 ds, \\ \|\nabla \times \nabla \times (\mathbf{E}(t) - \mathbf{E}_n(t))\|^2 &+ \int_0^t \|\nabla \times \nabla \times \nabla \times (\mathbf{E}(s) - \bar{\mathbf{E}}_n(s))\|^2 ds, \end{aligned}$$


---

are estimated. The partial estimates are then combined together and Grönwall's lemma gives the final argument to conclude the proof. Several auxiliary estimates are used in order to perform the individual steps of this proof. Among them

$$\|\mathbf{E}_n(t) - \overline{\mathbf{E}}_n(t)\| = \|\mathbf{e}_{i-1} + (t - t_{i-1})\delta\mathbf{e}_i - \mathbf{e}_i\| \leq \|\delta\mathbf{e}_i\| \tau, \quad \forall t \in [t_{i-1}, t_i],$$

which is valid for any norm.

### Step I.

First subtract the measurement problem (MP) (2.10) from the discretized measurement problem (DMP) (2.17) to get

$$\begin{aligned} \overline{h}_n(t) - h(t) &= \\ &= \frac{1}{\int_{\gamma} \mathbf{f} \cdot \boldsymbol{\nu} d\gamma} \left( \overline{m}'_n(t) - m(t) + \int_{\gamma} \nabla \times \nabla \times (\overline{\mathbf{E}}_n(t - \tau) - \mathbf{E}(t)) \cdot \boldsymbol{\nu} d\gamma \right). \end{aligned}$$

Using the assumptions  $m \in C^2([0, T])$  and  $\int_{\gamma} \mathbf{f} \cdot \boldsymbol{\nu} d\gamma \neq 0$  gives

$$|\overline{h}_n(t) - h(t)| \leq C \left( \tau + \int_{\gamma} |\nabla \times \nabla \times (\overline{\mathbf{E}}_n(t - \tau) - \mathbf{E}(t)) \cdot \boldsymbol{\nu}| d\gamma \right).$$

An easy calculation yields

$$|\overline{h}_n(t) - h(t)|^2 \leq C \left( \tau^2 + \|\nabla \times \nabla \times (\overline{\mathbf{E}}_n(t - \tau) - \mathbf{E}(t))\|_{\mathbf{L}^2(\gamma)}^2 \right),$$

which can be rewritten as

$$\begin{aligned} |\overline{h}_n(t) - h(t)|^2 &\leq C \left( \tau^2 + \|\nabla \times \nabla \times (\overline{\mathbf{E}}_n(t - \tau) - \overline{\mathbf{E}}_n(t))\|_{\mathbf{L}^2(\gamma)}^2 \right. \\ &\quad \left. + \|\nabla \times \nabla \times (\overline{\mathbf{E}}_n(t) - \mathbf{E}(t))\|_{\mathbf{L}^2(\gamma)}^2 \right). \end{aligned} \quad (2.22)$$

The second term on the right-hand side can be further estimated using the facts that  $\gamma \subset \partial G$ ,  $G \subset \Omega$ , the trace theorem and [3] as follows

$$\begin{aligned} &\|\nabla \times \nabla \times (\overline{\mathbf{E}}_n(t - \tau) - \overline{\mathbf{E}}_n(t))\|_{\mathbf{L}^2(\gamma)}^2 \leq \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|_{\mathbf{L}^2(\gamma)}^2 \\ &\leq \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|_{\mathbf{L}^2(\partial G)}^2 \\ &\leq \varepsilon \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|_{\mathbf{H}^1(G)}^2 + C_{\varepsilon} \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|_{\mathbf{L}^2(G)}^2 \\ &\leq \varepsilon \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|_{\mathbf{H}^1(\Omega)}^2 + C_{\varepsilon} \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|^2 \\ &\leq \varepsilon \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|_{X_N}^2 + C_{\varepsilon} \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|^2 \\ &\leq \varepsilon \|\nabla \times \nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|^2 + C_{\varepsilon} \|\nabla \times \nabla \times (\mathbf{e}_i - \mathbf{e}_{i-1})\|^2. \end{aligned} \quad (2.23)$$



The third term in (2.22) can be estimated in a similar manner as

$$\begin{aligned}
& \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|_{\mathbf{L}^2(\gamma)}^2 \leq \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|_{\mathbf{L}^2(\partial G)}^2 \\
& \leq \varepsilon \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|_{\mathbf{H}^1(G)}^2 + C_\varepsilon \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|_{\mathbf{L}^2(G)}^2 \\
& \leq \varepsilon \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|_{\mathbf{H}^1(\Omega)}^2 + C_\varepsilon \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|^2 \\
& \leq \varepsilon \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|_{X_N}^2 + C_\varepsilon \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|^2 \\
& \leq \varepsilon \|\nabla \times \nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|^2 + C_\varepsilon \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(t) - \mathbf{E}(t))\|^2.
\end{aligned} \tag{2.24}$$

Substituting (2.23) and (2.24) into (2.22) and integrating over the time frame  $\int_0^t ds$  results in

$$\begin{aligned}
\int_0^t |\bar{h}_n(s) - h(s)|^2 ds & \leq C_\varepsilon \tau + \varepsilon \int_0^t \|\nabla \times \nabla \times \nabla \times (\bar{\mathbf{E}}_n(s) - \mathbf{E}(s))\|^2 ds \\
& + C_\varepsilon \int_0^t \|\nabla \times \nabla \times (\bar{\mathbf{E}}_n(s) - \mathbf{E}(s))\|^2 ds.
\end{aligned} \tag{2.25}$$

### Step II.

In order to administer the second step, it is first needed to subtract (2.16) from (2.5), multiply the result with test function  $\varphi$ , and integrate over  $\Omega$  to get the following formulation

$$(\partial_t(\mathbf{E}(t) - \mathbf{E}_n(t)), \varphi) + (\nabla \times (\mathbf{E}(t) - \bar{\mathbf{E}}_n(t)), \nabla \times \varphi) = (h(t) - \bar{h}_n(t))(\mathbf{f}, \varphi).$$

Using the definition of the norm in  $\mathbf{H}(\mathbf{curl}; \Omega)^*$  we are allowed to write

$$\begin{aligned}
\|\partial_t(\mathbf{E} - \mathbf{E}_n)\|_{\mathbf{H}(\mathbf{curl}; \Omega)^*} & = \sup_{\substack{\varphi \in \mathbf{H}(\mathbf{curl}; \Omega) \\ \|\varphi\|_{\mathbf{H}(\mathbf{curl}; \Omega)} \leq 1}} (\partial_t(\mathbf{E} - \mathbf{E}_n), \varphi) \\
& = \sup_{\substack{\varphi \in \mathbf{H}(\mathbf{curl}; \Omega) \\ \|\varphi\|_{\mathbf{H}(\mathbf{curl}; \Omega)} \leq 1}} [(h - \bar{h}_n)(\mathbf{f}, \varphi) - (\nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n), \nabla \times \varphi)] \\
& \leq C (|h - \bar{h}_n| + \|\nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|).
\end{aligned}$$

For the sake of clarity, the argument  $t$  is omitted for the functions  $\mathbf{E}(t)$ ,  $\mathbf{E}_n(t)$ ,

$\bar{\mathbf{E}}_n(t)$ ,  $h(t)$  and  $\bar{h}_n(t)$ . Choose the test function  $\varphi$  to be  $\varphi = \mathbf{E} - \bar{\mathbf{E}}_n$  to get

$$\begin{aligned} & (\partial_t(\mathbf{E} - \mathbf{E}_n), \mathbf{E} - \bar{\mathbf{E}}_n) + (\nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n), \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)) \\ & = (h - \bar{h}_n) (\mathbf{f}, \mathbf{E} - \bar{\mathbf{E}}_n). \end{aligned}$$

Adding  $\pm \mathbf{E}_n$  and subsequent use of Cauchy's inequality results in

$$\begin{aligned} & \frac{1}{2} \partial_t \|\mathbf{E} - \mathbf{E}_n\|^2 + \|\nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 \leq \\ & \leq |(\partial_t(\mathbf{E} - \mathbf{E}_n), \mathbf{E}_n - \bar{\mathbf{E}}_n)| + |h - \bar{h}_n| \|\mathbf{f}\| \|\mathbf{E} - \bar{\mathbf{E}}_n\| \\ & \leq |(\partial_t(\mathbf{E} - \mathbf{E}_n), \mathbf{E}_n - \bar{\mathbf{E}}_n)| + C |h - \bar{h}_n| \|\mathbf{E} - \mathbf{E}_n + \mathbf{E}_n - \bar{\mathbf{E}}_n\| \\ & \leq |(\partial_t(\mathbf{E} - \mathbf{E}_n), \mathbf{E}_n - \bar{\mathbf{E}}_n)| \\ & \quad + C \left( |h - \bar{h}_n|^2 + \|\mathbf{E} - \mathbf{E}_n\|^2 + \|\mathbf{E}_n - \bar{\mathbf{E}}_n\|^2 \right) \\ & \leq C \|\partial_t(\mathbf{E} - \mathbf{E}_n)\|_{\mathbf{H}(\text{curl}; \Omega)^*} \|\mathbf{E}_n - \bar{\mathbf{E}}_n\|_{\mathbf{H}(\text{curl}; \Omega)} \\ & \quad + C |h - \bar{h}_n|^2 + C \|\mathbf{E} - \mathbf{E}_n\|^2 + C \tau^2 \|\delta \mathbf{e}_i\|^2 \\ & \leq C [|h - \bar{h}_n| + \|\nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|] \|\mathbf{E}_n - \bar{\mathbf{E}}_n\|_{\mathbf{H}(\text{curl}; \Omega)} \\ & \quad + C |h - \bar{h}_n|^2 + C \|\mathbf{E} - \mathbf{E}_n\|^2 + C \tau^2 \|\delta \mathbf{e}_i\|^2 \\ & \leq \vartheta \|\nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 + C_\vartheta \tau^2 \|\delta \mathbf{e}_i\|_{\mathbf{H}(\text{curl}; \Omega)}^2 \\ & \quad + C |h - \bar{h}_n|^2 + C \|\mathbf{E} - \mathbf{E}_n\|^2 + C \tau^2 \|\delta \mathbf{e}_i\|^2. \end{aligned}$$

Integrating the result over time  $\int_0^t ds$  leads to

$$\begin{aligned} & \|\mathbf{E}(t) - \mathbf{E}_n(t)\|^2 + (1 - \vartheta) \int_0^t \|\nabla \times (\mathbf{E}(s) - \bar{\mathbf{E}}_n(s))\|^2 ds \\ & \leq C \left[ \int_0^t |h(s) - \bar{h}_n(s)|^2 ds + \int_0^t \|\mathbf{E}(s) - \mathbf{E}_n(s)\|^2 ds \right] + C_\vartheta \tau^2, \end{aligned} \tag{2.26}$$

which rounds out the second part of the proof.

### Step III.

The third step begins similarly as the previous step by subtracting (2.16) from (2.5). After that, the curl  $\nabla \times$  operator is applied, the result is multiplied by the

test function  $\varphi$  and integrated over  $\Omega$ .

$$\begin{aligned} (\nabla \times \partial_t(\mathbf{E} - \mathbf{E}_n), \varphi) + (\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n), \varphi) &= \\ &= (h - \bar{h}_n) (\nabla \times \mathbf{f}, \varphi). \end{aligned}$$

The test function is set to be  $\varphi = \nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)$ , Green's theorem is utilized, the artificial zero  $\mathbf{0} = \pm \mathbf{E}_n$  is added and the Cauchy-Schwarz inequality is used to get

$$\begin{aligned} \frac{1}{2} \partial_t \|\nabla \times \nabla \times (\mathbf{E} - \mathbf{E}_n)\|^2 + \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 \\ \leq |(\partial_t \nabla \times \nabla \times (\mathbf{E} - \mathbf{E}_n), \nabla \times \nabla \times (\mathbf{E}_n - \bar{\mathbf{E}}_n))| \\ + |h - \bar{h}_n| \|\nabla \times \mathbf{f}\| \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|. \end{aligned}$$

The right-hand-side (RHS) of this expression can be further estimated using similar techniques as in previous step as follows

RHS

$$\begin{aligned} &\leq \|\partial_t \nabla \times \nabla \times (\mathbf{E} - \mathbf{E}_n)\|_{\mathbf{H}(\mathbf{curl}; \Omega)^*} \|\nabla \times \nabla \times (\mathbf{E}_n - \bar{\mathbf{E}}_n)\|_{\mathbf{H}(\mathbf{curl}; \Omega)} \\ &\quad + C |h - \bar{h}_n| \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\| \\ &\leq \|\partial_t \nabla \times \nabla \times (\mathbf{E} - \mathbf{E}_n)\|_{\mathbf{H}(\mathbf{curl}; \Omega)^*} \|\nabla \times \nabla \times (\mathbf{E}_n - \bar{\mathbf{E}}_n)\|_{\mathbf{H}(\mathbf{curl}; \Omega)} \\ &\quad + \eta \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 + C_\eta |h - \bar{h}_n|^2 \\ &\leq C [|h - \bar{h}_n| + \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|] \|\nabla \times \nabla \times \delta \mathbf{e}_i\|_{\mathbf{H}(\mathbf{curl}; \Omega)} \tau \\ &\quad + \eta \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 + C_\eta |h - \bar{h}_n|^2 \\ &\leq \theta |h - \bar{h}_n|^2 + \theta \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 + C_\theta \tau^2 \|\nabla \times \nabla \times \delta \mathbf{e}_i\|^2 \\ &\quad + \eta \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 + C_\eta |h - \bar{h}_n|^2. \end{aligned}$$

Supposing  $\theta$  and  $\eta$  are small enough, we can write

$$\begin{aligned} \partial_t \|\nabla \times \nabla \times (\mathbf{E} - \mathbf{E}_n)\|^2 + \|\nabla \times \nabla \times \nabla \times (\mathbf{E} - \bar{\mathbf{E}}_n)\|^2 \\ \leq C \left( |h - \bar{h}_n|^2 + \|\nabla \times \nabla \times \delta \mathbf{e}_i\|^2 \tau^2 \right). \end{aligned} \quad (2.27)$$

Integrating (2.27) in time  $\int_0^t ds$  provides us with the final result of this part of the

---

proof.

$$\begin{aligned} & \|\nabla \times \nabla \times (\mathbf{E}(t) - \mathbf{E}_n(t))\|^2 + \int_0^t \|\nabla \times \nabla \times \nabla \times (\mathbf{E}(s) - \bar{\mathbf{E}}_n(s))\|^2 ds \\ & \leq C \left[ \tau^2 + \int_0^t |h(s) - \bar{h}_n(s)|^2 ds + \int_0^t \|\nabla \times \nabla \times (\mathbf{E}(s) - \mathbf{E}_n(s))\|^2 ds \right]. \end{aligned} \tag{2.28}$$

**Step IV.**

The last step is to combine the previous results together:

- add (2.26) to (2.28),
- substitute  $\int_0^t |h(s) - \bar{h}_n(s)|^2 ds$  from (2.25),
- assume the constants  $\varepsilon$  and  $\vartheta$  are small enough,
- apply Grönwall's lemma,
- use the result in (2.25) to get the estimate for  $|h - \bar{h}_n|$ .

□

## 2.7 Numerical Computations

A few tangible examples are offered in this section to demonstrate that the above-proved theoretical results are also valid in practice. The main objective of the experiments was to test whether the computational scheme gives reasonable results and if it converges as postulated.

---

### 2.7.1 First Experiment

We consider the inverse source problem (2.3)-(2.5) in  $\Omega = (0, 1) \times (0, 1) \times (0, 1)$  and  $T = 1$ . Furthermore, we set

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &= (\sin x_1 + 2 \sin x_2, \sin x_2 + 2 \sin x_3, \sin x_3 + 2 \sin x_1), \\ \gamma &= \{\mathbf{x} \in \partial\Omega : x_1 = 0\}, \\ m(t) &= (\cos 1 - 1)e^t. \end{aligned}$$

It can be easily seen that the exact solution we are looking for is

$$\begin{aligned} \mathbf{E}(t, \mathbf{x}) &= (e^t(\sin x_1 + 2 \sin x_2), e^t(\sin x_2 + 2 \sin x_3), e^t(\sin x_3 + 2 \sin x_1)), \\ h(t) &= e^t. \end{aligned}$$

We split the time interval  $(0, T)$  into 100 equidistant subintervals meaning that the time step is  $\tau = 0.01$ . We create a uniform mesh in our space domain as in Figure 2.1. We uniformly divide the unit cube into 1000 small cubes and each one of the cubes is split into 6 tetrahedra. We employ Lagrange FEM of second order, which results into a system with 27783 degrees of freedom (DoF).

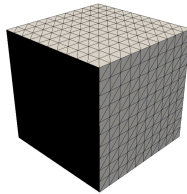


Figure 2.1: Mesh in domain: The left hand side of the cube (colored black) is denoted  $\gamma \subset \partial\Omega$ . Measurement  $m(t)$  is performed only on this subset of the boundary.

Since we know the exact solution  $\mathbf{E}$ , we are able to calculate the error of our numerical scheme. This can be seen on Figures 2.2 and 2.3. Using the numerical scheme we proposed and proved in the previous sections are able to reconstruct the function  $h(t)$  very precisely, which is demonstrated in Figure 2.4 (left). According to our proofs, we expect that the decay in error between exact function  $h(t)$  and its numerical reconstruction correlates with the diminishing of the time

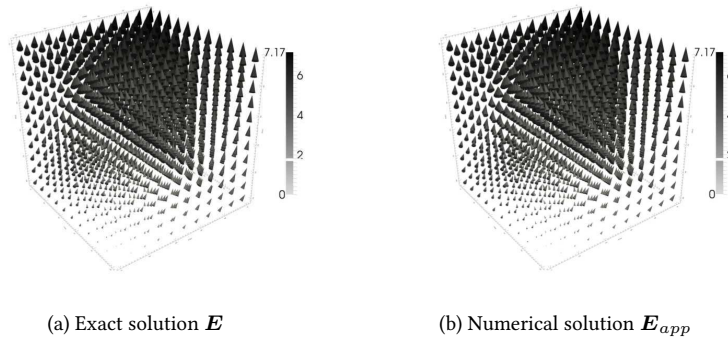


Figure 2.2: Exact and numerical solutions for  $\mathbf{E}$

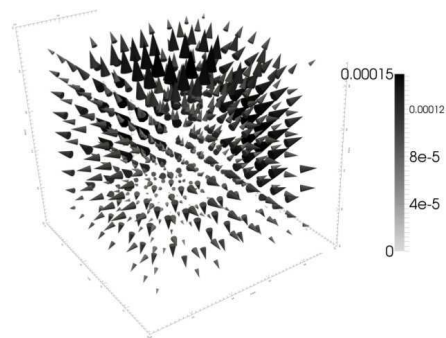
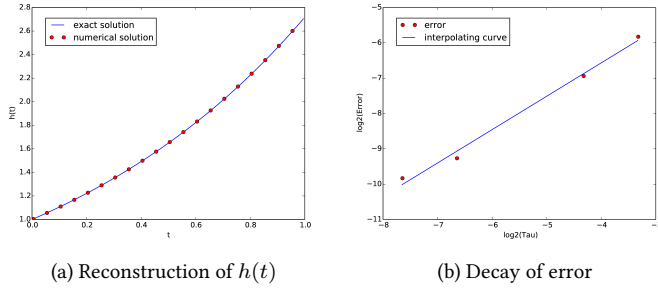


Figure 2.3: Relative error  $\frac{\mathbf{E} - \mathbf{E}_{app}}{|\mathbf{E}|}$  depicted for  $t = T$ . Its absolute value is denoted by color.

Figure 2.4: Reconstruction of the function  $h(t)$  and its error

step  $\tau$ . The errors for several values of  $\tau$  are plotted in Figure 2.4 (right). The linear regression line through data points is given by  $0.9444 \log_2 \tau - 0.78948$ , which is even better result than expected. To prove the optimal convergence rate  $O(\tau)$ , a better a priori estimates for the approximate solution are needed.

## 2.7.2 Second Experiment

In the second example, we take  $\mathbf{x} \in \mathbb{R}^3$ ,  $\mathbf{x} \in \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq 1\}$  and  $t \in (0, T)$  where  $T = 1$ . Furthermore we set

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &= (x_1, x_2, x_3), \\ \gamma &= \{\mathbf{x} \in \partial\Omega : x_1 \geq 0\}, \\ m(t) &= 2\pi e^t. \end{aligned}$$

The exact solution we are using for testing is

$$\begin{aligned} \mathbf{E}(t, \mathbf{x}) &= (e^t x_1, e^t x_2, e^t x_3) = e^t \mathbf{x}, \\ h(t) &= e^t. \end{aligned}$$

We split the time interval  $(0, T)$  into 100 equidistant subintervals (with the time step of the length  $\tau = 0.01$ ). We create then a quasi-uniform mesh on our space domain consisting of tetrahedra as can be seen on Figure 2.5. We use the Lagrange FEM of second order, which creates a system with 17871 DoF.

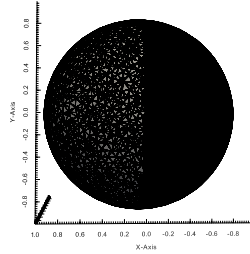


Figure 2.5: Mesh in domain.

The right-hand-side of the sphere (colored black) is denoted  $\gamma \subset \partial\Omega$ . Measurement  $m(t)$  is performed only on this subset of the boundary

The approximation of  $\mathbf{E}$  and its error are depicted in Figures 2.6 and 2.7. Our numerical scheme is able to reconstruct the function  $h(t)$  almost perfectly, see Figure 2.8 (left). The errors for various values of  $\tau$  are plotted in Figure 2.8. The linear regression line through data points is given by  $0.97942 \log_2 \tau - 0.3418$ , which is also better than the theoretical result.

**Noisy data** One important aspect that is worth a short discussion is the term “measurement” itself. The measurement is in practice never exact either due to the imperfect measuring devices or due to the computer precision limit. Certain level of noise is always introduced into the computational process. Therefore, it is very important to know how (or more importantly if) the proposed numerical scheme can handle such inaccuracy. To test this ability we run our model several times. Each time the measurement  $m(t)$  is slightly modified.

The modification reads as follows. First a random noise of a certain level, represented by the function  $\delta(t)$ , is added to the measurement

$$m_{noisy}(t) := m(t) + \delta(t).$$

Then, since the assumption  $m \in C^1([0, T])$  was made, smoothing the data is the next step. This can be done very easily by approximating the noisy data with a smooth function. Most of the modern solvers can handle this task automatically.

---



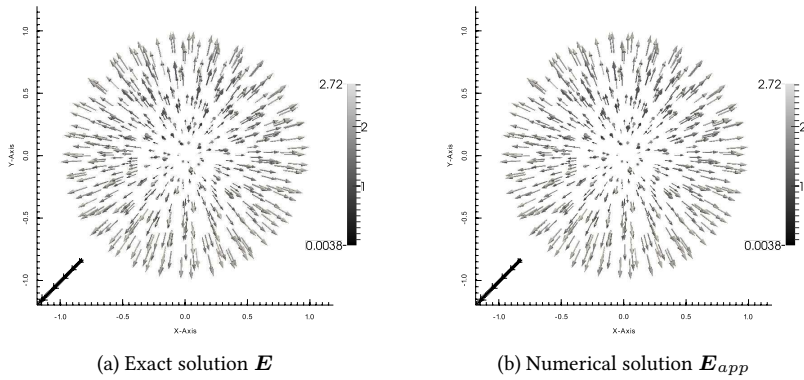
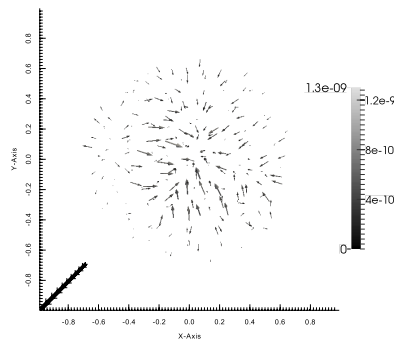
Figure 2.6: Exact and numerical solutions for  $E$ 

Figure 2.7: Relative error  $\frac{E - E_{app}}{|E|}$  depicted for  $t = T$ . Its absolute value is denoted by color.

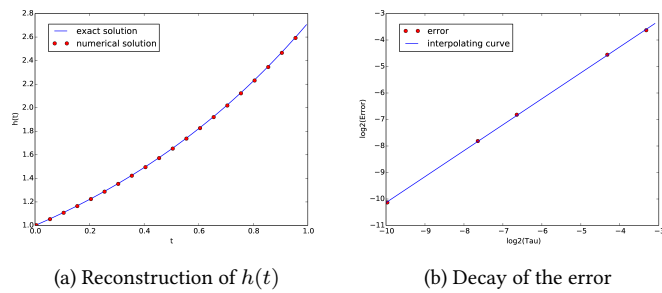


Figure 2.8: Reconstruction of the function  $h(t)$  and its error

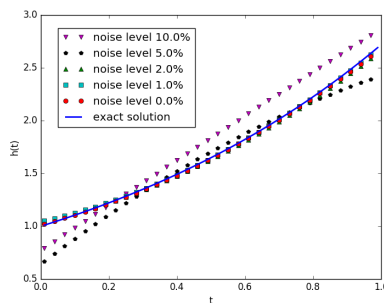


Figure 2.9: Reconstruction of the function  $h(t)$  when the measurement  $m(t)$  contain a certain level of noise

Figure 2.9 demonstrates, that our scheme is able to deal with the noisy data providing that the noise level is low.

## Chapter 3

# Determination of a time-dependent convolution kernel from a boundary measurement in nonlinear Maxwell's equations

### 3.1 Problem formulation

This chapter, similarly to Chapter 2.1, deals with a problem that is derived from the standard Maxwell's equations (1.14), (1.15), (1.16), (1.17), which are described in Section 1.5. The work presented in this chapter has been published [79] in the journal *Computers & Mathematics with applications*.

The assumptions of smooth or convex domain  $\Omega$  remains valid in this part of our thesis as well as all other notations mentioned in previous chapter.

Physical laws binding the four unknowns vectorial fields  $B, D, E, H$  from

Maxwell's equations are usually accompanied by constitutive relations of the type

$$\mathbf{D} = \mathbf{D}(\mathbf{E}, \mathbf{H}), \quad \mathbf{B} = \mathbf{B}(\mathbf{E}, \mathbf{H}),$$

where the exact form of the relationships depends on the situation under consideration. The dependencies can be linear (in linear materials) or nonlinear (in superconductors, nonlinear optics, ...). Note that in the previous chapter we assumed linear homogeneous magnetic material, where the relation  $\mathbf{B} = \mu\mathbf{H}$  holds true. However, in practical situations the relation between magnetic and electric fields is often much more general. In some situations, the present values of solutions depend on their past behaviour. Hysteresis may be expressed by using a memory term in the form of a time integral. Applications can be found in chiral media [83], meta-materials [47, 48] or polarized media [85]. The authors of [8] have considered a nonlinear memory effect for polarization  $\mathbf{P}$  of the type

$$\mathbf{P}(t) = (g * [\mathbf{E} + \mathbf{f}(\mathbf{E})])(t).$$

Here the symbol  $*$  stands for the convolution in time, which is usually understood as  $(K * \mathbf{u}(\mathbf{x}))(t) = \int_0^t K(t-s)\mathbf{u}(\mathbf{x}, s) ds$ . The formulation from [8] can be interpreted as a generalization of the Debye or Lorentz polarization models [7] in the sense that the polarization dynamics is driven by a nonlinear function of the electric field.

In our work, we adopt a generalized Ohm's law<sup>1</sup> of the following form

$$\mathbf{J} = \sigma * \mathbf{E} - 1 * \mathbf{g}(\mathbf{E}).$$

Further, we assume that

$$\mathbf{D} = \varepsilon\mathbf{E}$$

with a constant absolute permittivity  $\varepsilon$ , and a nonlinear magnetic material, i.e.

$$\mathbf{B} = \mu(\mathbf{H} - 1 * \mathbf{f}(\mathbf{E}))$$

with a positive variable permeability  $\mu$ . Elimination of  $\mathbf{H}$  from the set of four Maxwell's equations ((1.14), (1.15), (1.16), (1.17)) leads to

$$\varepsilon\mathbf{E}_{tt} + (\sigma * \mathbf{E})_t + \nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{E} \right) = \mathbf{g}(\mathbf{E}) + \nabla \times \mathbf{f}(\mathbf{E}) - \mathbf{J}_t^{app}.$$

---

<sup>1</sup>Note that in previous chapter we assumed Ohm's law in standard form  $\mathbf{J} = \sigma\mathbf{E}$ .

---

The conductivity term  $\sigma$  is assumed to be separable, i.e.

$$\sigma(\mathbf{x}, t) = \alpha(t)\tilde{\sigma}(\mathbf{x}), \quad (3.1)$$

where the given  $\tilde{\sigma}(\mathbf{x})$  describes the heterogeneity of the material. We assume that  $\tilde{\sigma}$  is constant along  $\Gamma$  with  $\tilde{\sigma}|_{\Gamma} = \sigma^{\Gamma}$ ,  $\sigma^{\Gamma} \neq 0$ . This means that the possible inhomogeneity is an interior one. The hereditary weight  $\alpha(t)$  is unknown and is has to be determined. A perfect contact on the boundary is assumed and the initial data are given.

### Direct problem

For the ease of explanation, it can be without any loss of generality assumed that  $\varepsilon = 1$  and that the term  $\mathbf{J}^{app}$  is omitted for a better readability<sup>2</sup>.

Collecting all previous premises gives

$$\mathbf{E}_{tt} + \tilde{\sigma}(\alpha * \mathbf{E})_t + \nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{E} - \mathbf{f}(\mathbf{E}) \right) = \mathbf{g}(\mathbf{E}), \quad (3.2)$$

$$\mathbf{E} \times \boldsymbol{\nu} = \mathbf{0}, \quad (3.3)$$

$$\mathbf{E}(\mathbf{x}, 0) = \mathbf{E}_0(\mathbf{x}), \quad (3.4)$$

$$\mathbf{E}_t(\mathbf{x}, 0) = \mathbf{V}_0(\mathbf{x}), \quad (3.5)$$

where (3.2) is the governing equation defined in  $\Omega \times [0, T]$ , (3.3) is a boundary condition modelling perfect conductivity on the boundary  $\Gamma$  and (3.4) together with (3.5) express initial conditions. As stated above, the weight function  $\alpha(t)$  is unknown, thus this problem can not be solved directly.

### Inverse problem

The inverse problem (IP) is to find a couple  $\{\mathbf{E}(\mathbf{x}, t), \alpha(t)\}$ . The missing data function  $\alpha(t)$  will be recovered by means of the following measurement along  $\Gamma$

$$\int_{\Gamma} \mathbf{E} \cdot \boldsymbol{\nu} \, d\gamma = m(t), \quad (\text{normal component measurement}). \quad (3.6)$$

---

<sup>2</sup>The term  $\mathbf{J}^{app}$  can be handled in a standard way adopting suitable regularity assumptions.

**Remark 3.1.1.** *As mentioned in Chapter 2, such an integral over-determination can be seen in e.g [50, 68, 80]. This appears mostly for diffusion processes as in [39, 40, 41, 68, 77, 78].*

**Overview of existing literature** Identification of the missing memory kernels in evolutionary PDEs was implemented in 1990's by the Russian and Italian schools. This started the research in this area, cf. [12, 17, 18, 19, 20, 30, 31, 34, 51, 52, 53]. The paper [31] studies a linear hyperbolic equation without a damping term, where the time convolution operator acts on the Laplacian of the solution. The unknown data are revealed from a point measurement. No constructive algorithm for finding a solution for this IP is presented. The paper [30] deals with identification of convolution kernels in abstract linear hyperbolic integro-differential problems. The local solvability in time of this IP is shown. However, there is no constructive algorithm for recovery of the missing convolution kernel. The article [51] is devoted to a one dimensional integro-differential linear hyperbolic problem. The error estimates (for a numerical scheme based on finite differences with dependent time and space mesh-steps) are derived under high regularity of solution. [12] presents a nice study of properties of Dirichlet-to-Neumann maps for memory reconstruction for linear settings. In [20] a global in time existence and uniqueness result for an inverse problem arising in the theory of heat conduction for materials with memory has been studied. The paper [18] derives some local and global in time existence results for the recovery of memory kernels.

A constructive and very interesting technique for the recovery of missing convolution kernels has been developed in [22, 23, 81] for scalar parabolic and hyperbolic equations. Here the additional measurement was a space integral of the solution over  $\Omega$ , namely  $\int_{\Omega} u(x, t) dx = m(t)$ . The goal of this chapter is an adaptation of this approach for vectorial electromagnetic fields. The natural function space (where the solution is expected to live) is  $\mathbf{H}(\mathbf{curl}; \Omega)$  or its subspace. Unfortunately,  $\mathbf{H}(\mathbf{curl}; \Omega)$  is not compactly embedded into  $\mathbf{L}^2(\Omega)$ , which causes troubles in the convergence analysis. This can be overcome with additional assumptions. For that purpose, vectorial function spaces of non-divergence free vectorial fields will be used.

Another very important highlight of this part of our thesis is that we consider

---

a “non invasive” measurement (3.6), where the integral is taken over the boundary  $\Gamma$ , which represents a new and essential aspect. Only the normal component of the electrical field is measured. We prove the existence and uniqueness of a solution of the IP given by (3.3), (3.4), (3.6) and (3.2). Moreover, we propose a numerical scheme for approximations based on the backward Euler’s method. Finally, we show the existence of approximations at each time step of the time partitioning and we establish some regularity results. The convergence of iterates towards the exact solution is obtained by arguments of functional analysis.

To see clearly how the the governing PDE (3.2) and the measurement (3.6) are associated with each other, let us first apply the divergence operator to (3.2) to get

$$\nabla \cdot (\mathbf{E}_{tt} + \tilde{\sigma} (\alpha * \mathbf{E})_t - \mathbf{g}(\mathbf{E})) = 0.$$

Integration over  $\Omega$  and the subsequent use of the mass balance argument (divergence theorem) imply

$$\int_{\Gamma} \mathbf{E}_{tt} \cdot \boldsymbol{\nu} + \int_{\Gamma} \tilde{\sigma} (\alpha * \mathbf{E})_t \cdot \boldsymbol{\nu} = \int_{\Gamma} \mathbf{g}(\mathbf{E}) \cdot \boldsymbol{\nu}.$$

Employing the assumption that  $\tilde{\sigma}|_{\Gamma} = \sigma^{\Gamma}$  leads to

$$\begin{aligned} \int_{\Gamma} \tilde{\sigma} (\alpha * \mathbf{E})_t \cdot \boldsymbol{\nu} &= \sigma^{\Gamma} \int_{\Gamma} (\alpha * \mathbf{E}_t + \alpha \mathbf{E}_0) \cdot \boldsymbol{\nu} \\ &= \sigma^{\Gamma} \left[ \alpha * \int_{\Gamma} (\mathbf{E} \cdot \boldsymbol{\nu})_t + \alpha \int_{\Gamma} \mathbf{E}_0 \cdot \boldsymbol{\nu} \right]. \end{aligned}$$

Thus, it can be written

$$m'' + \sigma^{\Gamma} [\alpha * m' + \alpha m_0] = (\mathbf{g}(\mathbf{E}) \cdot \boldsymbol{\nu}, 1)_{\Gamma}. \quad (\text{MP})$$

In the following text the standard notation of functional spaces for vectorial fields  $X$  and  $X_N$ <sup>3</sup> is used as defined in [3, 63].

*Note:* All other symbols standing for constants, norms, spaces, etc. have exactly the same meaning as in the previous chapter.

---

<sup>3</sup>The domain  $\Omega$  is supposed to be  $\Omega \in C^{1,1}$  or convex, that’s why we have  $X_N \subset \mathbf{H}^1(\Omega)$ . This follows from [3, Thm. 2.12] if  $\Omega \in C^{1,1}$ , and from [3, Thm. 2.17] if  $\Omega$  is convex.

---

**Variational formulation** The variational form of (3.2) for  $\varphi \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  and a.e. in  $[0, T]$  reads as

$$\begin{aligned} (\mathbf{E}_{tt}, \varphi) + (\alpha * \mathbf{E}_t + \alpha \mathbf{E}_0, \tilde{\sigma} \varphi) + \left( \frac{1}{\mu} \nabla \times \mathbf{E}, \nabla \times \varphi \right) \\ = (\mathbf{f}(\mathbf{E}), \nabla \times \varphi) + (\mathbf{g}(\mathbf{E}), \varphi). \end{aligned} \quad (\text{P})$$

The weak formulation of the IP given by (3.3), (3.4), (3.6) and (3.2) is to find a couple  $\{\mathbf{E}(\mathbf{x}, t), \alpha(t)\}$  obeying (P), (MP) and (3.4) such that

- $\alpha \in L^2((0, T))$ ,
- $\mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty((0, T), X_N)$ ,
- $\mathbf{E}_t \in L^2((0, T), \mathbf{L}^2(\Omega)) \cap C([0, T], X_N^*)$ ,
- $\mathbf{E}_{tt} \in L^2((0, T), \mathbf{H}_0(\mathbf{curl}; \Omega)^*)$

## 3.2 Time discretization

Using a simple discretization in time (backward Euler), a time-dependent problem is approximated by a sequence of elliptic problems, which have to be solved successively for increasing  $t_i$ . Solutions of these steady-state settings approximate the transient solution at the points of the time partitioning. The advantage of using Rothe's method is twofold: next to the proof of existence and possible uniqueness of a solution to the original problem, also a numerical algorithm is contained in this approach.

Let us first define the discrete convolution in time as follows <sup>4</sup>

$$(K * v)_i := \sum_{k=0}^i K_{i-k} v_k \tau.$$

---

<sup>4</sup>For ease of explanation, we consider an equidistant time-partitioning of the time frame  $[0, T]$  with a step  $\tau = T/n$ , for any  $n \in \mathbb{N}$ . We use the notation  $t_i = i\tau$  and for any function  $z$  we write  $z_i = z(t_i)$ ,  $\delta z_i = \frac{z_i - z_{i-1}}{\tau}$ .

---



A simple calculation yields<sup>5</sup>

$$\delta(K * v)_i = \frac{(K * v)_i - (K * v)_{i-1}}{\tau} = K_0 v_i + \sum_{k=0}^{i-1} \delta K_{i-k} v_k \tau, \quad i \geq 1. \quad (3.7)$$

Consider a system with unknowns  $(e_i, \alpha_i)$  for  $i = 1, \dots, n$ . At time layer  $t_i$  the problem (P) can be approximated by

$$\begin{aligned} (\delta^2 e_i, \varphi) + ((\alpha * \delta e)_i + \alpha_i e_0, \tilde{\sigma} \varphi) + \left( \frac{1}{\mu} \nabla \times e_i, \nabla \times \varphi \right) & \quad (\text{DP}i) \\ = (\mathbf{f}(e_{i-1}), \nabla \times \varphi) + (\mathbf{g}(e_{i-1}), \varphi) \end{aligned}$$

with  $e_0 = \mathbf{E}_0$  and  $\delta e_0 = \mathbf{V}_0$ . The measurement problem (MP) can be approximated by

$$m''_i + \sigma^\Gamma [(\alpha * m')_i + \alpha_i m_0] = (\mathbf{g}(e_{i-1}) \cdot \boldsymbol{\nu}, 1)_\Gamma \quad (\text{DMP}i)$$

with  $m_0 = m(0)$ .

**Remark 3.2.1.** *The initial and boundary conditions may naturally disagree at  $t = 0$ , which leads to a shock at the beginning of the process. Compatibility conditions for PDEs are relations between the initial conditions, the PDE, and the boundary conditions, which are necessary for getting better regularity of a solution. This in fact means that the governing PDE is fulfilled at  $t = 0$ , see e.g. [57]. Thus, the compatibility condition – following the derivation of (MP) – allows us to put  $t = 0$  in (MP) to get*

$$m''(0) + \sigma^\Gamma \alpha(0) m_0 = (\mathbf{g}(\mathbf{E}_0) \cdot \boldsymbol{\nu}, 1)_\Gamma.$$

*This implies the relation*

$$\alpha_0 := \alpha(0) = \frac{(\mathbf{g}(\mathbf{E}_0) \cdot \boldsymbol{\nu}, 1)_\Gamma - m''(0)}{\sigma^\Gamma m_0}$$

*tacitly assuming that  $m_0 = m(0) \neq 0$ .*

The system (DP*i*) and (DMP*i*) represents a linear decoupled scheme. The algorithm for finding a solution follows the scheme

---

<sup>5</sup>If the upper summation bound in a sum is less than the lower summation bound, the sum is empty and it vanishes per definition.

---

1. set  $i = 1$ ,
2. solve the discrete measured problem (DMP $i$ ) to get  $\alpha_i$ ,
3. then solve (DP $i$ ) to get  $e_i$ ,
4. increase the index  $i$  by one and jump to step 2.

### 3.3 A priori estimates

**Lemma 3.3.1.** *Let  $\mathbf{f}$  and  $\mathbf{g}$  be global Lipschitz continuous functions. Suppose that  $\mathbf{E}_0 \in \mathbf{L}^2(\Omega)$ ,  $\mathbf{V}_0 \in \mathbf{L}^2(\Omega)$ ,  $0 < \sigma_* \leq \tilde{\sigma} \leq \sigma^*$ ,  $0 < \mu_* \leq \mu \leq \mu^*$ ,  $m \in C^2([0, T])$  and  $m(0) \neq 0$ . Assume the validity of (MP) at  $t = 0$ . There exists  $\tau_0 > 0$  such that for any  $0 < \tau \leq \tau_0$  and any  $i = 1, \dots, n$ , there exists a unique couple  $\{e_i, \alpha_i\} \in \{\mathbf{H}_0(\mathbf{curl}; \Omega), \mathbb{R}\}$ , which solves (DP $i$ ) and (DMP $i$ ).*

*Proof.* For a given  $e_0, \dots, e_{i-1} \in \mathbf{L}^2(\Omega)$  and  $\alpha_0, \dots, \alpha_{i-1} \in \mathbb{R}$  the  $\alpha_i$  can be eliminated from (DMP $i$ ). The relation

$$m_i'' + \sigma^\Gamma \left[ \sum_{k=0}^i \alpha_{i-k} m_k' \tau + \alpha_i m_0 \right] = (\mathbf{g}(e_{i-1}) \cdot \boldsymbol{\nu}, 1)_\Gamma$$

can be rewritten as

$$\sigma^\Gamma \alpha_i [m_0 + m_0' \tau] = (\mathbf{g}(e_{i-1}) \cdot \boldsymbol{\nu}, 1)_\Gamma - m_i'' - \sigma^\Gamma \sum_{k=1}^i \alpha_{i-k} m_k' \tau,$$

which implies

$$\alpha_i = \frac{(\mathbf{g}(e_{i-1}) \cdot \boldsymbol{\nu}, 1)_\Gamma - m_i'' - \sigma^\Gamma \sum_{k=1}^i \alpha_{i-k} m_k' \tau}{\sigma^\Gamma [m_0 + m_0' \tau]}. \quad (3.8)$$

That means one can get  $\alpha_i \in \mathbb{R}$  for any sufficiently small time step  $\tau \leq \tau_0$ .

---

Let us rewrite (DP<sub>i</sub>) as

$$\begin{aligned} & \left( \frac{\mathbf{e}_i}{\tau^2}, \boldsymbol{\varphi} \right) + (\alpha_0 \mathbf{e}_i, \tilde{\sigma} \boldsymbol{\varphi}) + \left( \frac{1}{\mu} \nabla \times \mathbf{e}_i, \nabla \times \boldsymbol{\varphi} \right) \\ &= (\mathbf{f}(\mathbf{e}_{i-1}), \nabla \times \boldsymbol{\varphi}) + (\mathbf{g}(\mathbf{e}_{i-1}), \boldsymbol{\varphi}) + \left( \frac{\mathbf{e}_{i-1}}{\tau^2} + \frac{\delta \mathbf{e}_{i-1}}{\tau}, \boldsymbol{\varphi} \right) \\ & \quad - \left( \alpha_i \mathbf{e}_0 - \alpha_0 \mathbf{e}_{i-1} + \sum_{k=0}^{i-1} \alpha_{i-k} \delta \mathbf{e}_k \tau, \tilde{\sigma} \boldsymbol{\varphi} \right). \end{aligned}$$

The left-hand-side (LHS) represents an elliptic continuous bilinear form in the space  $\mathbf{H}_0(\mathbf{curl}; \Omega)$ . The right-hand-side (RHS) is a linear bounded functional on  $\mathbf{H}_0(\mathbf{curl}; \Omega)$ . We apply the Lax-Milgram lemma to get the existence of a unique  $\mathbf{e}_i \in \mathbf{H}_0(\mathbf{curl}; \Omega)$ .  $\square$

**Lemma 3.3.2.** *Let  $\mathbf{f}$  and  $\mathbf{g}$  be global Lipschitz continuous functions. Assume that  $|\mathbf{g}(\cdot)| \leq C$ ,  $\mathbf{E}_0 \in \mathbf{H}_0(\mathbf{curl}; \Omega)$ ,  $\mathbf{V}_0 \in \mathbf{L}^2(\Omega)$ ,  $0 < \sigma_* \leq \tilde{\sigma} \leq \sigma^*$ ,  $0 < \mu_* \leq \mu \leq \mu^*$ ,  $m \in C^2([0, T])$  and  $m(0) \neq 0$ . Moreover, assume the validity of (MP) at  $t = 0$ .*

*Then there exist a positive constant  $C$  and  $\tau_0$  such that for any  $0 < \tau \leq \tau_0$  we have*

$$(i) \quad \max_{1 \leq j \leq n} |\alpha_j| \leq C,$$

$$(ii) \quad \max_{1 \leq j \leq n} \|\delta \mathbf{e}_j\|^2 + \max_{1 \leq j \leq n} \|\nabla \times \mathbf{e}_j\|^2 + \sum_{i=1}^n \|\delta \mathbf{e}_i - \delta \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^n \|\nabla \times \mathbf{e}_i - \nabla \times \mathbf{e}_{i-1}\|^2 \leq C,$$

$$(iii) \quad \max_{1 \leq j \leq n} \|\delta^2 \mathbf{e}_j\|_{\mathbf{H}_0(\mathbf{curl}; \Omega)^*} \leq C.$$

*Proof.* (i) The relation (3.8) implies

$$|\alpha_i| \leq C \left( 1 + \sum_{k=1}^{i-1} |\alpha_k| \tau \right).$$

We employ the Grönwall argument (cf. [5]) to conclude the proof.

---

(ii) Set  $\varphi = \delta \mathbf{e}_i \tau$  in (DP*i*). Then sum (DP*i*) up for  $i = 1, \dots, j$  to obtain

$$\begin{aligned} \sum_{i=1}^j (\delta^2 \mathbf{e}_i, \delta \mathbf{e}_i) \tau + \sum_{i=1}^j \left( \frac{1}{\mu} \nabla \times \mathbf{e}_i, \nabla \times \delta \mathbf{e}_i \right) \tau &= \sum_{i=1}^j (\mathbf{f}(\mathbf{e}_{i-1}), \nabla \times \delta \mathbf{e}_i) \tau \\ &+ \sum_{i=1}^j (\mathbf{g}(\mathbf{e}_{i-1}), \delta \mathbf{e}_i) \tau - \sum_{i=1}^j ((\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0, \tilde{\sigma} \delta \mathbf{e}_i) \tau. \end{aligned} \quad (3.9)$$

Let  $\{a_i\}$  be any sequence of real numbers, then the following obvious identity holds true (Abel's summation - see Appendix B)

$$\sum_{i=1}^j a_i (a_i - a_{i-1}) = \frac{1}{2} \left[ a_j^2 - a_0^2 + \sum_{i=1}^j (a_i - a_{i-1})^2 \right].$$

Thus, the LHS of (3.9) can be estimated from below by

$$\sum_{i=1}^j (\delta^2 \mathbf{e}_i, \delta \mathbf{e}_i) \tau = \frac{1}{2} \left[ \|\delta \mathbf{e}_j\|^2 - \|\mathbf{V}_0\|^2 + \sum_{i=1}^j \|\delta \mathbf{e}_i - \delta \mathbf{e}_{i-1}\|^2 \right]$$

and

$$\begin{aligned} \sum_{i=1}^j \left( \frac{1}{\mu} \nabla \times \mathbf{e}_i, \nabla \times \delta \mathbf{e}_i \right) \tau \\ \geq \frac{1}{2\mu^*} \left( \|\nabla \times \mathbf{e}_j\|^2 + \sum_{i=1}^j \|\nabla \times \mathbf{e}_i - \nabla \times \mathbf{e}_{i-1}\|^2 \right) - \frac{1}{2\mu^*} \|\nabla \times \mathbf{e}_0\|^2. \end{aligned}$$

For any real sequences  $\{z_i\}_{i=1}^{\infty}$  and  $\{w_i\}_{i=1}^{\infty}$  the following (summation by parts) identity takes place

$$\sum_{i=1}^j z_i (w_i - w_{i-1}) = z_j w_j - z_0 w_0 - \sum_{i=1}^j (z_i - z_{i-1}) w_{i-1}. \quad (3.10)$$

For the first term on the RHS of (3.9), we use (3.10) together with Cauchy's and

---

Young's inequalities to deduce

$$\begin{aligned}
& \sum_{i=1}^j (\mathbf{f}(\mathbf{e}_{i-1}), \nabla \times \delta \mathbf{e}_i) \tau \\
&= (\mathbf{f}(\mathbf{e}_j), \nabla \times \mathbf{e}_j) - (\mathbf{f}(\mathbf{e}_0), \nabla \times \mathbf{e}_0) - \sum_{i=1}^j (\delta \mathbf{f}(\mathbf{e}_i), \nabla \times \mathbf{e}_i) \tau \\
&\leq \varepsilon \|\nabla \times \mathbf{e}_j\|^2 + C_\varepsilon \|\mathbf{e}_j\|^2 + C_\varepsilon + C \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau + C \sum_{i=1}^j \|\nabla \times \mathbf{e}_i\|^2 \tau \\
&\leq \varepsilon \|\nabla \times \mathbf{e}_j\|^2 + C_\varepsilon + C \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau + C \sum_{i=1}^j \|\nabla \times \mathbf{e}_i\|^2 \tau.
\end{aligned}$$

The last terms on the RHS of (3.9) can be estimated in a similar way as

$$\left| \sum_{i=1}^j (\mathbf{g}(\mathbf{e}_{i-1}), \delta \mathbf{e}_i) \right| \tau \leq C + C \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau$$

and

$$\left| \sum_{i=1}^j ((\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0, \tilde{\sigma} \delta \mathbf{e}_i) \tau \right| \leq C + C \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau.$$

Putting estimates together and fixing a sufficiently small  $\varepsilon > 0$ , we obtain

$$\begin{aligned}
& \|\delta \mathbf{e}_j\|^2 + \|\nabla \times \mathbf{e}_j\|^2 + \sum_{i=1}^j \|\delta \mathbf{e}_i - \delta \mathbf{e}_{i-1}\|^2 + \sum_{i=1}^j \|\nabla \times \mathbf{e}_i - \nabla \times \mathbf{e}_{i-1}\|^2 \\
&\leq C + C \sum_{i=1}^j \|\delta \mathbf{e}_i\|^2 \tau + C \sum_{i=1}^j \|\nabla \times \mathbf{e}_i\|^2 \tau.
\end{aligned}$$

Applying the discrete Grönwall lemma and taking the maximum over  $1 \leq j \leq n$ , we conclude the proof of this part.

(iii) Please note that  $\mathbf{e}_0 \in \mathbf{L}^2(\Omega)$ , Lemma 3.3.2 (ii) together with

$$\mathbf{e}_j = \mathbf{e}_0 + \sum_{i=1}^j \delta \mathbf{e}_i \tau$$

imply

$$\max_{1 \leq j \leq n} \|\mathbf{e}_j\| \leq C.$$


---

Relation (DP $i$ ) implies

$$\begin{aligned} (\delta^2 \mathbf{e}_i, \boldsymbol{\varphi}) &= -((\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0, \tilde{\sigma} \boldsymbol{\varphi}) \\ &\quad - \left( \frac{1}{\mu} \nabla \times \mathbf{e}_i, \nabla \times \boldsymbol{\varphi} \right) + (\mathbf{f}(\mathbf{e}_{i-1}), \nabla \times \boldsymbol{\varphi}) + (\mathbf{g}(\mathbf{e}_{i-1}), \boldsymbol{\varphi}). \end{aligned}$$

A simple calculation yields

$$|(\delta^2 \mathbf{e}_i, \boldsymbol{\varphi})| \leq C \|\boldsymbol{\varphi}\| + C \|\nabla \times \boldsymbol{\varphi}\|.$$

Therefore

$$\|\delta^2 \mathbf{e}_i\|_{\mathbf{H}_0(\mathbf{curl}; \Omega)^*} = \sup_{\substack{\boldsymbol{\varphi} \in \mathbf{H}_0(\mathbf{curl}; \Omega) \\ \|\boldsymbol{\varphi}\|_{\mathbf{H}_0(\mathbf{curl}; \Omega)} \leq 1}} (\delta^2 \mathbf{e}_i, \boldsymbol{\varphi}) \leq C.$$

□

**Lemma 3.3.3.** *Let  $\mathbf{f}$  and  $\mathbf{g}$  be global Lipschitz continuous functions. Suppose that  $|\mathbf{g}(\cdot)| \leq C$ ,  $\mathbf{E}_0 \in \mathbf{H}^1(\Omega)$ ,  $\mathbf{V}_0 \in \mathbf{H}(\operatorname{div}; \Omega)$ ,  $0 < \sigma_* \leq \tilde{\sigma} \leq \sigma^*$ ,  $0 < \mu_* \leq \mu \leq \mu^*$ ,  $\tilde{\sigma} \in H^{1,\infty}(\Omega)$ ,  $m \in C^2([0, T])$  and  $m(0) \neq 0$ . Assume the validity of (MP) at  $t = 0$ . There exist  $C, \tau_0 > 0$  such that for any  $0 < \tau \leq \tau_0$  we have*

$$\max_{1 \leq j \leq n} \|\nabla \cdot \delta \mathbf{e}_j\|^2 + \sum_{i=1}^n \|\nabla \cdot \delta \mathbf{e}_i - \nabla \cdot \delta \mathbf{e}_{i-1}\|^2 \leq C.$$

*Proof.* Lemma 3.3.2 says that  $\mathbf{e}_i \in \mathbf{H}_0(\mathbf{curl}; \Omega)$ , which implies

$\nabla \cdot \mathbf{e}_i \in H^{-1}(\Omega) = (H_0^1(\Omega))^*$ . First show that  $\nabla \cdot \mathbf{e}_i \in L^2(\Omega)$ . Let  $\Phi \in H_0^1(\Omega)$ .

Set  $\boldsymbol{\varphi} = \nabla \Phi$  in (DP $i$ ) to get

$$(\delta^2 \mathbf{e}_i, \nabla \Phi) + ((\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0, \tilde{\sigma} \nabla \Phi) = (\mathbf{g}(\mathbf{e}_{i-1}), \nabla \Phi), \quad (3.11)$$

which can be rewritten as

$$\begin{aligned} \left( \frac{\mathbf{e}_i}{\tau^2}, \nabla \Phi \right) + (\alpha_0 \mathbf{e}_i, \tilde{\sigma} \nabla \Phi) &= (\mathbf{g}(\mathbf{e}_{i-1}), \nabla \Phi) + \left( \frac{\mathbf{e}_{i-1}}{\tau^2} + \frac{\delta \mathbf{e}_{i-1}}{\tau}, \nabla \Phi \right) \\ &\quad - \left( \alpha_i \mathbf{e}_0 - \alpha_0 \mathbf{e}_{i-1} + \sum_{k=0}^{i-1} \alpha_{i-k} \delta \mathbf{e}_k \tau, \tilde{\sigma} \nabla \Phi \right). \end{aligned}$$

Now, we apply the Green theorem and

$$\nabla \cdot (v \mathbf{u}) = v \nabla \cdot \mathbf{u} + \mathbf{u} \cdot \nabla v$$

to obtain

$$\begin{aligned}
& \left( \frac{\nabla \cdot \mathbf{e}_i}{\tau^2}, \Phi \right) + \alpha_0 (\tilde{\sigma} \nabla \cdot \mathbf{e}_i + \mathbf{e}_i \cdot \nabla \tilde{\sigma}, \Phi) \\
&= (\nabla \cdot \mathbf{g}(\mathbf{e}_{i-1}), \Phi) + \left( \frac{\nabla \cdot \mathbf{e}_{i-1}}{\tau^2} + \frac{\nabla \cdot \delta \mathbf{e}_{i-1}}{\tau}, \Phi \right) \\
&\quad - \left( \alpha_i \nabla \cdot \mathbf{e}_0 - \alpha_0 \nabla \cdot \mathbf{e}_{i-1} + \sum_{k=0}^{i-1} \alpha_{i-k} \delta \nabla \cdot \mathbf{e}_k \tau, \tilde{\sigma} \Phi \right) \\
&\quad - \left( \alpha_i \mathbf{e}_0 - \alpha_0 \mathbf{e}_{i-1} + \sum_{k=0}^{i-1} \alpha_{i-k} \delta \mathbf{e}_k \tau, \Phi \nabla \tilde{\sigma} \right).
\end{aligned}$$

Keeping just the terms containing  $\nabla \cdot \mathbf{e}_i$  on the left, we arrive at

$$\begin{aligned}
\left( \left[ \frac{1}{\tau^2} + \alpha_0 \tilde{\sigma} \right] \nabla \cdot \mathbf{e}_i, \Phi \right) &= (\nabla \cdot \mathbf{g}(\mathbf{e}_{i-1}), \Phi) - \alpha_0 (\mathbf{e}_i \cdot \nabla \tilde{\sigma}, \Phi) \\
&\quad + \left( \frac{\nabla \cdot \mathbf{e}_{i-1}}{\tau^2} + \frac{\nabla \cdot \delta \mathbf{e}_{i-1}}{\tau}, \Phi \right) \\
&\quad - \left( \alpha_i \nabla \cdot \mathbf{e}_0 - \alpha_0 \nabla \cdot \mathbf{e}_{i-1} + \sum_{k=0}^{i-1} \alpha_{i-k} \delta \nabla \cdot \mathbf{e}_k \tau, \tilde{\sigma} \Phi \right) \\
&\quad - \left( \alpha_i \mathbf{e}_0 - \alpha_0 \mathbf{e}_{i-1} + \sum_{k=0}^{i-1} \alpha_{i-k} \delta \mathbf{e}_k \tau, \Phi \nabla \tilde{\sigma} \right).
\end{aligned}$$

If  $\mathbf{E}_0 \in \mathbf{H}^1(\Omega)$ ,  $\mathbf{V}_0 \in \mathbf{H}(\operatorname{div}; \Omega)$ ,  $\tilde{\sigma} \in H^{1,\infty}(\Omega)$  then the RHS for  $i = 1$  is a linear bounded functional on  $L^2(\Omega)$ . The space  $H_0^1(\Omega)$  is dense in  $L^2(\Omega)$ . Thus, by the Hahn-Banach theorem we get  $\left[ \frac{1}{\tau^2} + \alpha_0 \tilde{\sigma} \right] \nabla \cdot \mathbf{e}_1 \in L^2(\Omega)$  and also  $\nabla \cdot \mathbf{e}_1 \in L^2(\Omega)$ . Applying Lemma 3.3.2 we have  $\mathbf{e}_1 \in X_N \subset \mathbf{H}^1(\Omega)$ , according to [3]. Repeating this bootstrap argument for increasing  $i$  we conclude that  $\mathbf{e}_i \in \mathbf{H}^1(\Omega)$  for  $i = 1, \dots, n$ .

The next goal is to get stability estimates for  $\nabla \cdot \mathbf{e}_i$ . Applying Green's theorem in (3.11) results in

$$\begin{aligned}
(\delta^2 \nabla \cdot \mathbf{e}_i, \Phi) + ((\alpha * \delta \nabla \cdot \mathbf{e})_i + \alpha_i \nabla \cdot \mathbf{e}_0, \tilde{\sigma} \Phi) &= (\nabla \cdot \mathbf{g}(\mathbf{e}_{i-1}), \Phi) \\
&\quad - ((\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0, \Phi \nabla \tilde{\sigma}).
\end{aligned}$$

This relation is valid for any  $\Phi \in L^2(\Omega)$ , which follows from the density argument  $\overline{H_0^1(\Omega)} = L^2(\Omega)$ .

---

Now, set  $\Phi = \delta \nabla \cdot \mathbf{e}_i \tau$  and sum up for  $i = 1, \dots, j$  to get

$$\begin{aligned}
& \sum_{i=1}^j (\delta^2 \nabla \cdot \mathbf{e}_i, \delta \nabla \cdot \mathbf{e}_i) \tau \\
&= \sum_{i=1}^j (\nabla \cdot \mathbf{g}(\mathbf{e}_{i-1}), \nabla \cdot \delta \mathbf{e}_i) \tau - \sum_{i=1}^j ((\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0, \nabla \cdot \delta \mathbf{e}_i \nabla \bar{\sigma}) \tau \\
&\quad - \sum_{i=1}^j ((\alpha * \delta \nabla \cdot \mathbf{e})_i + \alpha_i \nabla \cdot \mathbf{e}_0, \bar{\sigma} \nabla \cdot \delta \mathbf{e}_i) \tau. \tag{3.12}
\end{aligned}$$

Abel's summation applied to the LHS gives

$$\begin{aligned}
& \sum_{i=1}^j (\delta^2 \nabla \cdot \mathbf{e}_i, \delta \nabla \cdot \mathbf{e}_i) \tau \\
&= \frac{1}{2} \left[ \|\nabla \cdot \delta \mathbf{e}_j\|^2 - \|\nabla \cdot \mathbf{V}_0\|^2 + \sum_{i=1}^j \|\nabla \cdot \delta \mathbf{e}_i - \nabla \cdot \delta \mathbf{e}_{i-1}\|^2 \right].
\end{aligned}$$

The terms on the RHS of (3.12) can be estimated as follows

$$\begin{aligned}
& \left| \sum_{i=1}^j (\nabla \cdot \mathbf{g}(\mathbf{e}_{i-1}), \nabla \cdot \delta \mathbf{e}_i) \tau \right| \\
&\leq \sum_{i=1}^j \|\nabla \cdot \mathbf{g}(\mathbf{e}_{i-1})\| \|\nabla \cdot \delta \mathbf{e}_i\| \tau \\
&\leq C \sum_{i=1}^j \|\mathbf{e}_{i-1}\|_{\mathbf{H}^1(\Omega)} \|\nabla \cdot \delta \mathbf{e}_i\| \tau \\
&\stackrel{[3]}{\leq} C \sum_{i=1}^j \|\mathbf{e}_{i-1}\|_{X_N} \|\nabla \cdot \delta \mathbf{e}_i\| \tau \\
&\stackrel{\text{Lemma 3.3.2}}{\leq} C \sum_{i=1}^j (1 + \|\nabla \cdot \mathbf{e}_{i-1}\|) \|\nabla \cdot \delta \mathbf{e}_i\| \tau \\
&\leq C + C \sum_{i=1}^j \|\nabla \cdot \delta \mathbf{e}_i\|^2 \tau,
\end{aligned}$$


---



$$\begin{aligned}
& \left| \sum_{i=1}^j ((\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0, \nabla \cdot \delta \mathbf{e}_i \nabla \tilde{\sigma}) \tau \right| \\
& \leq C \sum_{i=1}^j \|(\alpha * \delta \mathbf{e})_i + \alpha_i \mathbf{e}_0\| \|\nabla \cdot \delta \mathbf{e}_i \nabla \tilde{\sigma}\| \tau \\
& \stackrel{\text{Lemma 3.3.2}}{\leq} C \sum_{i=1}^j \|\nabla \cdot \delta \mathbf{e}_i\| \tau \\
& \leq C + C \sum_{i=1}^j \|\nabla \cdot \delta \mathbf{e}_i\|^2 \tau,
\end{aligned}$$

and

$$\begin{aligned}
& \left| \sum_{i=1}^j ((\alpha * \delta \nabla \cdot \mathbf{e})_i + \alpha_i \nabla \cdot \mathbf{e}_0, \tilde{\sigma} \nabla \cdot \delta \mathbf{e}_i) \tau \right| \\
& \leq \sum_{i=1}^j \|(\alpha * \delta \nabla \cdot \mathbf{e})_i + \alpha_i \nabla \cdot \mathbf{e}_0\| \|\tilde{\sigma} \nabla \cdot \delta \mathbf{e}_i\| \tau \\
& \stackrel{\text{Lemma 3.3.2}}{\leq} C + C \sum_{i=1}^j \|\nabla \cdot \delta \mathbf{e}_i\|^2 \tau.
\end{aligned}$$

Putting things together we arrive at

$$\|\nabla \cdot \delta \mathbf{e}_j\|^2 + \sum_{i=1}^j \|\nabla \cdot \delta \mathbf{e}_i - \nabla \cdot \delta \mathbf{e}_{i-1}\|^2 \leq C \left( 1 + \sum_{i=1}^j \|\nabla \cdot \delta \mathbf{e}_i\|^2 \tau \right).$$

The rest of the proof follows from Grönwall's argument.  $\square$

### 3.4 Existence of a solution

Let us introduce the following piece-wise linear in time function, where  $0 \leq i \leq n$ ,

$$\mathbf{E}_n : [0, T] \rightarrow \mathbf{L}^2(\Omega) : t \mapsto \begin{cases} \mathbf{E}_0 & t = 0 \\ \mathbf{e}_{i-1} + (t - t_{i-1}) \delta \mathbf{e}_i & t \in (t_{i-1}, t_i] \end{cases},$$


---

and the piece-wise constant function

$$\overline{\mathbf{E}}_n : [0, T] \rightarrow \mathbf{L}^2(\Omega) : t \mapsto \begin{cases} \mathbf{E}_0 & t = 0 \\ \mathbf{e}_i & t \in (t_{i-1}, t_i] \end{cases}, \quad 0 \leq i \leq n.$$

Analogously the iterates for  $\mathbf{E}_t$  are defined

$$\mathbf{V}_n : [0, T] \rightarrow \mathbf{L}^2(\Omega) : t \mapsto \begin{cases} \mathbf{V}_0 & t = 0 \\ \delta \mathbf{e}_{i-1} + (t - t_{i-1})\delta^2 \mathbf{e}_i & t \in (t_{i-1}, t_i] \end{cases},$$

and

$$\overline{\mathbf{V}}_n : [0, T] \rightarrow \mathbf{L}^2(\Omega) : t \mapsto \begin{cases} \mathbf{V}_0 & t = 0 \\ \delta \mathbf{e}_i & t \in (t_{i-1}, t_i] \end{cases}, \quad 0 \leq i \leq n.$$

Similarly, we define  $\overline{m}'_n$  and  $\overline{m}''_n$ . Using Rothe's functions, it is possible to rewrite (DPi) and (DMPi) on the whole time frame as (for  $t \in (t_{i-1}, t_i]$ )

$$\begin{aligned} & (\partial_t \mathbf{V}_n(t), \boldsymbol{\varphi}) + ((\overline{\alpha}_n * \overline{\mathbf{V}}_n)(t_i) + \overline{\alpha}_n(t) \mathbf{E}_0, \tilde{\sigma} \boldsymbol{\varphi}) + \left( \frac{1}{\mu} \nabla \times \overline{\mathbf{E}}_n(t), \nabla \times \boldsymbol{\varphi} \right) \\ & = (\mathbf{f}(\overline{\mathbf{E}}_n(t - \tau)), \nabla \times \boldsymbol{\varphi}) + (\mathbf{g}(\overline{\mathbf{E}}_n(t - \tau)), \boldsymbol{\varphi}) \end{aligned} \quad (\text{DP})$$

and (MP) as

$$\overline{m}''_n(t) + \sigma^\Gamma [(\overline{\alpha}_n * \overline{m}'_n)(t_i) + \overline{\alpha}_n(t) m_0] = (\mathbf{g}(\overline{\mathbf{E}}_n(t - \tau)) \cdot \boldsymbol{\nu}, 1)_\Gamma. \quad (\text{DMP})$$

The objective of this section is to prove the existence of a solution to (P) and (MP), which represents a weak solution to (3.3), (3.4), (3.6) and (3.2).

**Theorem 3.4.1.** *Let  $\Omega \in C^{1,1}$  or  $\Omega$  be convex. Let  $\mathbf{f}$  and  $\mathbf{g}$  be global Lipschitz continuous functions. Suppose  $|\mathbf{g}(\cdot)| \leq C$ ,  $\mathbf{E}_0 \in \mathbf{H}^1(\Omega)$ ,  $\mathbf{V}_0 \in \mathbf{H}(\text{div}; \Omega)$ ,  $0 < \sigma_* \leq \tilde{\sigma} \leq \sigma^*$ ,  $0 < \mu_* \leq \mu \leq \mu^*$ ,  $\tilde{\sigma} \in H^{1,\infty}(\Omega)$ ,  $m \in C^2([0, T])$  and  $m(0) \neq 0$ . Assume the validity of (MP) at  $t = 0$ . Then there exists a weak solution  $\{\mathbf{E}, \alpha\}$  to (P) and (MP). Moreover, it holds that  $\alpha \in L^2((0, T))$ ,  $\mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty((0, T), X_N)$  with  $\partial_t \mathbf{E} \in L^2((0, T), \mathbf{L}^2(\Omega)) \cap C([0, T], X_N^*)$  and  $\mathbf{E}_{tt} \in L^2((0, T), \mathbf{H}_0(\text{curl}; \Omega)^*)$ .*

---

*Proof.* Bounded sets in  $L^2((0, T))$  are weakly closed. Therefore the stability estimate (cf. Lemma 3.3.2)

$$\int_0^T \bar{\alpha}_n^2(t) dt \leq C$$

gives for a subsequence (denoted by the same symbol again)

$$\bar{\alpha}_n \rightharpoonup \alpha \quad \text{in } L^2((0, T)). \quad (3.13)$$

According to [3, Thm. 2.8]  $X_N$  is compactly embedded ( $\Subset$ ) in  $\mathbf{L}^2(\Omega)$ , so the following diagram is valid

$$\begin{array}{ccccc} \mathbf{H}^1(\Omega) & & \mathbf{H}_0(\mathbf{curl}; \Omega) & & \\ \cup & \subset & & \subset & \\ X_N & & \Subset & & \mathbf{L}^2(\Omega) \\ & & & & \Downarrow \\ X_N^* & & \ni & & \mathbf{L}^2(\Omega)^* \\ \cup & \supset & & \supset & \\ \mathbf{H}^1(\Omega)^* & & \mathbf{H}_0(\mathbf{curl}; \Omega)^* & & \end{array}$$

Here the Hilbert space  $\mathbf{L}^2(\Omega)$  is identified with its dual by the Riesz representation theorem.

A priori estimates from Lemmas 3.3.2 and 3.3.3 imply

$$\int_0^T \|\partial_t \mathbf{E}_n(t)\|^2 dt \leq C, \quad \|\mathbf{E}_n(t)\|_{X_N} \leq C, \quad \forall t \in [0, T].$$

Now, Kačur's lemma [56, Lemma 1.3.13] is applied to get existence of a vector field  $\mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty((0, T), X_N)$  with  $\partial_t \mathbf{E} \in L^2([0, T], \mathbf{L}^2(\Omega))$  ( $\mathbf{E}$  is differentiable a.e. in  $[0, T]$ ) and a subsequence of  $\mathbf{E}_n$  (which we denote by the same symbol again) such that

$$\left\{ \begin{array}{ll} \mathbf{E}_n \rightarrow \mathbf{E}, & \text{in } C([0, T], \mathbf{L}^2(\Omega)) \end{array} \right. \quad (3.14a)$$

$$\left\{ \begin{array}{ll} \mathbf{E}_n(t) \rightharpoonup \mathbf{E}(t), & \text{in } X_N, \quad \forall t \in [0, T] \end{array} \right. \quad (3.14b)$$

$$\left\{ \begin{array}{ll} \overline{\mathbf{E}_n}(t) \rightharpoonup \mathbf{E}(t), & \text{in } X_N, \quad \forall t \in [0, T] \end{array} \right. \quad (3.14c)$$

$$\left\{ \begin{array}{ll} \overline{\mathbf{V}_n} = \partial_t \mathbf{E}_n \rightharpoonup \partial_t \mathbf{E}, & \text{in } L^2((0, T), \mathbf{L}^2(\Omega)). \end{array} \right. \quad (3.14d)$$

Lemma 3.3.2 gives the estimate  $\|\partial_t \mathbf{V}_n\|_{X_N^*} \leq C \|\partial_t \mathbf{V}_n\|_{\mathbf{H}_0(\mathbf{curl}; \Omega)^*} \leq C$ . This

together with  $X_N \in \mathbf{L}^2(\Omega) \implies \mathbf{L}^2(\Omega) \in (X_N)^*$  imply

$$\|\mathbf{V}_n\|_{X_N^*} \leq \|\mathbf{V}_n\| \leq C$$

and for  $\varphi \in X_N$  also

$$\begin{aligned} |(\mathbf{V}_n(t) - \mathbf{V}_n(s), \varphi)| &= \left| \int_s^t (\partial_t \mathbf{V}_n, \varphi) \right| \\ &\leq \left| \int_s^t \|\partial_t \mathbf{V}_n\|_{X_N^*} \|\varphi\|_{X_N} \right| \leq C |t - s| \|\varphi\|_{X_N}. \end{aligned}$$

Therefore, the sequence  $\mathbf{V}_n$  must be equi-bounded and equi-continuous in the space  $C([0, T], X_N^*)$ , hence it is compact there.

For  $\varphi \in X_N$  and  $t \in [t_{i-1}, t_i]$  we have

$$|(\overline{\mathbf{V}}_n - \mathbf{V}_n, \varphi)| = \left| \int_t^{t_i} (\partial_t \mathbf{V}_n, \varphi) \right| \leq \mathcal{O}(\tau) \|\varphi\|_{X_N} \xrightarrow{n \rightarrow \infty} 0$$

and we also see that

$$\overline{\mathbf{V}}_n \rightarrow \mathbf{V} \quad \text{in} \quad C([0, T], X_N^*). \quad (3.15)$$

Combining this with (3.14d) we deduce that  $\mathbf{V} = \partial_t \mathbf{E}$ . Lemma 3.3.2 yields  $\partial_t \mathbf{V}_n \in L^2((0, T), \mathbf{H}_0(\mathbf{curl}; \Omega)^*)$  and due to the reflexivity of this space, it is true that (for a subsequence)

$$\partial_t \mathbf{V}_n \rightharpoonup \mathbf{E}_{tt} \quad \text{in} \quad L^2((0, T), \mathbf{H}_0(\mathbf{curl}; \Omega)^*).$$

We are allowed to write for  $\varphi \in X_N$  and  $t \in [t_{i-1}, t_i]$  that

$$\begin{aligned} |(\overline{\alpha}_n * (\overline{\mathbf{V}}_n, \varphi))(t_i) - (\overline{\alpha}_n * (\overline{\mathbf{V}}_n, \varphi))(t)| &\leq \left| \int_t^{t_i} (\overline{\mathbf{V}}_n(t_i - s), \varphi) \overline{\alpha}_n(s) \, ds \right| \\ &+ \left| \int_0^t ((\overline{\mathbf{V}}_n(t_i - s), \varphi) - (\overline{\mathbf{V}}_n(t - s), \varphi)) \overline{\alpha}_n(s) \, ds \right| \leq \mathcal{O}(\tau) \|\varphi\|_{X_N} \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

because of  $\|\partial_t \mathbf{V}_n\|_{X_N^*} + |\overline{\alpha}_n| \leq C$ .

The relations (3.13), (3.15) and the Lebesgue dominated theorem yield for  $\varphi \in X_N$  that

$$(\overline{\alpha}_n * (\overline{\mathbf{V}}_n, \varphi))(t) \rightarrow (\alpha * (\partial_t \mathbf{E}, \varphi))(t) \quad \text{as} \quad n \rightarrow \infty.$$

A simple calculation gives

$$|(\mathbf{g}(\overline{\mathbf{E}}_n(t - \tau)) - \mathbf{g}(\overline{\mathbf{E}}_n(t)), \boldsymbol{\varphi})| \leq C \|\overline{\mathbf{E}}_n(t - \tau) - \overline{\mathbf{E}}_n(t)\| \|\boldsymbol{\varphi}\| \leq \mathcal{O}(\tau) \|\boldsymbol{\varphi}\|$$

and

$$\begin{aligned} |(\mathbf{f}(\overline{\mathbf{E}}_n(t - \tau)) - \overline{\mathbf{E}}_n(t), \nabla \times \boldsymbol{\varphi})| &\leq \\ &\leq C \|\overline{\mathbf{E}}_n(t - \tau) - \overline{\mathbf{E}}_n(t)\| \|\nabla \times \boldsymbol{\varphi}\| \leq \mathcal{O}(\tau) \|\nabla \times \boldsymbol{\varphi}\|. \end{aligned}$$

Now, we integrate the relation (DP) in time over  $t \in [0, \eta] \subset [0, T]$ . Based on the considerations above and we may pass to the limit for  $n \rightarrow \infty$  and  $\boldsymbol{\varphi} \in X_N$  to get

$$\begin{aligned} &(\mathbf{E}_t(\eta) - \mathbf{V}_0, \boldsymbol{\varphi}) + \int_0^\eta (\alpha * \mathbf{E}_t + \alpha \mathbf{E}_0, \tilde{\sigma} \boldsymbol{\varphi}) + \int_0^\eta \left( \frac{1}{\mu} \nabla \times \mathbf{E}, \nabla \times \boldsymbol{\varphi} \right) \\ &= \int_0^\eta (\mathbf{f}(\mathbf{E}), \nabla \times \boldsymbol{\varphi}) + \int_0^\eta (\mathbf{g}(\mathbf{E}), \boldsymbol{\varphi}). \end{aligned}$$

The process is straightforward, therefore we omit further details. Differentiation with respect to  $\eta$  brings us to

$$\begin{aligned} &(\mathbf{E}_{tt}, \boldsymbol{\varphi}) + (\alpha * \mathbf{E}_t + \alpha \mathbf{E}_0, \tilde{\sigma} \boldsymbol{\varphi}) + \left( \frac{1}{\mu} \nabla \times \mathbf{E}, \nabla \times \boldsymbol{\varphi} \right) \\ &= (\mathbf{f}(\mathbf{E}), \nabla \times \boldsymbol{\varphi}) + (\mathbf{g}(\mathbf{E}), \boldsymbol{\varphi}), \end{aligned}$$

which is valid a.e. in  $[0, T]$  and for  $\boldsymbol{\varphi} \in X_N$ . Taking into account the density of  $X_N$  in  $\mathbf{H}_0(\mathbf{curl}; \Omega)$  we conclude the validity of (P) for any  $\boldsymbol{\varphi} \in \mathbf{H}_0(\mathbf{curl}; \Omega)$  and  $\mathbf{E}_{tt}(t) \in (\mathbf{H}_0(\mathbf{curl}; \Omega))^*$  a.e. in  $[0, T]$ .

The next step is to check the validity of (MP). We know that  $|\bar{\alpha}_n| \leq C$  and  $m \in C^2([0, T])$ . Therefore we successively deduce for  $t \in [t_{i-1}, t_i]$  that

$$\begin{aligned} &|(\bar{\alpha}_n * \bar{m}'_n)(t_i) - (\bar{\alpha}_n * \bar{m}'_n)(t)| \\ &\leq \left| \int_t^{t_i} \bar{m}'_n(t_i - s) \bar{\alpha}_n(s) \, ds \right| + \left| \int_0^t (\bar{m}'_n(t_i - s) - \bar{m}'_n(t - s)) \bar{\alpha}_n(s) \, ds \right| \\ &\leq \mathcal{O}(\tau) + C \int_0^t |\bar{m}'_n(t_i - s) - \bar{m}'_n(t - s)| \, ds \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Using (3.13), it is easy to observe that

$$(\bar{\alpha}_n * \bar{m}'_n)(t) \rightarrow (\alpha * m')(t) \quad \text{as } n \rightarrow \infty.$$

Lemma's 3.3.2 and 3.3.3 together with  $X_N \subset \mathbf{H}^1(\Omega)$  give

$$\|\mathbf{E}_n(t)\|_{\mathbf{H}^1(\Omega)} \leq C.$$

From the Nečas inequality [65], [6, (7.116)] or B.0.11 in Appendix.

$$\|z\|_{\Gamma}^2 \leq \varepsilon \|\nabla z\|^2 + C_\varepsilon \|z\|^2, \quad \forall z \in H^1(\Omega), \quad 0 < \varepsilon < \varepsilon_0 \quad (3.16)$$

and the strong convergence in  $C([0, T], \mathbf{L}^2(\Omega))$  cf. (3.14a) follows the strong convergence on the boundary

$$\begin{aligned} \|\mathbf{E}_n(t) - \mathbf{E}(t)\|_{\Gamma}^2 &\leq \varepsilon \|\mathbf{E}_n(t) - \mathbf{E}(t)\|_{\mathbf{H}^1(\Omega)}^2 + C_\varepsilon \|\mathbf{E}_n(t) - \mathbf{E}(t)\|^2 \\ &\leq \varepsilon + C_\varepsilon \|\mathbf{E}_n(t) - \mathbf{E}(t)\|^2 \end{aligned}$$

i.e.

$$\mathbf{E}_n \rightarrow \mathbf{E} \quad \text{in} \quad C([0, T], \mathbf{L}^2(\Gamma)).$$

Analogously it can be deduced that

$$\max_{t \in [0, T]} \|\mathbf{E}_n(t) - \mathbf{E}_n(t - \tau)\|_{\Gamma} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

Now, the relation (DMP) has to be integrated in time over  $t \in [0, \eta] \subset [0, T]$ . Once the considerations above are taken into account, it is possible to pass to the limit for  $n \rightarrow \infty$  to get

$$\int_0^\eta m'' + \int_0^\eta \sigma^\Gamma [\alpha * m' + \alpha m_0] = \int_0^\eta (\mathbf{g}(\mathbf{E}) \cdot \boldsymbol{\nu}, 1)_\Gamma.$$

Differentiation with respect to  $\eta$  confirms the validity of (MP), and we conclude the proof.  $\square$

### 3.5 Uniqueness

The existence of a solution to problems (P) and (MP) is proven in the previous section. In order to guarantee that there exists at most one solution a proof has to be made. Before the proof itself, a handful of estimates are made that are used later in this section.

---

Let  $u : \mathbb{R}^3 \rightarrow \mathbb{R}$  be a differentiable function. Fix vectorial fields  $\mathbf{a}, \mathbf{b} : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$  and define  $v(t) = u(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$ . Since  $v$  is a differentiable function in one variable, the mean value theorem gives

$$v(1) - v(0) = v'(c)$$

for some  $c$  between 0 and 1. But since  $v(1) = u(\mathbf{b})$  and  $v(0) = u(\mathbf{a})$ , computing  $v'(c)$  explicitly gives

$$u(\mathbf{b}) - u(\mathbf{a}) = \nabla u(\mathbf{a} + c(\mathbf{b} - \mathbf{a})) \cdot (\mathbf{b} - \mathbf{a}).$$

If  $u \in C^2$  and  $\mathbf{a}, \mathbf{b} \in \mathbf{H}^{1,\infty}(\Omega)$  it holds that

$$|\partial_{x_i}(u(\mathbf{b}) - u(\mathbf{a}))| \leq C(|\mathbf{b} - \mathbf{a}| + |\partial_{x_i}(\mathbf{b} - \mathbf{a})|).$$

Applying these considerations to each component of a 3-dimensional vector function  $\mathbf{h} = (h_1, h_2, h_3)$  it can be concluded that

$$\|\nabla \times [\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b})]\| \leq C \|\mathbf{b} - \mathbf{a}\|_{\mathbf{H}^1(\Omega)} \quad \text{if } \mathbf{h} \in \mathbf{C}^2; \mathbf{a}, \mathbf{b} \in \mathbf{H}^{1,\infty}(\Omega) \quad (3.17)$$

and

$$\|\nabla \cdot [\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b})]\| \leq C \|\mathbf{b} - \mathbf{a}\|_{\mathbf{H}^1(\Omega)} \quad \text{if } \mathbf{h} \in \mathbf{C}^2; \mathbf{a}, \mathbf{b} \in \mathbf{H}^{1,\infty}(\Omega). \quad (3.18)$$

Now, we address the uniqueness of solution to the inverse problem (P), (MP). We are able to prove it within a class of regular solutions.

**Theorem 3.5.1.** *[uniqueness] Let  $\mathbf{f}, \mathbf{g} \in \mathbf{C}^2$ . There exists at most one solution  $\{\mathbf{E}, \alpha\}$  to the (P), (MP) obeying*

- $\alpha \in L^2((0, T))$ ,
- $\mathbf{E} \in C([0, T], \mathbf{L}^2(\Omega)) \cap L^\infty((0, T), \mathbf{H}^{1,\infty}(\Omega))$ ,
- $\mathbf{E}_t \in L^2((0, T), \mathbf{L}^2(\Omega)) \cap C([0, T], X_N^*)$ ,
- $\mathbf{E}_{tt} \in L^2((0, T), \mathbf{H}_0(\mathbf{curl}; \Omega)^*)$

*Proof.* Having two solutions  $\{\mathbf{E}_1, \alpha_1\}$  and  $\{\mathbf{E}_2, \alpha_2\}$  to the (P), (MP), we denote

$$\mathbf{E} = \mathbf{E}_1 - \mathbf{E}_2, \quad \alpha = \alpha_1 - \alpha_2.$$

We split the proof into a few steps.

---

**Estimate for  $|\alpha|$ :** Subtracting (MP) for both solutions from each other, we readily get

$$|\alpha(t)| \leq C \left( \|\mathbf{E}(t)\|_{\Gamma} + \int_0^t |\alpha| \right).$$

Grönwall's argument implies that

$$\begin{aligned} |\alpha(t)| &\leq C \left( \|\mathbf{E}(t)\|_{\Gamma} + \int_0^t \|\mathbf{E}\|_{\Gamma} \right) \\ &\implies \int_0^t \alpha^2 \leq C \int_0^t \|\mathbf{E}\|_{\Gamma}^2 \leq C \int_0^t \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2. \end{aligned} \quad (3.19)$$

**Estimate for  $\|\mathbf{E}_t\|$  and  $\|\nabla \times \mathbf{E}\|$ :** Taking a difference of corresponding relations (P) for both solutions, we get

$$\begin{aligned} (\mathbf{E}_{tt}, \varphi) + \left( \frac{1}{\mu} \nabla \times \mathbf{E}, \nabla \times \varphi \right) &= (\mathbf{f}(\mathbf{E}_1) - \mathbf{f}(\mathbf{E}_2), \nabla \times \varphi) + \\ &+ (\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2), \varphi) - (\alpha_1 * \mathbf{E}_t + \alpha * \mathbf{E}_{2t} + \alpha \mathbf{E}_0, \tilde{\sigma} \varphi). \end{aligned}$$

Setting  $\varphi = \mathbf{E}_t$  and subsequent integration in time result in

$$\begin{aligned} \frac{1}{2} \|\mathbf{E}_t(\eta)\|^2 + \frac{1}{2} \left( \frac{1}{\mu} \nabla \times \mathbf{E}(\eta), \nabla \times \mathbf{E}(\eta) \right) &= \\ = \int_0^{\eta} (\nabla \times [\mathbf{f}(\mathbf{E}_1) - \mathbf{f}(\mathbf{E}_2)], \mathbf{E}_t) + \int_0^{\eta} (\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2), \mathbf{E}_t) \\ - \int_0^{\eta} (\alpha_1 * \mathbf{E}_t + \alpha * \mathbf{E}_{2t} + \alpha \mathbf{E}_0, \tilde{\sigma} \mathbf{E}_t). \end{aligned}$$

The lower bound of the LHS is

$$\frac{1}{2} \|\mathbf{E}_t(\eta)\|^2 + \frac{1}{2\mu^*} \|\nabla \times \mathbf{E}(\eta)\|^2.$$

It can be easily seen that

$$\begin{aligned} \left| \int_0^{\eta} (\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2), \mathbf{E}_t) \right| &\leq \int_0^{\eta} \|\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2)\| \|\mathbf{E}_t\| \\ &\leq C \int_0^{\eta} \|\mathbf{E}\| \|\mathbf{E}_t\| \leq C \int_0^{\eta} (\|\mathbf{E}_t\|^2 + \|\mathbf{E}\|^2). \end{aligned}$$



We successively deduce that

$$\begin{aligned}
 & \left| \int_0^\eta (\nabla \times [\mathbf{f}(\mathbf{E}_1) - \mathbf{f}(\mathbf{E}_2)], \mathbf{E}_t) \right| \\
 & \leq \int_0^\eta \|\nabla \times [\mathbf{f}(\mathbf{E}_1) - \mathbf{f}(\mathbf{E}_2)]\| \|\mathbf{E}_t\| \\
 & \stackrel{(3.17)}{\leq} C \int_0^\eta \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)} \|\mathbf{E}_t\| \\
 & \leq C \int_0^\eta \left( \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2 + \|\mathbf{E}_t\|^2 \right),
 \end{aligned}$$

$$\begin{aligned}
 & \left| \int_0^\eta (\alpha_1 * \mathbf{E}_t, \tilde{\sigma} \mathbf{E}_t) \right| \\
 & \leq C \int_0^\eta \|\alpha_1 * \mathbf{E}_t\| \|\mathbf{E}_t\| \\
 & \leq \int_0^\eta (|\alpha_1| * \|\mathbf{E}_t\|) \|\mathbf{E}_t\| \\
 & \leq \sqrt{\int_0^\eta (|\alpha_1| * \|\mathbf{E}_t\|)^2} \sqrt{\int_0^\eta \|\mathbf{E}_t\|^2} && \text{Cauchy's inequality} \\
 & \leq \int_0^\eta |\alpha_1| \int_0^\eta \|\mathbf{E}_t\|^2 && \text{Young's inequality} \\
 & \leq C \int_0^\eta \|\mathbf{E}_t\|^2,
 \end{aligned}$$

and

$$\begin{aligned}
 & \left| \int_0^\eta (\alpha * \mathbf{E}_{2t}, \tilde{\sigma} \mathbf{E}_t) \right| \\
 & \leq C \int_0^\eta \|\alpha * \mathbf{E}_{2t}\| \|\mathbf{E}_t\| \\
 & \leq C \int_0^\eta (|\alpha| * \|\mathbf{E}_{2t}\|) \|\mathbf{E}_t\| \\
 & \leq C \sqrt{\int_0^\eta (|\alpha| * \|\mathbf{E}_{2t}\|)^2} \sqrt{\int_0^\eta \|\mathbf{E}_t\|^2} && \text{Cauchy's inequality} \\
 & \leq C \int_0^\eta |\alpha| \sqrt{\int_0^\eta \|\mathbf{E}_{2t}\|^2} \sqrt{\int_0^\eta \|\mathbf{E}_t\|^2} && \text{Young's inequality} \\
 & \leq C \int_0^\eta \left( \alpha^2 + \|\mathbf{E}_t\|^2 \right).
 \end{aligned}$$


---

Putting estimates (within this step) together we arrive at

$$\begin{aligned} \|\mathbf{E}_t(\eta)\|^2 + \|\nabla \times \mathbf{E}(\eta)\|^2 &\leq C \int_0^\eta \left( \alpha^2 + \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2 + \|\mathbf{E}_t\|^2 \right) \\ &\stackrel{(3.19)}{\leq} C \int_0^\eta \left( \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2 + \|\mathbf{E}_t\|^2 \right). \end{aligned} \quad (3.20)$$

**Estimate for  $\|\nabla \cdot \mathbf{E}\|$ :** Taking the difference of the corresponding relations (P) for both solutions and integrating in time for  $\varphi = \nabla \Phi$ ,  $\Phi \in H_0^1(\Omega)$  produces

$$(\mathbf{E}_t, \nabla \Phi) = \left( \int_0^t [\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2)], \nabla \Phi \right) - (\alpha_1 * \mathbf{E} + \alpha * \mathbf{E}_2, \tilde{\sigma} \nabla \Phi).$$

From the Green theorem and

$$\nabla \cdot (v\mathbf{u}) = v \nabla \cdot \mathbf{u} + \mathbf{u} \cdot \nabla v,$$

it follows that

$$\begin{aligned} (\nabla \cdot \mathbf{E}_t, \Phi) &= \left( \int_0^t \nabla \cdot [\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2)], \Phi \right) \\ &\quad - (\alpha_1 * \nabla \cdot \mathbf{E} + \alpha * \nabla \cdot \mathbf{E}_2, \tilde{\sigma} \Phi) - (\alpha_1 * \mathbf{E} + \alpha * \mathbf{E}_2, \Phi \nabla \tilde{\sigma}). \end{aligned}$$

This relation is valid for any  $\Phi \in H_0^1(\Omega)$ , thus by the density argument it holds true for  $\Phi \in L^2(\Omega)$ . If  $\varphi = \nabla \cdot \mathbf{E}$  is set, the integration in time yields

$$\begin{aligned} \frac{1}{2} \|\nabla \cdot \mathbf{E}(\eta)\|^2 &= \int_0^\eta \left( \int_0^t \nabla \cdot [\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2)], \nabla \cdot \mathbf{E}(t) \right) \\ &\quad - \int_0^\eta (\alpha_1 * \nabla \cdot \mathbf{E} + \alpha * \nabla \cdot \mathbf{E}_2, \tilde{\sigma} \nabla \cdot \mathbf{E}) \\ &\quad - \int_0^\eta (\alpha_1 * \mathbf{E} + \alpha * \mathbf{E}_2, \nabla \cdot \mathbf{E} \nabla \tilde{\sigma}). \end{aligned}$$

Consecutive deduction gives

$$\begin{aligned} \int_0^\eta \left( \int_0^t \nabla \cdot [\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2)], \nabla \cdot \mathbf{E}(t) \right) &\leq \int_0^\eta \|\nabla \cdot \mathbf{E}(t)\| \int_0^t \|\nabla \cdot [\mathbf{g}(\mathbf{E}_1) - \mathbf{g}(\mathbf{E}_2)]\| \\ &\stackrel{(3.18)}{\leq} C \int_0^\eta \|\nabla \cdot \mathbf{E}(t)\| \int_0^t \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)} \\ &\leq C \int_0^\eta \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2, \end{aligned}$$

$$\begin{aligned}
& \left| \int_0^\eta (\alpha_1 * \nabla \cdot \mathbf{E}, \tilde{\sigma} \nabla \cdot \mathbf{E}) \right| \\
& \leq C \int_0^\eta \|\alpha_1 * \nabla \cdot \mathbf{E}\| \|\nabla \cdot \mathbf{E}\| \\
& \leq \int_0^\eta (|\alpha_1| * \|\nabla \cdot \mathbf{E}\|) \|\nabla \cdot \mathbf{E}\| \\
& \leq \sqrt{\int_0^\eta (|\alpha_1| * \|\nabla \cdot \mathbf{E}\|)^2} \sqrt{\int_0^\eta \|\nabla \cdot \mathbf{E}\|^2} \quad \text{Cauchy's inequality} \\
& \leq \int_0^\eta |\alpha_1| \int_0^\eta \|\nabla \cdot \mathbf{E}\|^2 \quad \text{Young's inequality} \\
& \leq C \int_0^\eta \|\nabla \cdot \mathbf{E}\|^2,
\end{aligned}$$

and

$$\begin{aligned}
\left| \int_0^\eta (\alpha * \nabla \cdot \mathbf{E}_2, \tilde{\sigma} \nabla \cdot \mathbf{E}) \right| & \leq C \int_0^\eta \|\alpha * \nabla \cdot \mathbf{E}_2\| \|\nabla \cdot \mathbf{E}\| \\
& \leq C \int_0^\eta (|\alpha| * \|\nabla \cdot \mathbf{E}_2\|) \|\nabla \cdot \mathbf{E}\| \\
& \leq C \int_0^\eta \alpha^2 + C \int_0^\eta \|\nabla \cdot \mathbf{E}\|^2.
\end{aligned}$$

It can be similarly obtained that

$$\left| \int_0^\eta (\alpha_1 * \nabla \cdot \mathbf{E}, \mathbf{E} \nabla \tilde{\sigma}) \right| \leq C \int_0^\eta (\|\mathbf{E}\|^2 + \|\nabla \cdot \mathbf{E}\|^2)$$

and

$$\left| \int_0^\eta (\alpha * \nabla \cdot \mathbf{E}_2, \mathbf{E} \nabla \tilde{\sigma}) \right| \leq C \int_0^\eta \alpha^2 + C \int_0^\eta \|\mathbf{E}\|^2.$$

Assembling the estimates (within this step) together we arrive at

$$\|\nabla \cdot \mathbf{E}(\eta)\|^2 \leq C \int_0^\eta \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2 + C \int_0^\eta \alpha^2 \stackrel{(3.19)}{\leq} C \int_0^\eta \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2. \quad (3.21)$$

**Summary :** Putting the relations (3.20) and (3.21) together gives

$$\|\mathbf{E}_t(\eta)\|^2 + \|\nabla \times \mathbf{E}(\eta)\|^2 + \|\nabla \cdot \mathbf{E}(\eta)\|^2 \leq C \int_0^\eta (\|\mathbf{E}_t\|^2 + \|\mathbf{E}\|_{\mathbf{H}^1(\Omega)}^2).$$

Using the embedding  $X_N \subset \mathbf{H}^1(\Omega)$  results in

$$\begin{aligned}
\|\mathbf{E}_t(\eta)\|^2 + \|\nabla \times \mathbf{E}(\eta)\|^2 + \|\nabla \cdot \mathbf{E}(\eta)\|^2 \\
\leq C \int_0^\eta (\|\mathbf{E}_t\|^2 + \|\nabla \times \mathbf{E}\|^2 + \|\nabla \cdot \mathbf{E}\|^2).
\end{aligned}$$

Finally, the application of the Grönwall lemma leads to observation that  $\mathbf{E} = \mathbf{0}$  a.e. in  $\Omega \times (0, T)$ .

□

### 3.6 Numerical computations

The last section of this chapter proposes an illustrative example, which is to support the theoretical outcomes from the previous sections. The main objective of these numerical experiments is to test the convergence of the algorithm.

We consider the following test case: Find  $\{\mathbf{E}(\mathbf{x}, t), \alpha(t)\}$  such that

$$\begin{aligned} \varepsilon \mathbf{E}_{tt} + (\sigma * \mathbf{E})_t + \nabla \times \left( \frac{1}{\mu} \nabla \times \mathbf{E} \right) &= \mathbf{F} + \mathbf{g}(\mathbf{E}) && \text{in } \Omega \times (0, T) \\ \mathbf{E} \times \boldsymbol{\nu} &= \mathbf{0}, && \text{in } \Gamma \times (0, T) \end{aligned} \quad (3.22)$$

when the additional measurement (3.6) is considered.

#### Numerical setting

We set  $\Omega = \{\mathbf{x} = (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 < 1\}$  and  $t \in (0, T)$  along with  $T = 1$ . To preserve the ability to compute and analyze the error of the obtained numerical results, we prescribe the exact solution  $\mathbf{E}$  and the time convolution kernel function  $\alpha(t)$  as

$$\begin{aligned} \mathbf{E}(\mathbf{x}, t) &= e^t \left( x - 1 + |\mathbf{x}|^2, y - 1 + |\mathbf{x}|^2, z - 1 + |\mathbf{x}|^2 \right), \\ \alpha(t) &= t^3 + t + 1. \end{aligned} \quad (3.23)$$

Furthermore, we set

$$\begin{aligned} \varepsilon(\mathbf{x}) &= 1, \\ \tilde{\sigma}(\mathbf{x}) &= 2 - |\mathbf{x}|^2, \\ \mu(\mathbf{x}) &= \frac{1}{x + y + z + 5}, \\ m(t) &= 4\pi e^t. \end{aligned}$$


---

and the nonlinear term  $\mathbf{g}(\mathbf{E}) = \mathbf{E}^2$ , where the square operation was defined component-wise<sup>6</sup>. The vectorial field  $\mathbf{F}$  in (3.22) takes the following form

$\mathbf{F}(\mathbf{x}) = (F_x, F_y, F_z)$ , where

$$\begin{aligned} F_x &= e^t(20 - 6x - 6y) + e^t(-1 + x + |\mathbf{x}|) + \\ &\quad + (-7 + 8e^t - 6t - 3t^2)(2 - |\mathbf{x}|)(-1 + x + |\mathbf{x}|) - e^{2t}(-1 + x + |\mathbf{x}|)^2, \\ F_y &= e^t(20 - 6x - 6z) + e^t(-1 + y + |\mathbf{x}|) + \\ &\quad + (-7 + 8e^t - 6t - 3t^2)(2 - |\mathbf{x}|)(-1 + y + |\mathbf{x}|) - e^{2t}(-1 + y + |\mathbf{x}|)^2, \\ F_z &= e^t(20 - 6y - 6z) + e^t(-1 + z + |\mathbf{x}|) + \\ &\quad + (-7 + 8e^t - 6t - 3t^2)(2 - |\mathbf{x}|)(-1 + z + |\mathbf{x}|) - e^{2t}(-1 + z + |\mathbf{x}|)^2. \end{aligned}$$

One can check that  $\{\mathbf{E}(\mathbf{x}, t), \alpha(t)\}$  defined in (3.23) is truly a solution of the equation (3.22).

In order to obtain a numerical solution, the time interval  $(0, 1)$  is split into 100 subintervals meaning the length of the time step  $\tau$  is 0.01. Consequently the domain – the unit cube – is automatically divided into 4010 cells (tetrahedra) whose diameter may vary in range between 0.206825306587 and 0.332772041371. On every time step we use the Lagrange FEM of second order, which resulted in a system with 18015 DoF.

The approximation of  $\mathbf{E}$  at time  $t = T$  and its error are shown in Figure 3.2 and Figure 3.3, respectively. The quality of reconstruction of the function  $\alpha(t)$  can be seen in Figure 3.4. Figure 3.1 displays the decay of error for various values of  $\tau$ . The linear regression lines through the data points are:

$2.447371 \log_2 \tau - 0.750418$  for the error of approximation of  $\mathbf{E}$  depending on  $\tau$

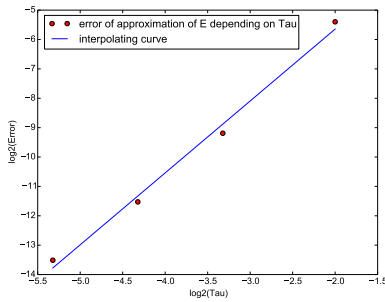
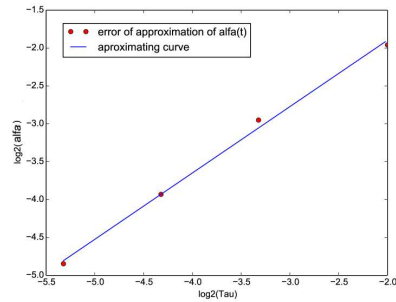
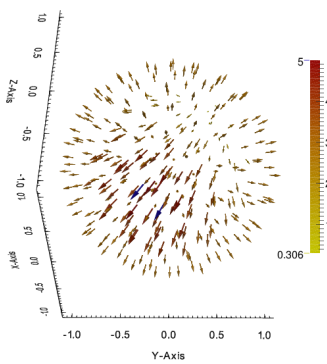
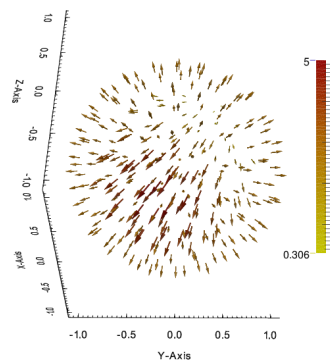
$0.875556 \log_2 \tau - 0.146030$  for the error of approximation of  $\alpha$  depending on  $\tau$ .

**Noisy data** Similarly, as in previous chapter, the ability of the proposed scheme to handle noisy data is tested. The noise is introduced in the same way as described in Section 2.7.2. Figure 3.5 shows how the numerical scheme performs when certain level of inaccuracy is introduced into the measurement process.

---

<sup>6</sup> $\mathbf{E}(\mathbf{x}, t) = (E_1(\mathbf{x}, t), E_2(\mathbf{x}, t), E_3(\mathbf{x}, t))$  and  $\mathbf{E}^2 = (E_1^2(\mathbf{x}, t), E_2^2(\mathbf{x}, t), E_3^2(\mathbf{x}, t))$

---

(a) Decay of error of  $\mathbf{E}$ .(b) Decay of error of  $\alpha$ .Figure 3.1: Decay of error of  $\mathbf{E}$  and  $\alpha(t)$  depending on the size of time step  $\tau$ (a) Exact solution  $\mathbf{E}$ (b) Numerical solution  $\mathbf{E}_{app}$ Figure 3.2: Exact and Numerical solutions of  $\mathbf{E}$  at time  $t = T$

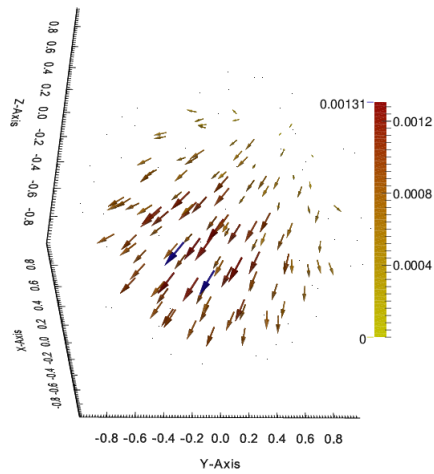


Figure 3.3: Relative error  $\frac{E - E_{app}}{|E|}$  at  $t = T$ . The absolute value of the error is indicated with color

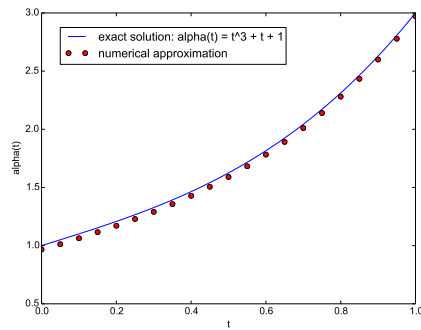


Figure 3.4: Reconstruction of function  $\alpha(t) = t^3 + t + 1$ . For the sake of clarity only every 5<sup>th</sup> time step is plotted.

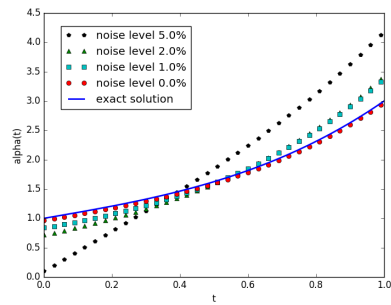


Figure 3.5: Reconstruction of the function  $\alpha(t)$  when the measurement  $m(t)$  contain a certain level of noise

---



# Part II



## Chapter 4

# Model order reduction

In this chapter, we propose a discussion about advantages and drawbacks that may occur when standard computational techniques are used for solving partial differential equations. A new class of techniques called Model Order Reduction is introduced and compared with standard techniques in later sections of this chapter.

Mathematical modelling and numerical experiments together with a new theory development form a core part of many branches of science. Discovering a new theory leads to a set of experiments, which can and usually are supplemented with a number of simulations, in order to save time and/or money. On the other hand, experiments may often inspire scientist who then propose a new theory, that is then validated by simulations. Performing simulations has become an important part of today's (not only) technological world. The relationship between these three feet, the modern science stands on, is illustrated in Figure 4.1. Computer simulations are nowadays performed frequently for many physical, chemical and other processes. The demand for more realistic simulations is ever increasing, which places a heavy burden not only onto the shoulders of researchers. Realistic simulations request the error between computer and physical approximation of some phenomena to be small. Moreover, many aspects of the physical world have to be taken into account. One aspect, which makes more realistic computations possible, is the decreasing price of computational power. However, there is still

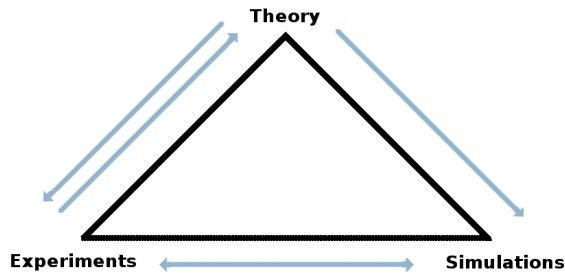


Figure 4.1: Theory - Experiments - Simulations

a lot of space left for mathematicians to develop models which correspond better with reality, and to develop newer, faster and more reliable numerical methods. In fact, the increase in computational power is tied to developments in numerical algorithms. Iterative solution techniques are the main cause of speed-up in algorithms. The main contributions in this area can be found in [10, 11, 42] or [72].

The counter side of having supercomputers available for a wide range of scientists is the decreased need for sophisticated solution methods. It is not uncommon that instead of constructing a special algorithm to solve a certain problem a brute force approach is used (e.g. using a very fine mesh or very small time step in time discretization methods). The question which would be good to know the answer to is whether all the knowledge can be used together. That is to say, the goal is to combine new algorithms and the power of modern computers to get even better results than ever in history and/or get the results in less time.

---

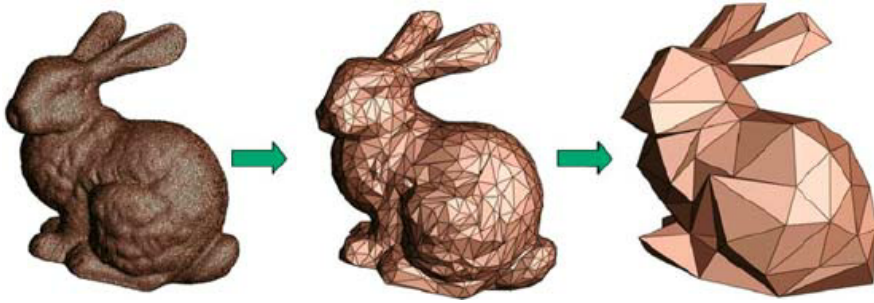


Figure 4.2: The concept of model order reduction

## 4.1 Model Order Reduction

The definition of model order reduction (MOR) depends on the context. A very detailed overview of current state-of-the-art MOR techniques can be found in [74]. It was first developed in the area of systems and control theory in order to reduce the complexity of the models, while their input-output behaviour was required to be preserved as much as possible. This approach has been adopted also by mathematicians, especially after [27] was published. The goal of MOR is to simplify computations to perform simulations within a reasonable amount of time and with limited storage capacity. MOR tries to capture the essential features of the structure of a model. This in fact means that the original model is substituted by a surrogate model of a smaller dimension. Moreover, the process of acquiring the essential parts of the model has to be done automatically. The concept of model order reduction is illustrated in Figure 4.2. The desire for making complicated formulas simpler is obvious throughout the history of mathematics. Just to name a few of the ideas: Fourier tried to approximate a function with a few trigonometric terms, Lanczos looked for a way how to reduce a matrix in tridiagonal form or Arnoldi came up with an idea that a smaller matrix could under some circumstances approximate the original matrix. The techniques based on these ideas have been incorporated into many numerical solvers used today. The integral methods in the field of model order reduction were published in the last two decades of previous century. Among them Moore's Truncated Balanced Realization in 1981, Hankel-norm reduction published by Glover in 1984 or Proper

---

Orthogonal Decomposition proposed by Sirovich in 1987. The last listed method is the subject of our deeper interest in later sections.

## 4.2 Proper Orthogonal Decomposition

The method known as Proper Orthogonal Decomposition (POD) is widely discussed in literature. The original concept was introduced by Pearson [66]. The method is also known under the name Karhunen-Loève decomposition [55, 61], or principal component analysis [46]. The essence of this method is that it gives an optimally ordered, orthonormal basis in the least-squares sense for a given set of theoretical, experimental or computational data [2]. Truncation of this optimal basis then yields the reduced order model (also labeled as surrogate model). The choice of data is obviously crucial. To choose the data properly one has to rely on intuition, simulations or guesswork. Very famous is the method of snapshots [75]. POD has recently been used for solving PDEs [1, 25, 43, 45].

POD is an element of a group of projection methods, where the dynamical system is projected onto a subspace of the original phase space. It can be combined with the Galerkin method [44, 45, 59]. Such a combination gives a powerful tool for deriving surrogate models for high-dimensional dynamical systems. The reason why this approach is so compelling is that the subspace obtained via POD is composed of basis functions that inherit the special characteristics of the solution. The choice of basis functions in standard finite element method is, on the contrary, independent of the system dynamics. One of the main advantages of POD is the fact, that it only uses standard matrix computations, even though it can be applied to very sophisticated problems.

### 4.2.1 POD basis construction

Once the basic idea of proper orthogonal decomposition is known, a more detailed principle of how this method works is proposed. Let's express the above-stated paragraphs in mathematical language. Imagine a vector space  $S$  of dimension  $n$  and a given set of data in the space  $S$ . If the goal is to cope with a dynamical system described by partial differential equations, the data set is in fact the phase

---

space of a differential system, which can be obtained after semi-discretization via a method of lines. For the sake of clarity assume  $S = \mathbb{R}^n$ . The sampled data set  $U = \{u_1(t), \dots, u_m(t)\}$ , where  $u_i(t) \in \mathbb{R}^n$  for  $i = 1, \dots, m$  and  $t \in [0, T]$ , are the trajectories obtained once the method of lines has been applied. The objective is to find a subspace  $S_d \subset S$ , of dimension  $d \ll n$  that approximates the data in some optimal least-square sense. E.g an orthogonal projection  $\Pi_d : S \rightarrow S_d$  is sought, which minimizes the total least-square distance

$$\|U - \Pi_d U\|^2 := \sum_{i=1}^m \int_0^T \|u_i(t) - \Pi_d u_i(t)\|^2 dt.$$

To find solution to this problem, the term *correlation matrix* needs to be introduced first. The correlation matrix  $K \in \mathbb{R}^{n \times n}$  is defined by

$$K = \sum_{i=1}^m \int_0^T u_i(t) u_i(t)^T dt.$$

From the definition it follows that  $K$  is a symmetric positive semi-definite matrix, whose eigenvalues are real, non-negative, and can be ordered as  $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ . Vectors  $\omega_j$  for which

$$K \omega_j = \lambda_j \omega_j, \quad \text{for } i = 1, \dots, n$$

holds, are called the eigenvectors. Thanks to the special structure of matrix  $K$ , they can be chosen as an orthonormal basis of space  $S$ . The subspace  $S_d$  is then given as  $S_d = \text{span}\{\omega_1, \dots, \omega_d\}$ . The vectors  $\omega_1, \dots, \omega_d$  are then used to compute the so-called *POD modes*  $\pi_i$ , which are in fact the columns of the projection matrix  $\Pi = [\pi_1, \dots, \pi_n]$ . To be more specific, the following result is achieved [73]:

**Theorem 4.2.1.** *Let  $K$  be the correlation matrix of the data and let  $\lambda_1 \geq \dots \geq \lambda_n \geq 0$  be the ordered eigenvalues of  $K$ . Then it holds that*

$$\min_{S_d} \|U - \Pi_d U\| = \sum_{j=n-d+1}^n \lambda_j,$$

where the minimum is taken over all subspaces  $S_d$  of dimension  $d$ . Further, the

---

optimal orthogonal projection  $\Pi_d : S \rightarrow S_d$  with  $\Pi_d \Pi_d^T = I$ , is given by

$$\Pi_d = \sum_{j=1}^d \omega_j \omega_j^T.$$

The above-stated result allows us to rewrite every data vector  $u_i(t) \in S$  as

$$u_i(t) = \sum_{j=1}^n c_{ij}(t) \omega_j,$$

where  $c_{ij}(t) = (u_i(t), \omega_j)$ . Then, since  $(\omega_i, \omega_j) = \delta_{ij}$ , it holds that

$$\Pi_d u_i(t) = \sum_{j=1}^d \omega_j \omega_j^T \left( \sum_{l=1}^n c_{il}(t) \omega_l \right) = \sum_{j=1}^d c_{ij}(t) \omega_j.$$

### Dimension of the basis

The question, how to determine the dimension  $d$  has not been answered yet. We only mentioned that  $d \ll n$ , which is natural to demand, because the goal is to get a system that is in some way reduced. The only tool to operate with is the ordered set of the eigenvalues  $\lambda_i$ . For dynamical systems, the larger eigenvalues correspond to the more important characteristics of the system. Therefore *relative information content* was defined in [1] as

$$I(d) = \frac{\sum_{j=1}^d \lambda_j}{\sum_{j=1}^n \lambda_j}.$$

If  $I(d)$  is near to 1 for some  $d$  (e.g.  $I(d) > 0.999$ ), then it is considered enough to take the subspace  $S_d$  of that dimension.

**Remark 4.2.1.** *In order to get  $d$  significantly smaller than  $n$ , the assumption of fast decaying eigenvalues must be made. For some applications (e.g. heat transfer) an exponential decrease can be observed, however, it is not the case for all applications.*

---



### 4.2.2 Method of Snapshots

For problems arising from a real world applications, one has to deal with many degrees of freedom. That means the dimension  $n$  of the phase space is often very large. To calculate the POD modes, it is required to solve a very large eigenvalue problem for a full matrix  $K \in \mathbb{R}^{n \times n}$ , which is a hard if even feasible task. The *method of snapshots* proposed by Sirovich [75] overcomes this problem. Instead of solving the eigensystem for matrix  $K$  of dimension  $n \times n$ , only a smaller matrix of dimension  $m \times m$ , where  $m$  is the number of snapshots, is considered.

Snapshots are built from the trajectories of the dynamical system by evaluating them at certain discrete time instances  $t_1, \dots, t_m \in [0, T]$ . For more detailed discussion, the reader is referred to [44, 45] and the references therein. Snapshots  $u_i = u(t_i) \in \mathbb{R}^n$  form a new correlation matrix  $K$  defined as

$$K = \sum_{i=1}^m u(t_i)u(t_i)^T.$$

The matrix  $U = (u(t_1), \dots, u(t_m)) \in \mathbb{R}^{n \times m}$  is created. The columns consist of the snapshots. That is to say, that the trajectories of the dynamical system can be in this constitution found in each row of the matrix  $U$ . The definition of correlation matrix can be now rewritten to  $K = UU^T$ . The trick method of snapshots proposes is to consider matrix  $U^T U \in \mathbb{R}^{m \times m}$  instead. Solving the eigenvalue problem

$$U^T U \omega_j = \lambda_j \omega_j, \quad j = 1, \dots, m, \quad \omega_j \in \mathbb{R}^m$$

is much easier task to do. Once the orthonormal basis of eigenvectors  $\{\omega_1, \dots, \omega_m\}$  is obtained, the POD modes are then given by

$$\pi_j = \frac{1}{\sqrt{\lambda_j}} U \omega_j, \quad j = 1, \dots, m.$$

The projection matrix  $\Pi$  is then given as

$$\Pi = \left[ \frac{1}{\sqrt{\lambda_1}} U \omega_1, \dots, \frac{1}{\sqrt{\lambda_m}} U \omega_m \right].$$

**Remark 4.2.2.** *The questions which remain unanswered are how many snapshots to take and which time instances to choose. There exists no exact instruction yet; however, the greedy algorithm [21], which chooses the snapshots independently, can be and in practice is very often (e.g. see [37]) used in practice.*

---

## Singular Value Decomposition

There exists a strong link between POD and singular value decomposition. From a standard SVD it is known that for a rectangular matrix  $U \in \mathbb{R}^{n \times m}$  there exist real numbers  $\sigma_1 \geq \dots \geq \sigma_d > 0$  and two unitary matrices  $\Phi \in \mathbb{R}^{n \times n}$  and  $\Psi \in \mathbb{R}^{m \times m}$  such that

$$\Phi^T Y \Psi = \begin{pmatrix} \Sigma_d & 0 \\ 0 & 0 \end{pmatrix} = \Sigma \in \mathbb{R}^{n \times m},$$

where  $\Sigma_d = \text{diag}(\sigma_1, \dots, \sigma_d) \in \mathbb{R}^{d \times d}$ .

The positive numbers  $\sigma_i$  are called singular values of  $U$  and the columns of matrices  $\Phi$  and  $\Psi$  are called left and right singular vectors, respectively. The following holds

$$U\psi_i = \sigma_i\phi_i \quad \text{and} \quad Y^T\phi_i = \sigma_i\psi_i, \quad i = 1, \dots, d.$$

The link between POD and SVD lies in fact that  $\phi_i$  and  $\psi_i$  are in fact the eigenvectors of matrices  $UU^T$  and  $U^TU$ , with eigenvalues  $\lambda_i = \sigma_i^2$  for  $i = 1, \dots, d$ . Thus the POD modes can be very practically determined as

$$\pi_i = \frac{1}{\sigma_i} U\psi_i.$$

## Method of snapshots applied to parabolic inverse problems

This paragraph is based on the technique proposed by [44, 45]. The standard approach how to cope with inverse problems where some parameter (e.g. term source, convolution kernel, diffusion coefficient, etc.) are sought is to use the minimization technique, where a functional  $J$  is defined and minimized in order to identify the missing parameter. In order to perform the minimization, the gradient descent method [64] is usually applied. This method can be applied for a wide variety of problems, even for infinite-dimensional ones. In that case, the search space is typically a function space and Fréchet derivative (Appendix A.0.8) is needed to determine the descent direction. Schematically the minimization algorithm reads as

$$J_{new} = J_{old} - \tau J'_{old},$$


---

where  $J'$  denotes the gradient or the Fréchet derivative and  $\tau$  is the length of the step. This process is repeated until we are satisfied with the new value of  $J$ .

To conduct the minimization procedure, an adjoint problem to the direct problem usually has to be derived and solved. Assume that the unknown parameter is denoted as  $p$ ,  $u$  is the unknown function from the direct problem, and  $v$  is the unknown in the correlating adjoint problem. The scheme to solve such a problem reads as

1. set an initial guess  $p = p_0$
2. until the stopping criteria are met
  - (a) use the value of  $p$  to evaluate  $u$  from the direct problem
  - (b) solve the adjoint problem for  $v$  using the value of  $u$  obtained in the previous step
  - (c) make use of the value  $v$  to determine the new approximation of the sought parameter  $p$
  - (d) check the stopping criteria<sup>1</sup>

The process, if successful — meaning if stopping criteria have been met— ends with function  $u_k$  and parameter  $p_k$  after  $k$  repetitions of the procedure, where both  $u_k$  and  $p_k$  approximate the function  $u$  and the missing parameter  $p$  adequately.

In order to incorporate the POD to the previous scheme, the algorithm needs to be modified as follows

1. define the functional  $J$
2. use an initial guess  $p_0$  as a value of the unknown parameter  $p$
3. do until the stopping criteria<sup>2</sup> are met

---

<sup>1</sup>There are many ways how to define the stopping criteria. A simple example is  $\|p_{new} - p_{old}\| < tolerance$ .

<sup>2</sup>A good way how to define the stopping criterium here is to check the value of the functional  $J$ . Since  $J$  is usually defined as a non-negative functional and it is supposed to be 0 if the exact solution is found, the stopping criterion can be  $J < \varepsilon$ , where  $\varepsilon$  is the tolerance we are willing to accept.

---

- (a) calculate the direct problem to get  $u(t_i)$  for  $i = 1, \dots, n$
- (b) use the solutions  $u_i = u(t_i)$  as a columns of the snapshot matrix  $U = [u_1, \dots, u_n]$
- (c) calculate the projection matrix  $\Pi$  using POD method
- (d) project the direct and adjoint problems to a subspace with smaller dimension to get  $\bar{p}, \bar{u}$  and  $\bar{v}$
- (e) do until stopping criteria<sup>3</sup> are met
  - i. use the value of  $\bar{p} = \Pi p$  to evaluate  $\bar{u}(t_i)$  for  $i = 1, \dots, n$
  - ii. solve the projected adjoint problem to get  $\bar{v}(t_i)$  for  $i = 1, \dots, n$
  - iii. make use of the value  $\bar{v}$  to determine new value of  $\bar{p}$
- (f) project the solutions  $\bar{u}$  and  $\bar{p}$  back to the original full-size dimension to get  $u$  and  $p$ , respectively

It is worth mentioning, that the adjoint problem is solved only on the reduced low-dimensional space, which lessens both time and the computational power. The reader is now referred to the next section, where this algorithm is put into practice, for better illustration.

### 4.3 Numerical Examples

In this section, we propose a couple of examples which are to demonstrate how model order reduction via proper orthogonal decomposition can be put into practice. In further parts of our thesis we assume a real valued function  $u = u(t, x)$  defined on  $[0, T] \times \Omega$ , where  $\Omega \subset \mathbb{R}^n$  and  $\Gamma = \partial\Omega$ .

---

<sup>3</sup>Even here exist more than one way how to define the stopping criterium. For example the relative error between the two last approximations of the parameter  $p$  must be less than tolerance:  $\frac{\|\bar{p}_{new} - \bar{p}_{old}\|}{\|\bar{p}_{old}\|} < \varepsilon$

---

### 4.3.1 Identification of a space-dependent source from final time data

Given the function  $g_N = g_N(x)$  and  $f = f(x)$ , the forward problem

$$\begin{aligned} \partial_t u - \Delta u &= f & \text{in } \Omega \times [0, T] \\ \partial_{\mathbf{n}} u &= g_N & \text{on } \Gamma \times [0, T] \\ u(0) &= u_0 & \text{in } \Omega, \end{aligned} \quad (4.1)$$

defines the mapping

$$f \mapsto [u(T)](f) \equiv u_T(f).$$

The associated optimization (inverse) problem is to find  $f$  such that  $u_T(f)$  approximates a given  $g_T$  as well as possible. This can be formulated as the minimization problem

$$J(f) = \frac{1}{2} \|u_T(f) - g_T\|^2 \rightarrow \min.$$

**Minimization problem** The variational formulation of (4.1) reads as

$$\begin{aligned} (\partial_t u, \varphi) + (\nabla u, \nabla \varphi) &= (g_N, \varphi)_\Gamma + (f, \varphi) \quad \forall \varphi \in H^1(\Omega) \text{ and a.e. in } [0, T] \\ u(0) &= u_0 \quad \text{in } \Omega. \end{aligned}$$

The directional derivative of the functional  $J$  at  $f$  in the direction  $g = g(x)$  is

$$\delta J(f; g) \equiv \lim_{\varepsilon \rightarrow 0} \frac{J(f + \varepsilon g) - J(f)}{\varepsilon} = (\delta u_T(f; g), u_T(f) - g_T),$$

where the variation (directional derivative)  $\delta u(f; g)$  (from now on referred to as  $\delta u$ ) solves the problem

$$\begin{aligned} \partial_t \delta u - \Delta \delta u &= g & \text{in } \Omega \times [0, T] \\ \partial_{\mathbf{n}} \delta u &= 0 & \text{on } \Gamma \times [0, T] \\ \delta u(0) &= 0 & \text{in } \Omega. \end{aligned} \quad (4.2)$$

We use the steepest descent minimization method to solve this problem, so we need Fréchet derivative  $J'(f)g$ , not only the variation  $\delta J(f; g)$ . To this end, the

---

adjoint problem is defined as follows: First multiply the equation (4.2) by  $\partial_t v = \partial_t v(t, x)$  and use integration by parts formula to get

$$\begin{aligned} \int_0^T (\partial_t \delta u - \Delta \delta u, \partial_t v) \, dt &= \int_0^T (\partial_t \delta u, \partial_t v + \Delta v) \, dt + (\nabla \delta u, \nabla v)|_0^T \\ &\quad - \int_0^T (\partial_t \delta u, \partial_{\mathbf{n}} v)_{\partial \Omega} \, dt = \int_0^T (g, \partial_t v) \, dt. \end{aligned}$$

Let  $v$  satisfy the adjoint problem

$$\begin{aligned} \partial_t v + \Delta v &= (u_T - g_T) \quad \text{in } \Omega \times [0, T] \\ \partial_{\mathbf{n}} v &= 0 \quad \text{on } \Gamma \times [0, T] \\ v(T) &= 0 \quad \text{in } \Omega, \end{aligned} \tag{4.3}$$

with its variational formulation

$$\begin{aligned} -(\varphi, \partial_t v) + (\nabla \varphi, \nabla v) &= -(\varphi, u_T - g_T) \quad \forall \varphi \in H^1(\Omega) \text{ and a.e. in } [0, T] \\ v(T) &= 0 \quad \text{in } \Omega. \end{aligned}$$

Then

$$\delta J(f; g) = \int_0^T (\partial_t \delta u, u_T - g_T) \, dt = \int_0^T (g, \partial_t v) \, dt = (g, -v(0)).$$

Therefore,

$$J'(f) = -v(0).$$

The algorithm for the method of steepest descent reads as

1. use the initial guess  $f_0$  as the value of the function  $f$ ,
  2. use  $f$  to calculate  $u$  in (4.1),
  3. employ  $u$  to evaluate  $v$  in (4.3),
  4.  $f_{new} = f + \tau v_k(0)$ ,
  5. update the approximation of  $f := f_{new}$ ,
-

6. check the stopping criteria and repeat the whole procedure if necessary.

where  $\tau > 0$  is a suitable step length and  $v$  is solution to (4.3) for the current approximation of  $u$ , which is in turn the solution of (4.1) when the presently known approximation of the function  $f$  is used.

**Minimization problem combined with POD** Assume (4.1) and its variational formulation. Space approximation for the finite element space  $V^n = \text{span}\{\varphi_1, \dots, \varphi_n\}$  gives

$$u(t, x) = \sum_{i=1}^n u^i(t) \varphi_i(x).$$

Set

$$S_{ij} = (\varphi_i, \varphi_j), \quad M_{ij} = (\nabla \varphi_i, \nabla \varphi_j).$$

Then the space discretized problem can be written as

$$\begin{aligned} Su' + Mu &= Sf, \quad u \in \mathbb{R}^n \\ u(0) &= u_0. \end{aligned}$$

Time discretization:

$$\delta u_i = \frac{u_i - u_{i-1}}{\tau},$$

where  $u_i = u(t_i)$ ,  $t_i = i\tau$ ,  $\tau = T/N$  for  $N \in \mathbb{N}$ . The time discretized system reads as

$$S\delta u_i + Mu_i = Sf, \quad \text{for } i = 1, \dots, N$$

or

$$Au_i = b, \quad \text{for } i = 1, \dots, N, \quad (4.4)$$

where

$$A = S + \tau M \quad \text{and} \quad b = S(\tau f + u_{i-1}).$$

The columns of the snapshot matrix  $U$  according to [45] are formed by either  $u_i$  or  $\delta u_i$ . The correlation matrix is then  $K = \frac{1}{N} U^T U$ . Solving the eigenvalue or singular value problem provides us with the POD modes necessary to build the projection matrix  $\Pi_d$  for a chosen  $d$ . Multiply then (4.4) by  $\Pi_d$  to get

$$\Pi_d A u_i = \Pi_d b$$


---

then multiply the result with identity matrix  $I = \Pi_d^T \Pi_d$  as

$$\Pi_d A \Pi_d^T \Pi_d u_i = \Pi_d b,$$

which results in

$$\bar{A} \bar{u}_i = \bar{b}, \quad (4.5)$$

where  $\bar{A} = \Pi_d A \Pi_d^T$ ,  $\bar{u}_i = \Pi_d u_i$  and  $\bar{b} = \Pi_d b$ . Hence, we have the surrogate problem to solve, where instead of problem (4.4), we only have to handle system (4.5) of much smaller dimension, which can be calculated by hand or in almost no time by computer.

**Remark 4.3.1.** *To get the snapshots, the full-scale problem has to be computed first. Thus, we in fact do not need to calculate the surrogate model, once the original problem has already been computed. However, it is perfectly suitable for the minimization problem where the forward and adjoint problem are calculated repeatedly for changing parameter function  $f_k$ .*

The overall algorithm reads as follows:

1. set  $i = 0$
  2. Calculate the full size problem (4.4) with  $f^i$  to get the snapshot matrix  $U = [u_1^i, \dots, u_N^i]$
  3. Use solutions obtained in the previous step to create projection matrix  $\Pi_d$  and project (4.4) to smaller subspace
  4. Administer the whole minimization procedure within the low dimensional space to get  $\bar{u}_k^i$  and  $\bar{f}_k^i$ .
  5. Project solutions  $\bar{u}_k^i$  and  $\bar{f}_k^i$  back to the high dimensional space as  $u^{i+1} = \Pi_d^T \bar{u}_k^i$  and  $f^{i+1} = \Pi_d^T \bar{f}_k^i$ .
  6. Check stopping criteria. If not fulfilled, increase the index  $i$  by one and repeat the algorithm.
-



	Minimization	Minimization + POD
# of iterations	6	3
total time	459sec	71 sec
$u(T) - u_T$	3.5%	7.2%

Table 4.1: Comparison of a standard minimization technique with POD method

**Numerical results** To test both schemes in practice, we solved (4.1) in the following setting.

We assumed unit square domain  $\Omega = [0, 1] \times [0, 1]$  and the time interval  $[0, T]$ , where  $T = 1$ . The domain  $\Omega$  was split into 2500 subdomains and each of them was split into two triangles by a diagonal cut. To calculate the elliptic problems the automated FEM solver using FEniCS libraries was used. Use of Lagrange elements of first order resulted in a system with 10210 degrees of freedom. The time interval was split into 100 subintervals of equal length. Apart from that,

$$\begin{aligned}
 g_N &= 0, \\
 u(0) &= \text{const}(2), \\
 f_{exact} &= e^{-\frac{(x-0.7)^2 + (y-0.6)^2}{0.6^2}}, \\
 f_{guess} &= \text{const}(1),
 \end{aligned}$$

was set, where the value of  $f_{exact}$  was first used to calculate  $u_i$  for all time instances. Numerical experiments showed following results

Table 4.1 demonstrates that in this setting, the minimization technique with implemented POD method was more than six times faster than the standard minimization technique. The reason for that is that the full-size system with 10210 unknowns was calculated only 3 times instead of 6. The other reason is that the adjoint problem was only calculated on the low-dimensional subspace, where the system with 10210 unknowns was substituted by a system with 4 unknowns. Despite the fact, that only “rough” calculations were made the final approximations of sought function  $f(x)$  gained by both methods are of equal quality and precision.

---

### 4.3.2 Identification of a time-dependent source from a surface measurement

The problem proposed in this example is very similar to the one from Chapter 2. However, for the sake of clarity, it is simpler. In this setting the direct problem is defined as

$$\begin{aligned}\partial_t u - \Delta u &= f(x)h(t) \quad \text{in } \Omega \times [0, T] \\ \partial_{\mathbf{n}} u &= g_N \quad \text{on } \Gamma \times [0, T] \\ u(0) &= u_0 \quad \text{in } \Omega.\end{aligned}\tag{4.6}$$

The measurement is given as

$$m(t) = \int_{\Gamma} u \, ds \tag{4.7}$$

#### Minimization of the functional

If the functional  $J$  that is to be minimized is defined as

$$J(h) = \frac{1}{2} \int_0^T \left[ \int_{\Gamma} u \, ds - m(t) \right]^2 dt,$$

using similar technique as in previous example, one the derivation of the adjoint problem reads as follows.

First, the problem where the variation (directional derivative)  $\delta u = \delta u(h; g)$  is defined as

$$\begin{aligned}\partial_t \delta u - \Delta \delta u &= f(x)g(t) \quad \text{in } \Omega \times [0, T] \\ \partial_{\mathbf{n}} \delta u &= 0 \quad \text{on } \Gamma \times [0, T] \\ \delta u(0) &= 0 \quad \text{in } \Omega.\end{aligned}\tag{4.8}$$

The equation 4.8 is then multiplied by the adjoint variable  $v = v(x, t)$  and the integration by parts in time and space is used to get

$$\int_0^T (\delta u, -\partial_t v - \Delta v) \, dt + (\delta u, v) \Big|_0^T + \int_0^T (\delta u, \partial_{\mathbf{n}} v)_{\partial\Omega} \, dt = \int_0^T g(f, v) \, dt.$$


---

Assume that  $v$  satisfies the problem

$$\begin{aligned} -\partial_t v - \Delta v &= 0 \quad \text{in } \Omega \times [0, T] \\ \partial_{\mathbf{n}} v &= \int_{\Gamma} u \, ds - m(t) \quad \text{on } \Gamma \times [0, T] \\ v(T) &= 0 \quad \text{in } \Omega. \end{aligned} \quad (4.9)$$

Then

$$\begin{aligned} \delta J(h; g) &= \int_0^T \int_{\Gamma} \left[ \int_{\Gamma} u \, ds - m(t) \right] \delta u \, ds \, dt \\ &= \int_0^T \left( \delta u, \left[ \int_{\Gamma} u \, ds - m(t) \right] \right)_{\Gamma} dt \\ &= \int_0^T g(t) (f, v(t)) \, dt \end{aligned} \quad (4.10)$$

thus

$$J'(h) = (f, v)(t).$$

The unmodified minimization technique gives the following algorithm

1. set  $h(t)$  to be an arbitrary initial guess
2. compute the forward problem (4.6) to get  $u$
3. make use of  $u$  to calculate  $v$
4. use  $v$  to calculate the new value  $h_{new}(t) = h(t) - \tau (f, v)(t)$ .

**Adjoint problem II.** There exists another way (and possibly even more) how to derive the adjoint problem, where the problem 4.8 is multiplied by the adjoint variable  $v_t$  and the integration by parts in time and space is done in a different order to get

$$\int_0^T (\partial_t \delta u, -\partial_t v - \Delta v) \, dt + (\delta u, \Delta v) \Big|_0^T + \int_0^T (\partial_t \delta u, \partial_{\mathbf{n}} v_t)_{\partial \Omega} \, dt = \int_0^T g(f, v_t) \, dt.$$


---

If the adjoint problem is then assumed to be

$$\begin{aligned}
 -\partial_t v - \Delta v &= 0 \quad \text{in } \Omega \times [0, T] \\
 \partial_{\mathbf{n}} v &= \int_0^t \left[ \int_{\Gamma} u \, ds - m(t) \right] dt \quad \text{on } \Gamma \times [0, T] \\
 v(T) &= 0 \quad \text{in } \Omega,
 \end{aligned} \tag{4.11}$$

the Fréchet derivative of the functional  $J$  is then defined as

$$J'(h) = (f, v_t)(t). \tag{4.12}$$

The (dis)advantages of this approach are discussed in the latest section 4.3.2 of this chapter.

**Modification of the algorithm using POD** Once the model order reduction via proper orthogonal decomposition method is employed, the above-stated algorithm is modified as follows

1. set  $h(t)$  to be an arbitrary initial guess, set the tolerance  $\varepsilon$
  2. compute the forward problem (4.6) to get  $u(t_i)$  for  $i = 1, \dots, n$
  3. if not  $J(h) < \varepsilon$  then
    - (a) make use of  $u_i = u(t_i)$  as snapshots to get the projection matrix  $\Pi_d$
    - (b) project the forward problem (4.6) and the adjoint problem (4.9) to a subspace of dimension  $d$
    - (c) do
      - i. set  $\bar{h}(t) = h(t)$
      - ii. calculate  $\bar{u}$  for all time instances
      - iii. calculate  $\bar{v}$  for all time instances
      - iv. evaluate  $\bar{h}_{new}(t) = \bar{h}(t) - \tau(f, \bar{v})(t)$
    - (d) until stopping criteria are met
    - (e) project  $\bar{u}$  back to full-dimensional space and assign it to  $u_{new}$
-

### Decoupling scheme

The same problem (4.6) with (4.7) can be approached from a different point of view. It is possible to get a decoupling scheme similarly as in Chapter 2. The application of the measurement operator to the equation (4.6) gives

$$h(t) = \frac{m'(t) - \int_{\Gamma} \Delta u \, ds}{\int_{\Gamma} f \, ds},$$

under assumption that  $\int_{\Gamma} f \, ds \neq 0$ . The time derivative approximated by backward Euler's difference yields

$$u_i - \tau \Delta u_i = \tau h(t_i) f(x) + u_{i-1} \quad (4.13)$$

and

$$h(t_{i+1}) = \frac{\delta m(t_{i+1}) - \int_{\Gamma} \Delta u_i \, ds}{\int_{\Gamma} f \, ds}. \quad (4.14)$$

The overall computational algorithm reads as

1. from the initial value  $u(0)$  calculate value of  $h(t_{i+1})$ ,
2. use  $h(t_{i+1})$  to calculate  $u(t_{i+1})$ ,
3. increase index by one.

**Numerical results** All three methods described above were tested for the following setting. The domain  $\Omega = [0, 1] \times [0, 1]$  was split into 5000 triangles of equal size and shape, which resulted into a linear system with 10210 degrees of freedom. The time domain  $[0, T]$ , where  $T = 1$  was divided into 100 subintervals. For the purpose of demonstration the function  $f$  in (4.6) was set to  $f(x, y) = -4 - x^2 - y^2$  and the initial condition  $u(0) = 2x^2 + 2y^2$ . It is easy to see that if  $h_{exact}(t) = 2e^{-t}$ , then  $u_{exact}(x, y, t) = 2e^{-t}(x^2 + y^2)$  is the solution to problem (4.6). For the method where the functional  $J$  was minimized, the initial guess of function  $h(t) = 0$ . The overview of results of all methods is shown in Table (4.2)

Figure 4.3 demonstrates how these three methods performed.

	Minimization	Minimization + POD	Decoupling scheme
# of iterations	30	7	1
total time	2106 sec	256 sec	324 sec
$\Delta h$	5.3%	7.1%	1.3%
DoF	10210	10210	32743

Table 4.2: Comparison of a standard minimization technique, POD method, and decoupling scheme

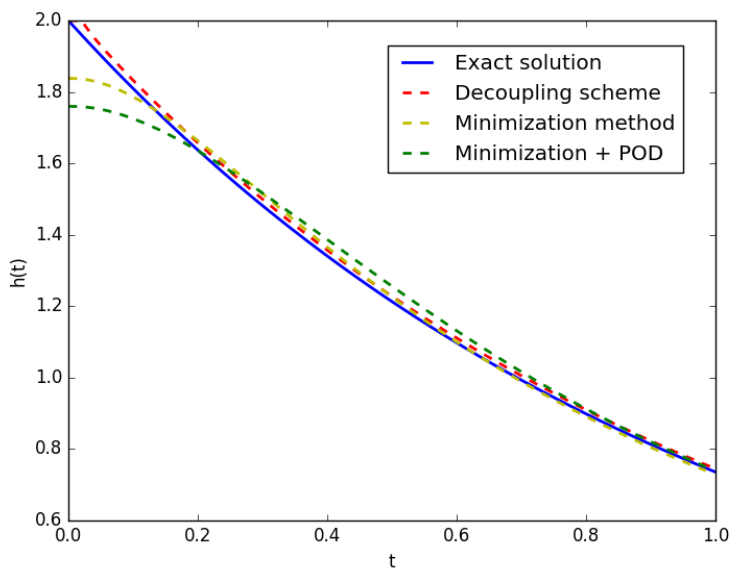


Figure 4.3: Reconstruction of the function  $h(t)$  by all three methods

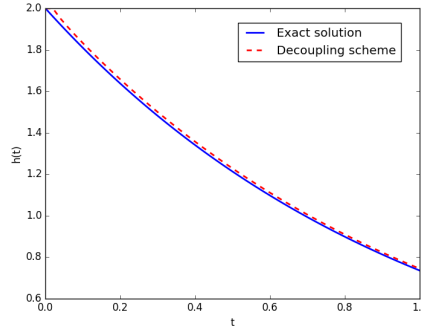


Figure 4.4: Reconstruction of the function  $h(t)$  by the decoupling scheme

**Remark 4.3.2.** *Number of iterations for Minimization + POD method is not the total number of iterations. The number stated in the table depicts the number of outer iterations, which means, how many times the full scale problem was solved.*

**Remark 4.3.3.** *The last row in the table stands for the relative error between exact and calculated solution, where  $\Delta h = \frac{\|h_{exact}(t) - h_{numerical}(t)\|}{\|h_{exact}(t)\|}$ .*

### Observations and conclusion

The decoupling scheme achieved better approximation than the other two methods, which makes it the winner of this imaginary competition. The goal of this comparison is in fact not to say which method is the best, but to highlight the advantages and disadvantages of every method.

From the above-stated table it is clear that the decoupling scheme offers the best approximating ability. (see Figure 4.4) The reason why this method gives such an accurate approximation of a solution is that it tries to approximate the solution on every time layer only once as well as possible. The method starts from time  $t = 0$  and ends for time  $t = T$ . No repetition of this procedure is needed nor possible, thus the number of calculations needed to finish the computational process is known beforehand. There is also no need to initially guess the solution, which is sometimes tricky. The only two things which influence the output is the size of the time step and how well the finite elements approximate the solution

in spatial dimensions. However, the cost for such a good approximation is that by design it requires finite elements of higher order, since in (4.14) the second derivative of  $u$  is computed, it is needed to use at least  $2^{nd}$  order finite elements to approximate the solution. This caused a system with 32743 degrees of freedom had to be calculated for every time layer, which is way more expensive operation than to compute a system with 10210 DoF. Moreover, the proof of convergence can be very likely concluded for highly regular solutions only.

On the other hand the minimization technique is more universal and can be applied to a wide variety of problems. The drawbacks of this approach are twofold. First the adjoint problem has to be derived, which is not always obvious. Second, due to the fact that the number of iterations is not fixed, one has to choose the stopping criteria very carefully to get reasonable results in acceptable time. In our experiments the minimization technique reached the limit of 30 iterations. For this particular example, the main drawback - time consumption - is obvious from the Table 4.2. The other thing which is worth mentioning is that the minimization technique does not give, in this case, a good approximation of the function  $h(t)$  for the whole time interval. This is caused due to the fact that, by definition of (4.9),  $v(T) = 0$ , which means that the correction  $(f, v)(t)$  in

$$h_{new}(t) = h(t) - \tau_k(f, v)(t)$$

decays to zero as  $t \rightarrow T$ . This effect can be seen in Figure 4.5 (b).

The only way to overcome this problem is to use the brute force method where the length of the time interval  $[0, T]$  is doubled tripled and then only the first half or third of the results, respectively, is taken into account. See Figure 4.5 (a). Such an approach increases the computational time inadequately; however, we are not aware of any better method. Deriving the adjoint problem as stated in (4.11) seems to solve the problem with the steady state at the final time  $t = T$  since the derivative of the function  $v$  at the final time  $v_t(T)$  rather than  $v(T)$  is used to compute the Fréchet derivative of the functional  $J(h)$ ; however, the numerical experiments showed the opposite. The solution near the final time  $T$  stays steady and the overall error is even worse. Possibly due to the fact, that the two extra steps – the numerical integration in the boundary condition of (4.11) and the numerical differentiation in (4.12) – introduce more inaccuracy into the process of numerical computation.

---



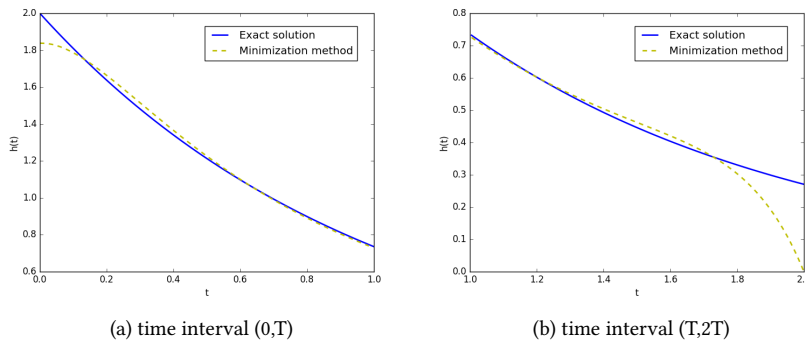


Figure 4.5: Minimization technique requires doubling of the time interval

If the time needed to finish the computations was the only criterium, then the minimization technique combined with proper orthogonal decomposition would be the best choice. This approach was even faster than the decoupling scheme partially due to the fact that only finite elements of 1st order were necessary to use. The use of finite elements of higher order only caused that the relative error between exact and computed solution improved by merely 0.1% (see Figure 4.6) for the cost of quadrupling the computational time, which is obviously not a reasonable compromise. All disadvantages of the standard minimization technique apply to this method as well. Apart from that, this approach seems to have a great potential to deliver good approximation in reasonable time.

The conclusion is that there is still a lot of space for improving and discovering new, better and faster methods for solving PDEs. For example, to find a way how to use the model order reduction techniques and their ability to decrease the time consumption of standard minimization method in combination with the decoupling scheme in order to get even better results in less time remains the subject of our further study.

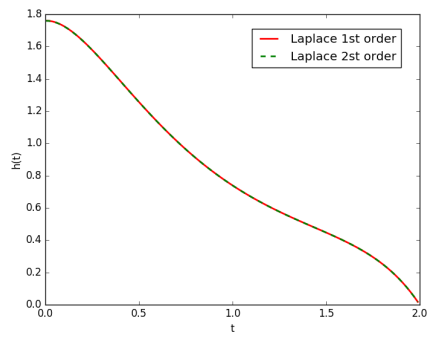


Figure 4.6: Difference between the minimization technique combined with 1st and 2nd order finite elements

---

# Conclusions and future research

This thesis was devoted to the study of advanced numerical methods for inverse problems in evolutionary PDEs. Inverse problems form a dynamically developing prolific field. Their applications are numerous, the problems diverse and the methods employed inevitably various. Particularly in their formulations and analysis, but also in their method of solution, there exists no “general” approach to inverse problems. Nevertheless, the formulations and solving of different inverse problems can give rise to ideas, procedures and observations that are useful from one application to the next.

The first chapter is a brief introduction to the mathematical terms which are used throughout this dissertation. It is crucial for the reader to be familiar with topics stated in that part, because they provide a suitable toolkit for handling equations and expressions occurring in the main part of this work.

The first part, covering Chapter 2 and Chapter 3, studies inverse problems for Maxwell’s equations. The reason why this part is devoted to the mathematical modelling of electromagnetism is the need for better, more detailed, and more realistic models. For example, memory effects occur in almost all electromagnetic phenomena and the models have to be adjusted accordingly. More specifically, a time-dependent source and a time-dependent convolution kernel are sought in Chapter 2 and Chapter 3, respectively. To solve such problems, an additional information is required. One of the highlights of this part is, that in both chapters we have assumed, unlike most authors, only two-dimensional measurements.

Non-invasive measurements correspond better with the physical properties and abilities of modeled phenomena. Theoretical results in both chapters, which include proofs of uniqueness and existence of a solution, have been concluded based on Rothe's method and the arguments of functional analysis. Besides the well-posedness result we have proposed a numerical scheme for finding the solution. The algorithm we have designed is, moreover, non-iterative in the sense that the solution for each time instance is evaluated only once. The results of our work are illustrated with a couple of numerical examples using exact measurements. How to deal with noisy data? In such a case we suggest to regularize the measurements first and then to apply the suggested scheme. It could be interesting to look closer on how exactly the data should be regularized, which opens a possibility for future work.

The second part of this thesis contains Chapter 4, which presents model order reduction techniques as a new set of tools, which allows the current state-of-the-art methods for solving PDEs to be improved in a manner that the computational time is boosted, and the ability to approximate the exact solution is preserved. The aim of this chapter was to test the abilities of the MOR technique, namely the proper orthogonal decomposition method, and to compare it with techniques considered to be standard, in order to raise a discussion about benefits and drawbacks of both approaches. The results and our observations are summarized at the end of this chapter. We suppose the combination of MOR techniques and decoupling scheme proposed in Part I of this dissertation, could bear fruit in achieving even better results. This remains our objective for further studies.

---

# Appendices



## Appendix A

# Differential operators and functions in n-dimensional spaces

**Definition A.0.1.** (The space  $\mathbb{R}^n$ ) Let  $n \in \mathbb{N}$  be the dimension of the space  $\mathbb{R}^n$ . Every element of this space is an  $n$ -tuple of the form  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , which can be uniquely expressed as

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i,$$

where  $x_i$  are called Cartesian coordinates and  $\mathbf{e}_i$  form the standard basis of a real space  $\mathbb{R}^n$ :

$$\mathbf{e}_1 = (1, 0, \dots, 0), \mathbf{e}_2 = (0, 1, \dots, 0), \dots, \mathbf{e}_n = (0, 0, \dots, 1).$$

The Euclidean inner product of two vectors  $\mathbf{x}$  and  $\mathbf{y}$  is defined as

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^n x_i y_i.$$

The cross (vector) product of two vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$  is defined as

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}.$$

**Definition A.0.2.** (Gradient, Divergence, and Rotor operators) Let  $\Omega \subset \mathbb{R}^n$ .

A scalar field  $\phi$  is defined as

$$\phi : \Omega \rightarrow \mathbb{F} : \mathbf{x} \rightarrow \phi(\mathbf{x}).$$

A vector field  $\mathbf{f}$  is defined as

$$\mathbf{f} : \Omega \rightarrow \mathbb{F}^d : \mathbf{x} \rightarrow \mathbf{f}(\mathbf{x}) := f_1(\mathbf{x})\mathbf{e}_1 + \dots + f_d(\mathbf{x})\mathbf{e}_d,$$

where  $\mathbb{F}$  stands for  $\mathbb{R}$  or  $\mathbb{C}$  and  $f_j$  are scalar fields defined on  $\Omega$  for  $j = 1, \dots, d$ .

The gradient is an operator represented by symbol  $\nabla$  and is defined as

$$\nabla = \mathbf{e}_1 \partial_{x_1} + \dots + \mathbf{e}_n \partial_{x_n} = (\partial_{x_1}, \dots, \partial_{x_n}).$$

The divergence operator applied to the vector field  $\mathbf{f}$  is defined as

$$\nabla \cdot \mathbf{f} = \sum_{i=1}^n \partial_{x_i} f_i.$$

The rotor (also called Curl) operator, represented by symbol  $\nabla \times$ , can only act on a 3-dimensional vector field  $\mathbf{f}$  as

$$\nabla \times \mathbf{f} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \partial_{x_1} & \partial_{x_2} & \partial_{x_3} \\ f_1 & f_2 & f_3 \end{vmatrix}.$$

**Definition A.0.3.** (Laplace operator for scalar and vector fields). The Laplace operator  $\Delta$  is defined as

$$\Delta \phi = \nabla^2 \phi = \nabla \cdot \nabla \phi = \sum_{i=1}^n \frac{\partial^2 \phi}{\partial x_i^2}$$

for the scalar field  $\phi$  and as

$$\Delta \mathbf{f} = (\Delta f_1, \dots, \Delta f_d)$$

for the vector field  $\mathbf{f}$ .

---



A list of vector calculus identities according to [28] is stated in Appendix A.

**Definition A.0.4.** A function  $f$  is a mapping from set  $X$  to set  $Y$  which associates every element  $x \in X$  with one and only one element  $y := f(x) \in Y$ . The set  $X$  is usually called the domain of the function  $f$ , whereas the set  $f(X) \subset Y$  is called the range of function  $f$ .

**Definition A.0.5.** Let  $X$  be a set. If there exists a function  $d_X$  defined on  $X$  with the properties

1.  $d_X(x_1, x_2) = d_X(x_2, x_1)$
2.  $d_X(x_1, x_2) \geq 0$  and  $d_X(x_1, x_2) = 0$  iff  $x_1 = x_2$
3.  $d_X(x_1, x_3) \leq d_X(x_1, x_2) + d_X(x_2, x_3)$

holding for every  $x_1, x_2, x_3 \in X$ . Then the pair  $(X, d_X)$  is called a metric space.

**Definition A.0.6.** A subset  $S$  of a metric space  $(X, d_X)$  is bounded if there exist  $x \in X$  and  $r > 0$  such that for all  $s \in S$  holds  $d_X(x, s) < r$ . The whole space  $X$  is called bounded if it is bounded as a subset of itself.

**Definition A.0.7.** (Important properties of functions). A mapping  $f$  between spaces  $(X, d_X)$  and  $(Y, d_Y)$  is

- bounded if  $f(X)$  is a bounded subset of  $Y$ ;
- continuous if

$$(\forall x_2 \in X)(\forall \varepsilon > 0)(\exists \delta(\varepsilon, x_2) > 0) \\ (\forall x_1 \in X : d_X(x_1, x_2) < \delta \Rightarrow d_Y(f(x_1), f(x_2)) < \varepsilon);$$

- Lipschitz continuous if there exists  $L > 0$  such that

$$d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2);$$

- a contraction if the mapping is Lipschitz continuous with  $L < 1$ .

For a more detailed overview of important properties of functions and multi-dimensional spaces, the reader is referred to [9].

**Definition A.0.8.** (*Fréchet derivative*) Let  $V$  and  $W$  be Banach spaces and  $U \subset V$  be an open subset of  $V$ . A function  $f : U \rightarrow W$  is called Fréchet differentiable at  $x \in U$  if there exists a bounded linear operator  $A : V \rightarrow W$  such that

$$\lim_{h \rightarrow 0} \frac{\|f(x+h) - f(x) - Ah\|_W}{\|h\|_V} = 0.$$

## Functional spaces

**Definition A.0.9.** ( $L^p$  space): Let  $p \in [1, \infty)$ . The space  $L^p(\Omega)$  is the set of all measurable functions  $f$  from the bounded domain  $\Omega$  to  $\mathbb{F}$ , for which

$$\|f\|_p = \left( \int_{\Omega} |f|^p dx \right)^{\frac{1}{p}} < \infty$$

**Definition A.0.10.** ( $L^p((0, T), X)$  space): The space  $L^p((0, T), X)$  is defined as a set of all measurable functions  $u(t, x)$  for which

$$\|f\|_{L^p((0, T), X)} = \left( \int_0^T \|u(t)\|_X^p dt \right)^{\frac{1}{p}} < \infty$$

**Definition A.0.11.** ( $H^{k,p}$  space): The space  $H^{k,p}$  is defined as a set of functions  $u \in L^p(\Omega)$  whose generalized derivatives up to the order of  $k$  exist and belong to the space  $L^p(\Omega)$ .

**Definition A.0.12.** (Sobolev vector fields): Let  $\Omega \subset \mathbb{R}^3$  be a bounded domain whose boundary  $\Gamma$  is Lipschitz continuous. Suppose furthermore that  $\mathbf{u} = (u_1, u_2, u_3) \in \mathbf{L}^2(\Omega) := (L^2(\Omega))^3$  and  $\mathbf{v} = (v_1, v_2, v_3) \in \mathbf{L}^2(\Omega)$ . Then the inner product in space  $\mathbf{L}^2(\Omega)$  is defined as

$$(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^3 (u_i, v_i)_{L^2(\Omega)}.$$

The spaces  $\mathbf{H}^1(\Omega)$ ,  $\mathbf{H}(\mathit{curl}, \Omega)$  and  $\mathbf{H}(\mathit{div}, \Omega)$  are defined as

$$\begin{aligned} \mathbf{H}^1(\Omega) &:= \left\{ \mathbf{u} \in \mathbf{L}^2(\Omega) : \nabla \mathbf{u} \in (L^2(\Omega))^{3 \times 3} \right\} \\ \mathbf{H}(\mathit{curl}; \Omega) &:= \left\{ \mathbf{u} \in \mathbf{L}^2(\Omega) : \nabla \times \mathbf{u} \in \mathbf{L}^2(\Omega) \right\} \\ \mathbf{H}(\mathit{div}; \Omega) &:= \left\{ \mathbf{u} \in \mathbf{L}^2(\Omega) : \nabla \cdot \mathbf{u} \in L^2(\Omega) \right\} \end{aligned}$$

The spaces  $\mathbf{H}_0^1(\Omega)$ ,  $\mathbf{H}_0(\mathbf{curl}; \Omega)$  and  $\mathbf{H}_0(\mathbf{div}; \Omega)$  are defined as the closure of  $(C_0^\infty(\Omega))^3$  in the  $\mathbf{H}^1(\Omega)$ ,  $\mathbf{H}(\mathbf{curl}; \Omega)$  and  $\mathbf{H}(\mathbf{div}; \Omega)$  norms, respectively.

---

## Differentiation identities

In the following part,  $\mathbf{f}$  denotes a vector field and  $\phi, \psi$  denote scalar fields. Distributive properties

$$\nabla(\phi + \psi) = \nabla\phi + \nabla\psi$$

$$\nabla \cdot (\mathbf{f} + \mathbf{g}) = \nabla \cdot \mathbf{f} + \nabla \cdot \mathbf{g}$$

$$\nabla \times (\mathbf{f} + \mathbf{g}) = \nabla \times \mathbf{f} + \nabla \times \mathbf{g}$$

Gradient of product

$$\nabla(\phi\psi) = \phi\nabla\psi + \psi\nabla\phi$$

Product of a scalar and a vector

$$\nabla \cdot (\phi\mathbf{f}) = \phi(\nabla \cdot \mathbf{f}) + \mathbf{f} \cdot (\nabla\phi)$$

$$\nabla \times (\phi\mathbf{f}) = \phi(\nabla \times \mathbf{f}) + (\nabla\phi) \times \mathbf{f}$$

Curl of the gradient

$$\nabla \times (\nabla\phi) = \mathbf{0}$$

Divergence of the gradient

$$\nabla^2\phi = \nabla \cdot (\nabla\phi)$$

Rotor of the rotor

$$\nabla \times (\nabla \times \mathbf{f}) = \nabla(\nabla \cdot \mathbf{f}) - \nabla^2\mathbf{f}$$

Divergence of the rotor

$$\nabla \cdot (\nabla \times \mathbf{f}) = 0$$


---

## Integration identities

In the following  $V$  denotes a 3d volume with corresponding 2d boundary  $S = \partial V$  and  $\boldsymbol{\nu}$  stands for normal unit vector. Divergence theorem

$$\oint_{\partial V} \mathbf{f} \, dS = \iiint_V (\nabla \cdot \mathbf{f}) \, dV$$

$$\oint_{\partial V} \phi \, dS = \iiint_V \nabla \phi \, dV$$

$$\oint_{\partial V} (\boldsymbol{\nu} \times \mathbf{f}) \, dS = \iiint_V (\nabla \times \mathbf{f}) \, dV$$

Green's first identity

$$\oint_{\partial V} \psi (\nabla \phi \cdot \boldsymbol{\nu}) \, dS = \iiint_V (\psi \nabla^2 \phi + \nabla \phi \cdot \nabla \psi) \, dV$$

Green's second identity

$$\begin{aligned} \oint_{\partial V} [(\psi \nabla \phi - \phi \nabla \psi) \cdot \boldsymbol{\nu}] \, dS &= \oint_{\partial V} \left[ \psi \frac{\partial \phi}{\partial \boldsymbol{\nu}} - \phi \frac{\partial \psi}{\partial \boldsymbol{\nu}} \right] \, dS \\ &= \iiint_V (\psi \nabla^2 \phi - \phi \nabla^2 \psi) \, dV \end{aligned}$$

Stokes' theorem

$$\oint_{\partial S} \mathbf{f} \cdot d\mathbf{l} = \iint_S (\nabla \times \mathbf{f}) \, dS$$

$$\oint_{\partial S} \phi \, d\mathbf{l} = \iint_S (\boldsymbol{\nu} \times \nabla \phi) \, dS$$



# Appendix B

## Important inequalities

**Lemma B.0.1.** For  $a, b \in \mathbb{R}$ , it holds that

$$(a + b)^2 \leq 2(a^2 + b^2)$$

**Lemma B.0.2. Cauchy-Schwarz inequality - discrete version** Suppose that  $a_1, a_2, \dots, a_n$  and  $b_1, b_2, \dots, b_n$ , for  $n \in \mathbb{N} \cup \{\infty\}$  are two sequences of elements satisfying conditions  $\sum_{j=1}^n |a_j|^2 < \infty$  and  $\sum_{j=1}^n |b_j|^2 < \infty$ . Then

$$\left| \sum_{j=1}^n a_j b_j \right|^2 \leq \left( \sum_{j=1}^n |a_j|^2 \right) \left( \sum_{j=1}^n |b_j|^2 \right).$$

**Lemma B.0.3. Cauchy-Schwarz inequality - continuous version** Let  $\Omega \subset \mathbb{R}^n$  be an open subset. Let  $f : \Omega \rightarrow \mathbb{F}$  and  $g : \Omega \rightarrow \mathbb{F}$  be measurable functions satisfying conditions  $\int_{\Omega} |f(\mathbf{x})|^2 d\mathbf{x} < \infty$  and  $\int_{\Omega} |g(\mathbf{x})|^2 d\mathbf{x} < \infty$ . Then

$$\left| \int_{\Omega} f(\mathbf{x}) \overline{g(\mathbf{x})} d\mathbf{x} \right|^2 \leq \left( \int_{\Omega} |f(\mathbf{x})|^2 d\mathbf{x} \right) \left( \int_{\Omega} |g(\mathbf{x})|^2 d\mathbf{x} \right)$$

**Lemma B.0.4. Hölder inequality - discrete version** Consider  $n \in \mathbb{N} \cup \{\infty\}$ ,  $p \in (1, \infty)$  and  $q$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Assume  $\{a_j\}_{j=1}^n$  and  $\{b_j\}_{j=1}^n$  are sequences

satisfying conditions  $\sum_{j=1}^n |a_j|^p < \infty$  and  $\sum_{j=1}^n |a_j|^q < \infty$ . Then

$$\left| \sum_{j=1}^n a_j b_j \right| \leq \left( \sum_{j=1}^n |a_j|^p \right)^{\frac{1}{p}} \left( \sum_{j=1}^n |b_j|^q \right)^{\frac{1}{q}}.$$

**Lemma B.0.5. Hölder inequality - continuous version** Let  $\Omega$  be an open subset of  $\mathbb{R}^n$ . Assume  $p \in (1, \infty)$  and  $q$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Moreover, let  $f : \Omega \rightarrow \mathbb{F}$  and  $g : \Omega \rightarrow \mathbb{F}$  be measurable functions satisfying conditions  $\int_{\Omega} |f(\mathbf{x})|^p d\mathbf{x} < \infty$  and  $\int_{\Omega} |g(\mathbf{x})|^q d\mathbf{x} < \infty$ . Then

$$\int_{\Omega} |f(\mathbf{x})g(\mathbf{x})| d\mathbf{x} \leq \left( \int_{\Omega} |f(\mathbf{x})|^p d\mathbf{x} \right)^{\frac{1}{p}} \left( \int_{\Omega} |g(\mathbf{x})|^q d\mathbf{x} \right)^{\frac{1}{q}}.$$

**Remark B.0.1.** Cauchy-Schwarz inequality is a special case of Hölder inequality, where  $p = q = 2$ .

**Lemma B.0.6. Young inequality** Assume  $a, b \in \mathbb{R}_0^+$ ,  $p \in (1, \infty)$  and  $q$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

**Lemma B.0.7.  $\varepsilon$ -Young inequality.** Assume  $a, b, p, q$  as in B.0.6. Furthermore, assume  $\varepsilon \in \mathbb{R}_0^+$ . Then

$$ab \leq \varepsilon a^p + C_{\varepsilon} b^q,$$

where  $C_{\varepsilon} = \frac{(\varepsilon p)^{-q/p}}{q}$ .

The following class of inequalities is based on the studies of Grönwall [32] and are of critical importance in process of studying various types of equations.

**Lemma B.0.8. Grönwall inequality - original version** Let  $u : [\alpha, \alpha + h] \rightarrow \mathbb{R}$  be a continuous function which satisfies

$$0 \leq u(t) \leq \int_{\alpha}^t [a + bu(s)] ds, \quad \text{for } t \in [\alpha, \alpha + h],$$

where  $a, b$  are nonnegative constants. Then

$$u(t) \leq ahe^{bh} \quad \text{for } t \in [\alpha, \alpha + h].$$



**Lemma B.0.9. Grönwall inequality - discrete version** Let  $q > 0$  be a real number and  $\{a_i\}_{i=1}^n, \{b_i\}_{i=1}^n$  be two sequences of real nonnegative numbers, where  $n \in \mathbb{N} \cup \{\infty\}$ . If

$$a_n \leq b_n + \sum_{j=1}^{n-1} qa_j, \quad \text{for all } n,$$

then

$$a_n \leq b_n + e^{nq} \sum_{j=1}^{n-1} qb_j, \quad \text{for all } n.$$

**Lemma B.0.10. Grönwall inequality - continuous version** Let  $u(t), a(t)$  and  $b(t)$  be continuous functions from  $I = [\alpha, \beta]$  to  $\mathbb{F}$ . Moreover, let  $b(t)$  be nonnegative in  $I$  and  $a(t)$  be nondecreasing in  $I$ . If

$$u(t) \leq a(t) + \int_{\alpha}^t b(s)u(s)ds, \quad t \in I,$$

then

$$u(t) \leq a(t)e^{\int_{\alpha}^t b(s)ds}, \quad t \in I.$$

**Lemma B.0.11. Nečas inequality** [65] Let  $\Gamma$  be a Lipschitz continuous boundary of  $\Omega$ . Then for all  $u \in H^1(\Omega)$ ,  $0 < \varepsilon < \varepsilon_0$  holds

$$\|u\|_{\Gamma}^2 \leq \varepsilon \|\nabla u\|^2 + C(\varepsilon) \|u\|^2.$$

**Lemma B.0.12. Triangle inequality** Let  $a$  and  $b$  be elements of a vector space  $X$ . Then

$$\|a + b\|_X \leq \|a\|_X + \|b\|_X$$

**Lemma B.0.13. Abel's summation rule** Let  $\{a_i\}_{i=0}^n, n \geq 1$  be a sequence of real numbers. Then

$$2 \sum_{i=1}^n (a_i - a_{i-1})a_i = a_n^2 - a_0^2 + \sum_{i=1}^n (a_i - a_{i-1})^2.$$

If two sequences  $\{w_i\}_{i=0}^n$  and  $\{z_i\}_{i=0}^n$  are considered, then a similar equality rule holds

$$\sum_{i=1}^n (z_i - z_{i-1})w_i = z_n w_n - z_0 w_0 + \sum_{i=1}^n (w_i - w_{i-1})z_{i-1}.$$

This rule is often referred to as the summation by parts formula or discrete per-partes formula.



# Appendix C

## Functional analysis

**Definition C.0.1.** (*Scalar product.*) Suppose  $X$  is a vector space over the complex numbers. A scalar product on  $X$  is a mapping  $(\cdot, \cdot)_X : X \times X \rightarrow \mathbb{C}$  fulfilling following criteria:

1. if  $a \in X$  then  $(a, a)_X = 0$  iff  $a = 0$ ,
2.  $\forall a, b \in X$  holds  $\overline{(a, b)_X} = (b, a)_X$ ,
3.  $\forall a, b, c \in X$  and  $\alpha, \beta \in \mathbb{C}$  holds

$$(\alpha a + \beta b, c)_X = \alpha (a, c)_X + \beta (b, c)_X$$

The norm associated with space  $X$  and scalar product  $(\cdot, \cdot)_X$  is defined as

$$\|\Phi\|_X = \sqrt{(\Phi, \Phi)_X}, \quad \forall \Phi \in X.$$

This norm naturally satisfies the usual triangle inequality (B.0.12).

**Definition C.0.2.** (*Hilbert space.*) Let  $X$  be a vector space equipped with the scalar product  $(\cdot, \cdot)_X$ . If  $X$  is complete with respect to the norm  $\|\cdot\|_X$ , then  $X$  is called a Hilbert space.

**Example 13.** A very basic example of a Hilbert space is the space of square-integrable functions on an open domain  $\Omega \subset \mathbb{R}^3$ , denoted by  $L^2(\Omega)$ . The scalar product is then

defined as

$$(\phi, \psi)_{L^2(\Omega)} = \int_{\Omega} \phi \psi d\Omega$$

The subscript  $L^2(\Omega)$  is often omitted for both norm and scalar product.

**Definition C.0.3.** (Convergent sequence.) A sequence  $\{u_i\}_{i=1}^{\infty}$  is said to be convergent to an element  $u$  of a normed space  $X$  if

$$\lim_{i \rightarrow \infty} \|u - u_i\|_X = 0.$$

This type of convergence is usually denoted as  $u_i \rightarrow u$  and is often called strong convergence or convergence in norm. This happens when it is necessary to distinguish between strong and weak convergence. A sequence  $\{v_i\}_{i=1}^{\infty}$  is said to converge weakly to the element  $v \in X$  if

$$(v_i, \Phi)_X \rightarrow (v, \Phi)_X, \quad \forall \Phi \in X^*,$$

which is in standard notation depicted as  $v_i \rightharpoonup v$ .

The definition of weak convergence is important because, unfortunately, in general bounded sets in Hilbert spaces do not contain convergent subsequences. This only holds true if talking about weak convergence.

**Definition C.0.4.** (Linear operator.) A mapping  $A$  between two Hilbert spaces  $X$  and  $Y$  is often called an operator. An operator  $A : X \rightarrow Y$  is labeled linear if

$$A(\alpha a + \beta b) = \alpha A(a) + \beta A(b),$$

where  $a, b \in X$  and  $\alpha, \beta \in \mathbb{C}$ .

(Bounded operator.) An operator  $A$  is bounded if a constant  $C$  exists such that

$$\|A(\Phi)\|_Y \leq C \|\Phi\|_X, \quad \forall \Phi \in X.$$

Note that if it is clear which symbol stands for an operator and which for an element of a space, the notation  $A(u)$  is abbreviated to  $Au$ .

**Lemma C.0.1.** A linear operator is bounded if and only if it is continuous.

**Definition C.0.5.** The norm (often marked as natural norm) of a bounded linear operator  $A : X \rightarrow Y$  is defined as follows

$$\|A\|_{X \rightarrow Y} = \sup_{u \neq 0, u \in X} \frac{\|A(u)\|_Y}{\|u\|_X}.$$

**Definition C.0.6.** (*Adjoint operator.*) Suppose  $A : X \rightarrow Y$  is an operator between two Hilbert spaces, which is linear and bounded. A unique operator  $A^* : Y \rightarrow X$ , which is called the adjoint operator exists, such that

$$(Ax, y)_Y = (x, A^*y)_X, \quad \forall x \in X, y \in Y.$$

Moreover, this operator is unique.

**Definition C.0.7.** (*Dual space.*) Let  $X$  be a Banach space. The dual space  $X'$  is the space consisting of linear and bounded functionals defined on  $X$ .  $f \in X'$  means  $f : X \rightarrow \mathbb{C}$  is a linear and bounded operator (functional) mapping elements from the space  $X$  to  $\mathbb{C}$ . For every  $f \in X'$ , its norm is defined as

$$\|f\|_{X'} = \sup_{x \in X, x \neq 0} \frac{|f(x)|}{\|x\|_X}.$$

(Dual pairing)  $\langle g, u \rangle_X$  is defined by

$$\langle g, u \rangle_X = g(u), \quad \forall u \in X, g \in X'.$$

**Definition C.0.8.** (*Dual operator.*) The operator  $A^T : Y' \rightarrow X'$  is called dual operator, where  $X'$  and  $Y'$  are the dual spaces to  $X$  and  $Y$ , respectively, when

$$\langle Ax, y \rangle_{Y'} = \langle x, A^T y \rangle_{X'}, \quad \forall x \in X, y \in Y'.$$

**Remark C.0.1.** Note that the terms dual and adjoint operator stand for two separate definitions. Even though they are very often denoted by the same symbol  $*$ .

**Theorem C.0.1. Hahn-Banach theorem.** Let  $f$  be a continuous linear functional defined on a linear subset  $M \subset X$ , where  $X$  is a normed linear space. Then there exists a continuous linear functional  $\tilde{f}$  defined on  $X$  such that

$$f(u) = \tilde{f}(u) \quad \forall u \in M$$

and

$$\|\tilde{f}\|_{X'} = \sup_{x \in X, \|x\|_X \leq 1} \langle \tilde{f}, x \rangle = \sup_{x \in M, \|x\|_X \leq 1} \langle f, x \rangle = \|f\|_{M'}.$$

**Theorem C.0.2. Riesz' representation theorem.** Suppose  $X$  is a Hilbert space. For every  $f \in X'$  a unique  $u \in X$  exists, such that

$$(u, v)_X = f(v), \quad \forall v \in X.$$

Moreover,  $\|u\|_X = \|f\|_{X'}$

**Definition C.0.9.** (*Bilinear form.*) A mapping  $a(.,.) : X \times X \rightarrow \mathbb{C}$  is called a bilinear form on  $X$  if

$$a(\alpha_1 u_1 + \alpha_2 u_2, v) = \alpha_1 a(u_1, v) + \alpha_2 a(u_2, v), \quad \forall u_1, u_2, v \in X, \alpha_1, \alpha_2 \in \mathbb{C},$$

and

$$a(u, \beta_1 v_1 + \beta_2 v_2) = \beta_1 a(u, v_1) + \beta_2 a(u, v_2), \quad \forall u, v_1, v_2 \in X, \beta_1, \beta_2 \in \mathbb{C}$$

A bilinear form is said to be

- symmetric if  $a(u, v) = a(v, u)$  for all  $u, v \in X$
- elliptic if  $\exists C_m > 0 : C_m \|u\|_X^2 \leq a(u, u), \quad \forall u \in X$
- bounded if  $\exists C_M > 0 : |a(u, v)| \leq C_M \|u\|_X \|v\|_X, \quad \forall u, v \in X$

**Theorem C.0.3. Lax-Milgram lemma.** Let  $a(.,.)$  be a elliptic, and continuous bilinear form on a Hilbert space  $X$ . Moreover, let  $f \in X'$ . Then there exists a unique weak solution to the variational formulation given by

$$a(u, v) = f(v), \quad \forall v \in X,$$

and it holds true that

$$\|x\|_X \leq \frac{1}{C_m} \|f\|_{X'}.$$


---

# Index

- Adjoint operator, 139
- Ampère's law, 27
- Bilinear form, 140
- Boundary Conditions, 3
- Bounded operator, 138
- Cauchy-Schwarz inequality, 133
- Cea's lemma, 17
- Convergent sequence, 138
- Coulomb's law, 27
- Degrees of freedom, 18
- Direct problem, 10
- Dirichlet boundary condition, 4
- Discrete Galerkin problem, 16
- Divergence, 126
- Dual operator, 139
- Dual pairing, 139
- Dual space, 139
- Faraday's law, 25
- Finite element method, 18
- Galerkin method, 15
- Gauss' law for magnetism, 29
- Gauss's law, 26
- Grönwall's inequality, 134
- Gradient, 126
- Hölder's inequality, 134
- Hadamard's criteria, 9
- Hahn-Banach theorem, 139
- Hilbert space, 137
- Ill-posed problem, 10
- Initial Conditions, 3
- Inverse problem, 10
- Laplace operator, 126
- Lax-Milgram lemma, 140
- Linear operator, 138
- Maxwell's equations, 23
- Mixed boundary condition, 4
- Model Order reduction, 99
- Neumann boundary condition, 4
- Ohm's law, 30
- Order of PDE, 2
- PDE, 1

- Perfect conductivity on a boundary,  
31
- Proper Orthogonal Decomposition,  
100
- Riesz' representation theorem, 139
- Robin boundary condition, 4
- Rothe's method, 14
- Rotor (Curl), 126
- Scalar field, 126
- Scalar product, 137
- Space  $\mathbb{R}^n$ , 125
- Strong solution, 5
- Triangle inequality, 135
- Vector field, 126
- Weak solution, 6
- Well-posed problem, 9
- Young's inequality, 134
-



# Bibliography

- [1] K. Afanasiev and M. Hinze. Adaptive control of a wake flow using proper orthogonal decomposition. *Lecture Notes in Pure and PPL. Math.*, 216:317–332, 2001.
- [2] V.R. Alghazi and D.J. Sakrison. On the optimality of the Karhunen-Loève expansion. *IEEE Trans. Information Theory*, 15:319–320, 1969.
- [3] C. Amrouche, C. Bernardi, M. Dauge, and V. Girault. Vector potentials in three-dimensional non-smooth domains. *Math. Methods Appl. Sci.*, 21(9):823–864, 1998.
- [4] D.N. Arnold, R.S. Falk, and R. Winther. Finite element exterior calculus: from hodge theory to numerical stability. *Bull. Amer. Math. Soc. (N.S.)*, pages 281–354, 2010.
- [5] D. Bainov and P. Simeonov. *Integral inequalities and applications*, volume 57 of *Mathematics and its applications*. Kluwer Academic Publishers, Dordrecht, 1992.
- [6] C. Baiocchi and A. Capelo. *Variational and quasivariational inequalities. Applications to free boundary problems. (Variatsionnye i kvazivariatsionnye neravenstva. Prilozheniya k zadacham so svobodnoj granitsej)*. Transl. from the English by V. I. Agoshkov, E. Eh. Ovchinnikov and V. P. Shutyaev. Transl. ed. by V. I. Agoshkov. Moskva , 1988.
- [7] H.T. Banks and N.L. Gibson. Well-posedness in maxwell systems with distributions of polarization relaxation parameters. *Applied Mathematics Letters*, 18(4):423 – 430, 2005.

- 
- [8] H.T. Banks and Gabriella A. Pinter. Maxwell-systems with nonlinear polarization. *Nonlinear Analysis: Real World Applications*, 4(3):483 – 501, 2003.
- [9] K. Van Bockstal. *Numerical Techniques for Partial Differential Equations in Superconductivity and Thermoelasticity*. Phd thesis. Ghent university, 2015.
- [10] A. Brandt. Multilevel adaptive solutions to boundary value problems. *Math. Comp.*, 31:333–390, 1977.
- [11] W.L. Briggs, Van Emden Henson, and S.F. McCormick. *A multigrid tutorial*. SIAM, 2000.
- [12] A.L. Bukhgeim and G.V. Dyatlov. Inverse problems for equations with memory. *SIAM J. MATH. FOOL.*, 1(2):1–17, 1998.
- [13] Florian Cajori. The early history of partial differential equations and of partial differentiation and integration. *The American Mathematical Monthly*, 35(9):459–467, 1928.
- [14] J.R. Cannon. Determination of an unknown heat source from overspecified boundary data. *SIAM J. Numer. Anal.*, 5:275–286, 1968.
- [15] M. Cessenat. *Mathematical methods in electromagnetism: linear theory and applications*, volume 41 of *Series on Advances in Mathematics for Applied Sciences*. Singapore: World Scientific Publishers, 1996.
- [16] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.
- [17] F. Colombo and D. Guidetti. A global in time existence and uniqueness result for a semilinear integrodifferential parabolic inverse problem in Sobolev spaces. *Math. Models Methods Appl. Sci.*, 17(4):537–565, 2007.
- [18] F. Colombo and D. Guidetti. Some results on the identification of memory kernels. In M. Ruzhansky and J. Wirth, editors, *Modern aspects of the theory of partial differential equations. Including mainly selected papers based on the presentations at the 7th international ISAAC congress, London, UK, July 13–18, 2009.*, pages 121–138. Operator Theory: Advances and Applications 216. Basel: Birkhäuser, 2011.
-

- 
- [19] F. Colombo, D. Guidetti, and A. Lorenzi. On applications of maximal regularity to inverse problems for integrodifferential equations of parabolic type. Ruiz Goldstein, Gisèle (ed.) et al., Evolution equations. Proceedings of the conference, Blaubeuren, Germany, June 11–17, 2001 in honor of the 60th birthdays of Philippe Bénilan, Jerome A. Goldstein and Rainer Nagel. New York, NY: Marcel Dekker. Lect. Notes Pure Appl. Math. 234, 77-89 (2003), 2003.
- [20] F. Colombo, D. Guidetti, and V. Vespri. Some global in time results for integrodifferential parabolic inverse problems. Favini, Angelo (ed.) et al., Differential equations. Inverse and direct problems. Papers of the meeting, Cortona, Italy, June 21–25, 2004. Boca Raton, FL: CRC Press. Lecture Notes in Pure and Applied Mathematics 251, 35-58 (2006), 2006.
- [21] T.H. Cormen, Ch.E. Leiserson, and R.L. Rivest. *Introduction to Algorithms (1st ed.)*. 1990.
- [22] R. H. De Staelen and M. Slodička. Reconstruction of a convolution kernel in a semilinear parabolic problem based on a global measurement. *Nonlinear Analysis Series A: Theory, Methods & Applications*, 112:43–57, 2015.
- [23] R. H. De Staelen, K. Van Bockstal, and M. Slodička. Error analysis in the reconstruction of a convolution kernel in a semilinear parabolic problem with integral overdetermination. *JCAM*, 275:382–391, 2015.
- [24] A. Erdem, D. Lesnic, and A. Hasanov. Identification of a spacewise dependent heat source. *Applied Mathematical Modelling*, 37(24):10231 – 10244, 2013.
- [25] M. Fahl and E.W. Sachs. Reduced order modelling approaches to PDE-constrained optimization based on proper orthogonal decomposition. *Lect. Notes Comput. Sci. Eng.*, 30:268–280, 2003.
- [26] A. Farcas and D. Lesnic. The boundary-element method for the determination of a heat source dependent on one variable. *J. Eng. Math.*, 54:375–388, 2006.
- [27] P. Feldmann and R. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Trans. Computer-Aided Design*, 28:137–158, 1993.
-

- 
- [28] R.P. Feynman, R.B. Leighton, and M. Sands. *The Feynman Lecture on Physics*. Addison-Wesley, 1964.
- [29] H. Gajewski, K. Gröger, and K. Zacharias. *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*. Mathematische Lehrbücher und Monographien. II. Abteilung. Band 38. Berlin: Akademie-Verlag, 1974.
- [30] M. Grasselli. An identification problem for an abstract linear hyperbolic integrodifferential equation with applications. *J. Math. Anal. Appl.*, 171:27–60, 1992.
- [31] M. Grasselli, S.I. Kabanikhin, and A. Lorenzi. An inverse hyperbolic integrodifferential problem arising in geophysics II. *Nonlinear Analysis: Theory, Methods & Applications*, 15(3):283 – 298, 1990.
- [32] T.H. Grönwall. Note on the Derivatives with Respect to a Parameter of the Solutions of a System of Differential Equations. *Annals of Mathematics*, 4:292–296, 1919.
- [33] J.L. Guermond and A. Ern. *Theory and practice of finite elements*. Springer, 2004.
- [34] D. Guidetti. Convergence to a stationary state for solutions to parabolic inverse problems of reconstruction of convolution kernels. *Differ. Integral Equ.*, 20(9):961–990, 2007.
- [35] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52, 1902.
- [36] R.F. Harrington. *Introduction to Electromagnetic Engineering*. Courier Dover Publications, 2003.
- [37] R. Hasan, R. Sabariego, C. Geuzaine, and Y. Paquay. Proper orthogonal decomposition versus Krylov subspace methods in reduced-order energy-converter models. *IEEE ENERGYCON*, pages 1–6, 2016.
- [38] A. Hasanov. Simultaneous determination of source terms in a linear parabolic problem from the final overdetermination: Weak solution approach. *J. Math. Anal. Appl.*, 330:766–779, 2007.
-

- 
- [39] A. Hasanov. An inverse source problem with single Dirichlet type measured output data for a linear parabolic equation. *Appl. Math. Lett.*, 24(7):1269–1273, 2011.
- [40] A. Hasanov and B. Pektaş. Identification of an unknown time-dependent heat source term from overspecified dirichlet boundary data by conjugate gradient method. *Computers & Mathematics with Applications*, 65(1):42 – 57, 2013.
- [41] A. Hasanov and M. Slodička. An analysis of inverse source problems with final time measured output data for the heat conduction equation: A semi-group approach. *Applied Mathematics Letters*, 26(2):207–214, 2013.
- [42] M.R. Hestens and E. Stiefel. Methods of conjugate gradients for the solution of linear systems. *J. Res.Res.Natl.Bur.Stand.*, 49:409–436, 1952.
- [43] M. Hinze and K. Kunisch. Three control methods for time-dependent fluid flow. *Flow, Turbulence and Combustion*, 65:273–298, 2000.
- [44] M. Hinze and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM J. Numer. Anal.*, 40:492–515, 2002.
- [45] M. Hinze and S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. *Lecture Notes in Computational and Applied Mathematics*, pages 261–306, 2005.
- [46] H. Hoetelling. Simplified calculation of principal component analysis. *Psychometrika*, 1:27–35, 1935.
- [47] Y. Huang, J. Li, and Q. Lin. Superconvergence analysis for time-dependent Maxwell’s equations in metamaterials. *Numerical Methods for Partial Differential Equations*, 28(6):1794–1816, 2012.
- [48] Y. Huang, J. Li, and W. Yang. Solving metamaterial Maxwell’s equations via a vector wave integro-differential equation. *Computers & Mathematics with Applications*, 63(12):1597 – 1606, 2012.
-

- 
- [49] V. Isakov. *Inverse source problems*. Mathematical Surveys and Monographs, 34. Providence, RI: American Mathematical Society (AMS), 1990.
- [50] M.I. Ismailov, F. Kanca, and D. Lesnic. Determination of a time-dependent heat source under nonlocal boundary and integral overdetermination conditions. *Appl. Math. Comput.*, 218(8):4138–4146, 2011.
- [51] J. Janno. Discretization and regularization of an inverse problem related to a quasilinear hyperbolic integrodifferential equation. *Inverse Problems*, 13(3):711–728, 1997.
- [52] J. Janno. Recovering a time- and space-dependent kernel in a hyperbolic integro-differential equation from a restricted dirichlet-to-neumann operator. *Electron. J. Diff. Eqns.*, 2004(67):1–16, 2004.
- [53] J. Janno, E.M. Kiss, and L. von Wolfersdorf. On Tikhonov regularization for identifying memory kernels in heat conduction and viscoelasticity. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 80(4):259–272, 2000.
- [54] T. Johansson and D. Lesnic. A variational method for identifying a spacewise dependent heat source. *IMA J. Appl. Math.*, 72:748–760, 2007.
- [55] K. Karhunen. Zur Spektraltheorie stochastischer Prozesse. *Ann.Acad. Sci. Fennicae. Ser. A.I. Math.-Phys.*, 34, 1946.
- [56] J. Kačur. *Method of Rothe in evolution equations*. Teubner-Texte zur Mathematik, 1985.
- [57] J. Kačur. Application of Rothe’s method to evolution integrodifferential equations. *J. Reine Angew. Math.*, 388:73–105, 1988.
- [58] J.B. Keller. Inverse problems. *The American Mathematical Monthly*, 83(2):107–118, 1976.
- [59] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90:117–148, 2001.
- [60] L D. Landau and E M. Lifshitz. *Electrodynamics of Continuous Media*, volume 8. Oxford: Pergamon Press, 1960.
-

- 
- [61] M. Loeve. Fonctions aleatoire de second ordre. *Revue*, pages 195–206, 1946.
- [62] M. Mojžiš. TEMPO - Teória elektromagnetického poľa. Lecture notes published by Comenius University, Bratislava, Slovakia, 2006.
- [63] P. Monk. *Finite Element Methods for Maxwell's Equations*. Oxford University Press Inc., New York, 2003.
- [64] A. Mordecai. *Nonlinear Programming: Analysis and Methods*. 2003.
- [65] J. Nečas. *Direct Methods in the Theory of Elliptic Equations*. Springer Monographs in Mathematics, 2012.
- [66] K. Pearson. On lines and planes of closest to points in space. *Philosophical Magazine*, 2:609–629, 1901.
- [67] A. I. Prilepko and V. V. Solov'ev. Solvability theorems and Rothe's method for inverse problems for a parabolic equation. *Differential Equations*, 23:1341–1349, 1988.
- [68] A.I. Prilepko, D.G. Orlovsky, and I.A. Vasin. *Methods for solving inverse problems in mathematical physics*. Pure and Applied Mathematics, Marcel Dekker. 231. New York, NY: Marcel Dekker, 2000.
- [69] K. Rektorys. *The method of discretization in time and partial differential equations*. Reidel Publishing Company, Dordrecht-Boston-London, 1982.
- [70] E. Rothe. Zweidimensionale parabolische Randwertaufgaben als Grenzfall eindimensionaler Randwertaufgaben. *Mathematische Annalen*, 1(102):650–670, 1930.
- [71] W. Rundell. Determination of an unknown non-homogeneous term in a linear partial differential equation from overspecified boundary data. *Applicable Analysis*, 10:231–242, 1980.
- [72] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.
- [73] W.H.A. Schilder and H. van der Vorst. *Model Order Reduction: Theory, Research Aspects and Applications*. 2008.
-

- 
- [74] W. Schilders. Introduction to Model Order Reduction. *Mathematics in industry*, 13:3–32, 2008.
- [75] L. Sirovich. Turbulence and the dynamics of coherent structures. *Quart. Appl. Math*, 45:561–590, 1987.
- [76] M. Slodička. Recovery of boundary conditions in heat transfer. In O. Fudym, J.-L. Battaglia, and G.S. Dulikravich et al., editors, *IPDO 2013 : 4th Inverse problems, design and optimization symposium, 2013 June 26-28, Albi, Ecole des Mines d’Albi-Carmaux*, 10p., 2013. ISBN 979-10-91526-01-2.
- [77] M. Slodička. Determination of a solely time-dependent source in a semilinear parabolic problem by means of boundary measurements. *JCAM*, 289:433–440, 2015.
- [78] M. Slodička. A parabolic inverse source problem with a dynamical boundary condition. *Applied Mathematics and Computation*, 256:529–539, 2015.
- [79] M. Slodička and M. Galba. Determination of a time-dependent convolution kernel from a boundary measurement in nonlinear Maxwell’s equations. *Computers & Mathematics with Applications*, 72:1484–1500, 2016.
- [80] M. Slodička and M. Galba. Recovery of a time dependent source from a surface measurement in Maxwell’s equations. *Computers & Mathematics with Applications*, 71:368–380, 2016.
- [81] M. Slodička and L. Šeliga. Determination of a time-dependent convolution kernel in a nonlinear hyperbolic equation. *Inverse Problems in Science and Engineering*, 24(6):1011–1029, 2016.
- [82] V. V. Solov’ev. Solvability of the inverse problems of finding a source, using overdetermination on the upper base for a parabolic equation. *Differential Equations*, 25:1114–1119, 1990.
- [83] I. G. Stratis and A. N. Yannacopoulos. Electromagnetic fields in linear and nonlinear chiral media: a time-domain analysis. *Abstract and Applied Analysis*, 2004(6):471–486, 2004.
-



- [84] A.N. Tikhonov. On the solution of incorrectly formulated problems and the regularization method. *Soviet Math. Doklady*, (4):1035–1038, 1963. English translation.
- [85] M. Zaslavsky and V. Druskin. Solution of time-convolutionary Maxwell's equations using parameter-dependent Krylov subspace reduction. *Journal of Computational Physics*, 229:4831–4839, 2010.
-