

VŠB–Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
Department of Applied Mathematics

Optimal Shape Design in Magnetostatics
Ph.D. Thesis

Author: **Ing. Dalibor Lukáš**

Supervisor: **Prof. RNDr. Zdeněk Dostál, CSc.**
Department of Applied Mathematics,
VŠB–Technical University of Ostrava

Referees: **Prof. RNDr. Jaroslav Haslinger, DrSc.**
Faculty of Mathematics and Physics,
Charles University Prague

Prof. RNDr. Michal Křížek, DrSc.
Mathematical Institute,
The Academy of Sciences of the Czech Republic

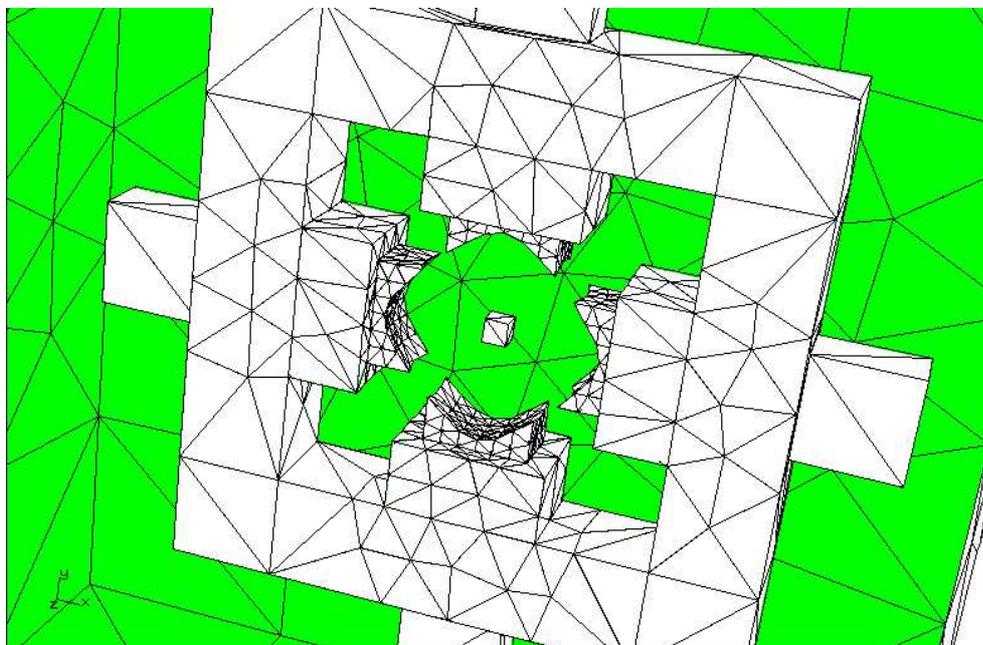
o.Univ.Prof. Dipl.-Ing. Dr. Ulrich Langer
Institute of Computational Mathematics,
Johannes Kepler University Linz

Submitted in September 2003

VŠB–Technical University of Ostrava
Faculty of Electrical Engineering and Computer Science
Department of Applied Mathematics



Optimal Shape Design in Magnetostatics
Ph.D. Thesis



September 2003

Ing. Dalibor Lukáš

Dedicated to Šárka

Abstract

This thesis treats with theoretical and computational aspects of three-dimensional optimal shape design problems that are governed by linear magnetostatics. The aim is to present a complete process of mathematical modelling in a well-balanced way. We step-by-step visit the world of physics, functional analysis, computational mathematics, and we end up with real-life applications. Nevertheless, the main emphasis is put on an efficient implementation of numerical methods for shape optimization which exploits an effective evaluation of gradients by the adjoint method and a just recently introduced multilevel optimization approach. We also emphasize numerical experiments with real-life problems of complex three-dimensional geometries.

We begin from a description of the electromagnetic phenomena by Maxwell's equations and we derive their three-dimensional (3d) and two-dimensional (2d) magnetostatic cases with the linear constitutive relation between the magnetic flux density and the magnetic strength density.

Then we start to develop a general theory that covers both 2d and 3d optimal interface-shape design problems that are constrained by a second-order linear elliptic boundary value problem (BVP). First we pose a weak formulation of the BVP with the homogeneous Dirichlet boundary condition. Whenever the kernel of the BVP operator is not trivial, we employ a regularization technique such that the regularized solutions converge to the true one. The continuous weak formulation of the abstract BVP is discretized by the first-order finite element method on triangles and tetrahedra, respectively. We set an abstract continuous shape optimization problem, the state problem of which involves one or more BVPs such that they only differ in the right-hand sides, i.e., different current excitations in case of magnetostatics. The design boundary is an interface between two materials, rather than a part of the computational domain boundary, as it is usual in optimal shape design for mechanics. We prove the existence of an optimal shape by checking the continuity of the cost functional and the compactness of the set of admissible shapes. Then we discretize the continuous optimization problem by the finite element method and prove the existence of the approximate solutions. The main theoretical result of this thesis is a proof of the convergence of the approximate optimized solutions to an optimal solution of the continuous problem, where we also involve an inner approximation of the original computational domain with a Lipschitz boundary by a polyhedral (in the 3d case) or polygonal (in the 2d case) domain. Throughout the abstract theory we introduce many assumptions that are checked for concrete applications afterwards. These assumptions show the scope of the theory.

Concerning the computational aspects in optimization, we use the sequential quadratic programming method with a successive approximation of the Hessian. To justify the use, we verify the smoothness of both the discretized cost and constraint functionals. Then we focus on the calculation of gradients by means of the adjoint method and we derive an efficient algorithm for that, including its Matlab implementation enclosed on the CD. We introduce a new multilevel optimization approach as a possible adaptive optimization method.

Finally, we end up with physics again. We present two real-life applications with rather com-

plex 3d geometries. After some motivation, we describe the optimization problem in terms of verifying the theoretical assumptions, and we give numerical results. We present the speedup of the adjoint method comparing to the numerical differentiation, and of the multilevel approach comparing to the classical optimization. One optimized design was manufactured, we are provided with measurements and, at the end, we discuss real improvements of the cost functional.

Acknowledgement

First of all I want to express my deep gratitude to my supervisor Prof. RNDr. Zdeněk Dostál, CSc., the head of the Department of Applied Mathematics at the Technical University of Ostrava (TUO), who introduced me to the magic of numerical mathematics and its applications in electromagnetism. I thank him for his kindness and patience during more than six years of my undergraduate and doctoral studies for when he has been supervising me.

Special thanks are devoted to my very good friend, teacher, and colleague Doc. RNDr. Jiří Bouchala, Ph.D., who is an associated professor at the Department of Applied Mathematics, TUO. Jirka convinced me that the mathematics is beautiful itself, even without applications. I thank to all the other friends, colleagues, and teachers of mine from the Department of Applied Mathematics, TUO, especially to Mgr. Vít Vondrák, Ph.D. and Mgr. Bohumil Krajc, Ph.D.

I am much obliged to o.Univ.Prof. Dipl.-Ing. Dr. Ulrich Langer, who is the head of the Institute for Numerical Mathematics and the co–speaker of the Special Research Initiative SFB F013 “Numerical and Symbolic Scientific Computing” both associated to the Johannes Kepler University Linz in Austria. Professor Langer offered me a Ph.D. position within the research project SFB F013, where I spent one year and where I have been currently working again. Here I have made a big progress in my research work and I have learned about efficient methods for solving large discretized direct simulation problems based on the finite element and/or boundary element discretizations. I am also much indebted to my current or former colleagues, especially, I would like to mention Dr. Dipl.-Ing. Joachim Schöberl, Dr. Dipl.-Ing. Michael Kuhn, MSc., and Dr. Dipl.-Ing. Wolfram Mühlhuber.

I further thank to Prof. Ing. Jaromír Pištora, CSc., who is with the Institute of Physics, TUO, and is the head of the Department of Education at the Faculty of Electrical Engineering. Professor Pištora asked me to cooperate on development of a new generation of electromagnets used in the research on magneto-optic effects. I also thank for the fruitful cooperation to the other colleagues from the Institute of Physics, namely to Dr. Mgr. Kamil Postava, Dr. RNDr. Dalibor Ciprian, Dr. Ing. Michal Lesňák, Ing. Martin Foldyna, and Ing. Igor Kopřiva.

I very appreciate the possibility to consult my work with Prof. RNDr. Jaroslav Haslinger, DrSc., who is a world leading expert in optimal shape design. I thank very much to Prof. RNDr. Michal Křížek, DrSc. from the Mathematical Institute of the Czech Academy of Sciences that I could learn a lot during several consultations in his office. I very much admire the enthusiasm of Prof. Křížek that he has devoted to math.

At the end let me express my deep grateness and love to my parents and to my girlfriend Šárka, with whom my life is complete, no matter what the career is about.

This work has been supported by the Austrian Science Fund FWF within the SFB “Numerical and Symbolic Computing” under the grant SFB F013, by the Czech Ministry of Education under the research project CEZ: J17/98:272400019, and by the Grant Agency of the Czech Republic under the grant 105/99/1698.

Notation

\mathbb{N}	non–negative integers
\mathbb{R}	real numbers
i	imaginary unit
\mathbb{C}	complex plane
C_1, \dots, C_{16}	fixed constants

Abbreviations

1d	one–dimensional
2d	two–dimensional
3d	three–dimensional
PDE	partial differential equation
BVP	boundary (vector–)value problem
FEM	finite element method
BFGS	update formula for the Hessian matrix named after Broyden, Fletcher, Goldfarb, and Shanno
SQP	sequential quadratic programming
AD	automatic differentiation

Chapter 2

B	magnetic flux density	p. 10
H	magnetic field	p. 10
μ	permeability	p. 10
J	direct electric current density	p. 10
u	magnetic vector potential	p. 10
Ω	three–dimensional computational domain	p. 10
Ω_{2d}	two–dimensional reduced computational domain which is the cross section of Ω with the plane $x_3 = 0$	p. 11
J	two–dimensional scalar direct electric current density	p. 11
u	two–dimensional scalar magnetic potential	p. 11

Chapter 3

$\ \cdot\ _U$	norm in the normed linear vector space U	p. 14
V/U	quotient space	p. 14
Ker(L)	kernel of the linear vector operator L	p. 15
U'	dual space to the normed linear vector space U	p. 15
$\langle \cdot, \cdot \rangle$	duality pairing	p. 15
(\cdot, \cdot)	scalar product	p. 15

U^\perp	orthogonal complement to the space U	p. 16
$H = U \oplus U^\perp$	orthogonal decomposition of the Hilbert space H	p. 17
\mathbb{R}^n	Euclidean space consisting of n -dimensional real vectors	p. 17
\mathbf{A}^T	transposed matrix	p. 18
$\det(\mathbf{A})$	determinant of the matrix \mathbf{A}	p. 18
$\tilde{\mathbf{A}}$	adjoint matrix	p. 18
\mathbf{A}^{-1}	inverse matrix	p. 18
m	dimension of the computational domain Ω , $m \in \{2, 3\}$	p. 19
Ω	domain, i.e., open, bounded, and connected subset of \mathbb{R}^m	p. 19
$\overline{\Omega}$	closure of the domain Ω	p. 19
$\partial\Omega$	boundary of the domain Ω	p. 19
\mathbf{n}	unit outer normal vector to the boundary $\partial\Omega$	p. 19
$C(\overline{\Omega})$	space of functions continuous over $\overline{\Omega}$	p. 20
$C^k(\overline{\Omega})$	space of functions which are continuous up to their k -th partial derivatives over Ω , $k \in \mathbb{N}$	p. 20
$C^\infty(\overline{\Omega})$	space of infinitely differentiable functions over $\overline{\Omega}$	p. 21
$\text{supp } v$	support of the function v	p. 21
$C_0^\infty(\Omega)$	space of infinitely differentiable functions with a compact support in $\overline{\Omega}$	p. 21
$C^{0,1}(\overline{\Omega})$	space of Lipschitz continuous functions over $\overline{\Omega}$	p. 21
\mathcal{L}	set of all the domains with Lipschitz continuous boundaries	p. 21
div	divergence operator	p. 22
\mathbf{grad}	gradient operator	p. 22
\mathbf{curl}	curl operator	p. 22
$\mathbf{n} \times \mathbf{u}$	cross product, tangential component of the function \mathbf{u} along the boundary $\partial\Omega$	p. 23
\mathbf{B}	linear vector first-order differential operator, $\mathbf{B} : [C^1(\overline{\Omega})]^{\nu_1} \mapsto [C(\overline{\Omega})]^{\nu_2}$, $\nu_1, \nu_2 \in \mathbb{N}$	p. 23
\mathbf{B}^*	adjoint operator related to \mathbf{B} by Green's theorem, $\mathbf{B}^* : [C^1(\overline{\Omega})]^{\nu_2} \mapsto [C(\overline{\Omega})]^{\nu_1}$	p. 23
γ	trace operator related to \mathbf{B} by Green's theorem, $\gamma : [C(\overline{\Omega})]^{\nu_1} \mapsto [C(\partial\Omega)]^{\nu_2}$	p. 23
$L^p(\Omega)$	Lebesgue space of measurable functions defined over Ω for which the Lebesgue integral of their p -th power is finite, $p \in [1, \infty)$	p. 24
$L^\infty(\Omega)$	Lebesgue space of measurable essentially bounded functions over Ω	p. 24
$\text{meas}(\Omega)$	Lebesgue measure of the domain Ω	p. 24
a.e.	almost everywhere	p. 24
$D^\alpha u$	the α -th generalized derivative of the function u , α is a multi-index	p. 25
$H^k(\Omega)$	Sobolev space of functions whose generalized derivatives up to the k -th order belong to $L^p(\Omega)$	p. 25
$(\cdot, \cdot)_{k,\Omega}$	scalar product in $H^k(\Omega)$	p. 25
$\ \cdot\ _{k,\Omega}$	norm in $H^k(\Omega)$	p. 25
$ \cdot _{k,\Omega}$	seminorm in $H^k(\Omega)$	p. 25
$[H^k(\Omega)]^n$	Cartesian product of Sobolev spaces, $n \in \mathbb{N}$	p. 25
$(\cdot, \cdot)_{n,k,\Omega}$	scalar product in $[H^k(\Omega)]^n$	p. 25
$H_0^1(\Omega)$	space of functions from $H^1(\Omega)$ whose traces vanish along $\partial\Omega$	p. 26

$H^{1/2}(\partial\Omega)$	space of traces of all functions from $H^1(\Omega)$	p. 26
$H^{-1/2}(\partial\Omega)$	dual space to $H^{1/2}(\partial\Omega)$	p. 26
$\mathbf{H}(\mathbf{B}; \Omega)$	space of functions from $[L^2(\Omega)]^{\nu_1}$ whose generalized operator \mathbf{B} (gradient, divergence, curl, etc.) is in $[L^2(\Omega)]^{\nu_2}$	p. 30
$(\cdot, \cdot)_{\mathbf{B}, \Omega}$	scalar product in $\mathbf{H}(\mathbf{B}; \Omega)$	p. 30
$\ \cdot\ _{\mathbf{B}, \Omega}$	norm in $\mathbf{H}(\mathbf{B}; \Omega)$	p. 31
$ \cdot _{\mathbf{B}, \Omega}$	seminorm in $\mathbf{H}(\mathbf{B}; \Omega)$	p. 31
$\mathbf{H}_0(\mathbf{B}; \Omega)$	space of functions from $\mathbf{H}(\mathbf{B}; \Omega)$ whose trace vanishes along $\partial\Omega$	p. 31
$\mathbf{Ker}(\mathbf{B}; \Omega)$	space of functions from $\mathbf{H}_0(\mathbf{B}; \Omega)$ whose belong to the kernel of \mathbf{B}	p. 31
$\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$	space of functions from $\mathbf{H}_0(\mathbf{B}; \Omega)$ that are orthogonal to $\mathbf{Ker}(\mathbf{B}; \Omega)$	p. 31
(S)	strong formulation of an abstract linear elliptic boundary vector–value problem	p. 32
\mathbf{D}	matrix function of material coefficients in (S)	p. 32
\mathbf{f}	vector function of the right–hand side in (S)	p. 32
$a(\cdot, \cdot)$	bilinear form in (W)	p. 33
$f(\cdot)$	linear functional in (W)	p. 33
(W)	weak formulation of an abstract linear elliptic boundary vector–value problem	p. 33
\mathbf{u}	solution to (W)	p. 33
ε	positive regularization parameter that regularizes the non–ellipticity of the bilinear form $a(\cdot, \cdot)$	p. 34
$a_\varepsilon(\cdot, \cdot)$	regularized bilinear form in (W_ε)	p. 34
(W_ε)	regularized weak formulation	p. 34
\mathbf{u}_ε	solution to (W_ε)	p. 34
Chapter 4		
h	positive discretization parameter	p. 39
\mathbf{V}^h	finite dimensional subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$	p. 39
n	dimension of the space \mathbf{V}^h	p. 39
\mathbf{D}^h	discretization of the matrix function \mathbf{D}	p. 39
$a_\varepsilon^h(\cdot, \cdot)$	discretization of the bilinear form $a_\varepsilon(\cdot, \cdot)$	p. 39
\mathbf{f}^h	discretization of the right–hand side \mathbf{f}	p. 39
$f^h(\cdot)$	discretization of the linear functional $f(\cdot)$	p. 39
(W_ε^h)	Galerkin discretization of the problem (W_ε)	p. 40
\mathbf{u}_ε^h	solution to (W_ε^h)	p. 40
\mathbf{A}_ε^n	system matrix that arises from the discretized bilinear form $a_\varepsilon^h(\cdot, \cdot)$	p. 42
\mathbf{f}^n	right–hand side vector that arises from the discretized linear functional $f^h(\cdot)$	p. 42
\mathbf{u}_ε^n	solution vector (corresponds to \mathbf{u}_ε^h) of the arising linear system	p. 42
Ω^h	polyhedral subdomain that approximates Ω from inner	p. 42
n_{Ω^h}	number of finite elements	p. 42
K_i, K^{e_i}	domain (triangle or tetrahedron) of the i –th element	p. 42
\mathcal{T}^h	discretization (e.g., triangulation) of Ω^h into elements	p. 43
\mathbf{x}^h	block vector of all the discretization nodes	p. 43
$n_{\mathbf{x}^h}$	number of the discretization nodes	p. 43

\mathbf{x}_i^h	coordinates of the i -th discretization node	p. 43
e_i	the i -th finite element	p. 43
\mathbf{P}^{e_i}	finite element space of the i -th element	p. 43
n_e	number of local degrees of freedom	p. 43
$\sigma_j^{e_i}$	the j -th local degree of freedom of the i -th element	p. 43
Σ^{e_i}	set of all the degrees of freedom of the i -th element	p. 43
E^h	set of all the finite elements	p. 43
σ_i^h	the i -th global degree of freedom	p. 43
Σ^h	set of all the n global degrees of freedom	p. 43
\mathcal{G}^{e_i}	mapping from local to global degrees of freedom for the i -th element, $\mathcal{G}^{e_i} : \{1, \dots, n_e\} \mapsto \{1, \dots, n\}$	p. 43
$\delta_{i,j}$	Kronecker's symbol	p. 44
$\xi_j^{e_i}$	the j -th local shape (base) function of the i -th element	p. 44
ξ_j^h	the j -th global shape (base) function	p. 44
\mathbf{P}^h	global finite element space	p. 44
\mathcal{I}_0^h	set of indices of those global degrees of freedom that determine the trace along $\partial\Omega^h$	p. 45
$\mathbf{H}_0(\mathbf{B}; \Omega^h)^h$	the finite element space \mathbf{V}^h	p. 45
$(W_\varepsilon(\Omega))$	the problem (W_ε) for varying computational domain Ω	p. 45
$\mathbf{u}_\varepsilon(\Omega)$	solution to $(W_\varepsilon(\Omega))$	p. 45
\mathbf{D}^{e_i}	the coefficient matrix \mathbf{D}^h restricted to the i -th element	p. 45
\mathbf{f}^{e_i}	the right-hand side vector \mathbf{f}^h restricted to the i -th element	p. 45
$(W_\varepsilon^h(\Omega^h))$	finite element discretization of the problem $(W_\varepsilon(\Omega^h))$	p. 45
$\mathbf{u}_\varepsilon^h(\Omega^h)$	solution to $(W_\varepsilon^h(\Omega^h))$	p. 45
E_i^h	set of the elements neighbouring with the i -th element	p. 46
$a_\varepsilon^{e_i}(\cdot, \cdot)$	contribution to the bilinear form $a_\varepsilon^h(\cdot, \cdot)$ from the i -th element	p. 46
$f^{e_i}(\cdot)$	contribution to the linear functional $f^h(\cdot)$ from the i -th element	p. 46
\mathbf{x}^{e_i}	block vector of all the corners of the i -th element	p. 46
$\mathbf{x}_j^{e_i}$	coordinates of the j -th corner of the i -th element	p. 46
\mathcal{H}^{e_i}	mapping from the element nodal indices to the global nodal indices, $\mathcal{H}^{e_i} : \{1, \dots, m+1\} \mapsto \{1, \dots, n_{\mathbf{x}^h}\}$	p. 46
r	reference element	p. 47
$\widehat{\mathbf{x}}^r$	block vector of all the reference element nodes	p. 47
$\widehat{\mathbf{x}}_j^r$	the i -th corner of the reference element	p. 47
$\mathcal{R}^{e_i}, \mathbf{R}^{e_i}$	linear mapping, the related matrix, from the reference element to the i -th element, $\mathcal{R}^{e_i} : \overline{K}^r \mapsto \overline{K}^{e_i}$	p. 47
$\mathcal{S}^{e_i}, \mathbf{S}^{e_i}$	linear mapping, the related matrix, of finite element functions defined over \overline{K}^r to the ones defined over \overline{K}^{e_i} , $\mathcal{S}^{e_i} : \mathbf{P}^r \mapsto \mathbf{P}^{e_i}$	p. 48
$\mathcal{S}_{\mathbf{B}}^{e_i}, \mathbf{S}_{\mathbf{B}}^{e_i}$	linear mapping, the related matrix, of the finite element functions defined over \overline{K}^r to the ones defined over \overline{K}^{e_i} under the operator \mathbf{B} , $\mathcal{S}_{\mathbf{B}}^{e_i} : [L^2(K^r)]^{\nu_2} \mapsto [L^2(K^{e_i})]^{\nu_2}$	p. 48
$\mathbf{B}_\varepsilon^{n,e_i}$	elementwise constant vector of the operator \mathbf{B} applied to the solution \mathbf{u}_ε^h	p. 50
\mathbf{B}_ε^n	block vector of the elementwise constant vectors $\mathbf{B}_\varepsilon^{n,e_i}$	p. 50
h^{e_i}	discretization parameter associated to the i -th element	p. 50
\bar{h}	the coarsest possible discretization parameter	p. 50

\mathbf{X}_ν^h	linear extension operator that extends functions by zero, $\mathbf{X}_\nu^h : [L^2(\Omega^h)]^\nu \mapsto [L^2(\Omega)]^\nu$, $\Omega^h \subset \Omega$, $\nu \in \mathbb{N}$	p. 51
$\mathbf{X}_0(\mathbf{B}; \Omega; \Omega^h)^h$	space of functions extended from $\mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ by $\mathbf{X}_{\nu_1}^h$	p. 51
$\Omega^h \nearrow \Omega$	approximation of Ω by polygonal domains Ω^h from inner	p. 51
χ_Ω	characteristic function of Ω	p. 52
π^{e_i}	interpolation operator associated to the i -th element, $\pi^{e_i} : [C^\infty(\overline{K^{e_i}})]^{\nu_1} \mapsto \mathbf{P}^{e_i}$	p. 52
π^h	global interpolation operator, $\pi^h : [C^\infty(\overline{\Omega^h})]^{\nu_1} \mapsto \mathbf{P}^h$	p. 52
π_0^h	global interpolation operator, $\pi_0^h : [C^\infty(\overline{\Omega^h})]^{\nu_1} \mapsto \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$	p. 52
$\Delta^h \Omega^h$	the most outer layer of finite elements	p. 53

Chapter 5

ω	nonempty polyhedral $(m - 1)$ -dimensional domain	p. 68
α	shape, $\alpha \in C(\overline{\omega})$	p. 68
α_l, α_u	lower and upper box constraints, $\alpha_l, \alpha_u \in \mathbb{R}$	p. 68
\mathcal{U}	set of admissible shapes	p. 68
$\alpha_n \rightrightarrows \alpha$	uniform convergence of shapes in \mathcal{U}	p. 68
n_Υ	number of design parameters	p. 68
Υ	set of admissible design parameters, $\Upsilon \subset \mathbb{R}^{n_\Upsilon}$	p. 68
F	parameterization of the admissible shapes, $F : \Upsilon \mapsto \mathcal{U}$	p. 68
$\Omega_0(\alpha), \Omega_1(\alpha)$	decomposition of Ω controlled by the shape α	p. 68
$\text{graph}(\alpha)$	graph of the shape α	p. 69
\mathbf{D}_α	material matrix function controlled by the shape α	p. 69
$\mathbf{D}_0, \mathbf{D}_1$	constant material matrices for the domains $\Omega_0(\alpha), \Omega_1(\alpha)$	p. 69
$a_\alpha(\cdot, \cdot)$	the bilinear form $a(\cdot, \cdot)$ controlled by the shape α	p. 69
n_v	number of variations of the right-hand side	p. 69
v	state index within the multistate problem, $v \in \{1, \dots, n_v\}$	p. 69
\mathbf{f}^v	one of the n_v right-hand side vectors	p. 69
$f^v(\cdot)$	one of the n_v linear functionals that corresponds to \mathbf{f}^v	p. 69
$(W^v(\alpha))$	multistate problem controlled by the shape α	p. 70
$\mathbf{u}^v(\alpha)$	solution to the multistate problem $(W^v(\alpha))$	p. 70
\mathcal{I}	cost functional, $\mathcal{I} : \mathcal{U} \times [[L^2(\Omega)]^{\nu_2}]^{n_v} \mapsto \mathbb{R}$	p. 71
\mathcal{J}	cost functional, $\mathcal{J} : \mathcal{U} \mapsto \mathbb{R}$	p. 71
(P)	continuous setting of the shape optimization problem	p. 72
α^*	solution to (P)	p. 72
$\tilde{\mathcal{J}}$	parameterized cost functional, $\tilde{\mathcal{J}} : \Upsilon \mapsto \mathbb{R}$	p. 72
(\tilde{P})	continuous setting of the shape optimization problem solved for design parameters	p. 72
\mathbf{p}^*	solution to (\tilde{P})	p. 72
$a_{\varepsilon, \alpha}(\cdot, \cdot)$	the regularized bilinear form $a_\varepsilon(\cdot, \cdot)$ controlled by the shape α	p. 72
$(W_\varepsilon^v(\alpha))$	the multistate problem $(W^v(\alpha))$ regularized by the regularization parameter ε	p. 72
\mathcal{J}_ε	regularized cost functional	p. 73
(P_ε)	regularized setting of the shape optimization problem	p. 73
α_ε^*	solution to (P_ε)	p. 73
$\tilde{\mathcal{J}}_\varepsilon$	regularized and parameterized cost functional, $\tilde{\mathcal{J}} : \Upsilon \mapsto \mathbb{R}$	p. 74

(\tilde{P}_ε)	regularized setting of the shape optimization problem solved for design parameters	p. 74
\mathbf{p}_ε^*	solution to (\tilde{P}_ε)	p. 74
n_ω^h	number of elements in the discretization of ω	p. 74
ω_i^h	the i -th element in the discretization of ω	p. 74
\mathcal{T}_ω^h	discretization of ω	p. 74
$P^1(\mathcal{T}_\omega^h)$	space of continuous functions that are linear over $\overline{\omega_i^h}$	p. 74
$\mathbf{x}_{\omega_i^h, j}^h$	the j -th corner of the i -th element in the discretization of ω	p. 74
$n_{\mathbf{x}_\omega^h}$	number of nodes in the discretization of ω	p. 74
$\mathbf{x}_{\omega, j}^h$	the j -th node in the discretization of ω	p. 74
\mathcal{U}^h	discretized set of admissible shapes	p. 74
α^h	discretized shape, $\alpha^h \in \mathcal{U}^h$	p. 74
π_ω^h	interpolation operator, $\pi_\omega^h : \mathcal{U} \mapsto P^1(\mathcal{T}_\omega^h)$	p. 75
$\Omega_0^h(\alpha^h), \Omega_1^h(\alpha^h)$	decomposition of Ω^h controlled by the discretized shape α^h	p. 75
$\mathcal{T}^h(\alpha^h)$	discretization of Ω^h controlled by the discretized shape α^h	p. 75
$(W_\varepsilon^{v, h}(\alpha^h))$	finite element discretization of the regularized multistate problem $(W_\varepsilon^v(\alpha^h))$ controlled by the discretized shape α^h	p. 76
$\mathbf{u}_\varepsilon^{v, h}(\alpha^h)$	solution to $(W_\varepsilon^{v, h}(\alpha^h))$	p. 76
$a_{\varepsilon, \alpha^h}^h(\cdot, \cdot)$	the discretized and regularized bilinear form $a_\varepsilon^h(\cdot, \cdot)$ controlled by the discretized shape α^h	p. 76
$\mathbf{D}_{\alpha^h}^h$	the discretized material matrix function \mathbf{D}^h controlled by the discretized shape α^h	p. 76
$f^{v, h}(\cdot)$	the discretized multistate linear functional $f^v(\cdot)$	p. 76
$\mathcal{J}_\varepsilon^h$	the discretized and regularized cost functional, $\mathcal{J} : \mathcal{U}^h \mapsto \mathbb{R}$	p. 80
$(\tilde{P}_\varepsilon^h)$	discretization of the regularized shape optimization problem (P_ε)	p. 80
α_ε^{h*}	solution to $(\tilde{P}_\varepsilon^h)$	p. 80
$\tilde{\mathcal{J}}_\varepsilon^h$	discretization of the regularized and parameterized cost functional $\tilde{\mathcal{J}}_\varepsilon, \tilde{\mathcal{J}}_\varepsilon^h : \Upsilon \mapsto \mathbb{R}$	p. 81
$(\tilde{P}_\varepsilon^h)$	discretization of the regularized setting (\tilde{P}_ε) of the shape optimization problem solved for design parameters	p. 81

Chapter 6

n_v	number of constraints	p. 84
\mathbf{v}	constraint function, $\mathbf{v} : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_v}$	p. 84
n_{α^h}	number of shape nodes, $n_{\alpha^h} := n_{\mathbf{x}_\omega^h}$	p. 84
α^h	design-to-shape mapping, $\alpha^h : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_{\alpha^h}}$	p. 84
\mathbf{x}^h	shape-to-mesh mapping, $\mathbf{x}^h : \mathbb{R}^{n_{\alpha^h}} \mapsto \mathbb{R}^{mn_{\mathbf{x}^h}}$	p. 84
\mathbf{x}_0^h	block vector of all the initial grid nodes, $\mathbf{x}_0^h \in \mathbb{R}^{mn_{\mathbf{x}^h}}$	p. 84
$\Delta \mathbf{x}^h$	block vector of all the grid displacements between the current and initial grid, $\Delta \mathbf{x}^h \in \mathbb{R}^{mn_{\mathbf{x}^h}}$	p. 84
\mathcal{M}^h	matrix that identically maps the shape displacements α^h onto the corresponding grid nodal coordinates \mathbf{x}^h , $\mathcal{M}^h : \mathbb{R}^{n_{\alpha^h}} \mapsto \mathbb{R}^{mn_{\mathbf{x}^h}}$	p. 84
$\mathbf{K}^h(\mathbf{x}_0^h)$	stiffness matrix for the initial grid \mathbf{x}_0^h of the auxiliary elasticity (shape-to-mesh) problem	p. 84
$\mathbf{b}^h(\alpha^h)$	right-hand side vector involving the inhomogeneous Dirichlet condition α^h of the auxiliary elasticity (shape-to-mesh) problem	p. 84

$\mathbf{f}^{v,n}(\mathbf{x}^h)$	right-hand side vector that arises from the discretized linear functional $f^{v,h}(\cdot)$ for the v -th state	p. 85
$\mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h)$	solution vector (corresponds to $\mathbf{u}_\varepsilon^{v,h}(\alpha^h)$) of the arising v -th linear system	p. 85
$\mathbf{u}_\varepsilon^{v,n,e_i}(\mathbf{x}^h)$	the solution vector $\mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h)$ restricted to the i -th element	p. 85
$\mathbf{B}_\varepsilon^{v,n,e_i}(\mathbf{x}^h)$	elementwise constant vector of the operator \mathbf{B} applied to the solution $\mathbf{u}_\varepsilon^{v,h}(\mathbf{x}^h)$	p. 85
$\mathbf{B}_\varepsilon^{v,n}(\mathbf{x}^h)$	block vector of the elementwise constant vectors $\mathbf{B}_\varepsilon^{v,n,e_i}(\mathbf{x}^h)$	p. 85
\mathcal{I}^h	revisited discretized cost functional, $\mathcal{I}^h : \mathbb{R}^{n_{\alpha^h}} \times \mathbb{R}^{mn_{\mathbf{x}^h}} \times [\mathbb{R}^{\nu_2 n_{\Omega^h}}]^{n_v} \mapsto \mathbb{R}$	p. 86
(\mathcal{QP})	quadratic programming problem	p. 90
$(\mathcal{QP}_1(\mathbf{p}_0))$	quadratic programming subproblem for the line search approach	p. 90
Hess	Hessian, matrix of all the second partial derivatives of a scalar function	p. 90
Grad	gradient of a vector function whose columns are gradients of the particular components of the function	p. 90
$(\mathcal{LS}(\mathbf{p}_0, \mathbf{s}_{\mathcal{QP}}^*))$	line search problem	p. 90
$(\mathcal{QP}_2(\mathbf{p}_0, d))$	quadratic programming subproblem for the trust region method	p. 91
\mathbf{H}_k	the k -th successive approximation of the Hessian, $k \in \mathbb{N}$	p. 92
\mathbf{G}	matrix involving the sensitivity of the multistate system upon the grid displacements	p. 96

Chapter 7

MC	Maltese Cross electromagnet	p. 105
O-Ring	O-Ring electromagnet	p. 105
Ω_m	magnetization area	p. 106
Ω_{yoke}	domain occupied by the ferromagnetic yoke	p. 108
$\Omega_{\text{westp}}, \dots$	domains occupied by the ferromagnetic poles	p. 108
$\Omega_{\text{westc}}, \dots$	domains occupied by the coils which complete the related poles	p. 108
$\mathbf{x}_{\omega,i,j}$	the (i, j) -th node in the (tensor product) regular discretization of the 2d shape domain ω	p. 111
$p_{i,j}$	design parameter related to the node $\mathbf{x}_{\omega,i,j}$, a component of \mathbf{p}	p. 111
β_i^n	Bernstein polynomial, $\beta_i^n : [0, 1] \mapsto \mathbb{R}$, $n, i \in \mathbb{N}$, $i \leq n$	p. 111
I	direct electric current	p. 112
n_I	number of turns	p. 112
S_c	area of the coil cross section	p. 112
\mathbf{J}^v	current density for the current variation v	p. 112
φ	the part of the cost functional that measures the homogeneity, $\varphi : [L^2(\Omega)]^m \mapsto \mathbb{R}$, $m \in \{2, 3\}$	p. 114
θ^v	the part of the cost functional that for the v -th state problem penalizes the magnetic field being below the minimal required magnitude, $\theta^v : [L^2(\Omega)]^m \mapsto \mathbb{R}$, $m \in \{2, 3\}$	p. 114
ρ	penalty for θ^v	p. 114
$B^{\text{avg},v}$	average normal component (to the magnetization plane) of the magnetic field of the v -th state problem, $B^{\text{avg},v} : [L^2(\Omega)]^m \mapsto \mathbb{R}$, $m \in \{2, 3\}$	p. 114
\mathbf{n}_m^v	unit outer normal vector to the v -th magnetization plane	p. 114

$B_{\min}^{\text{avg},v}$	minimal required magnitude of the magnetic field of the v -th state problem	p. 114
φ^h	discretization of φ , $\varphi^h : \mathbb{R}^{mn_{\Omega^h}} \mapsto \mathbb{R}$	p. 116
$\theta^{v,h}$	discretization of θ^v , $\theta^{v,h} : \mathbb{R}^{mn_{\Omega^h}} \mapsto \mathbb{R}$	p. 116
$B^{\text{avg},v,n}$	discretization of $B^{\text{avg},v}$	p. 116
Γ_{α^h}	design interface	p. 117
$\mathbf{x}_{\omega,i}$	the i -th node in the regular discretization of the 1d shape domain ω	p. 118
p_i	design parameter related to the node $\mathbf{x}_{\omega,i}$, a component of \mathbf{p}	p. 118

Contents

Abstract	v
Acknowledgement	vii
Notation	ix
Contents	xvii
1 Introduction	1
1.1 General aspects of optimization	2
1.1.1 Optimization problems: Classification and connections	2
1.1.2 Optimization methods	3
1.1.3 Iterative methods for linear systems of equations	3
1.1.4 Commercial versus academic software tools	4
1.2 Optimal shape design	5
1.3 Computational electromagnetism	6
1.4 Structure of the thesis	6
2 Mathematical modelling in magnetostatics	9
2.1 Maxwell's equations	9
2.2 Three-dimensional linear magnetostatics	10
2.3 Two-dimensional linear magnetostatics	11
3 Abstract boundary vector-value problems	13
3.1 Preliminaries from linear functional analysis	13
3.1.1 Normed linear vector spaces	13
3.1.2 Linear operators	14
3.1.3 Hilbert spaces	15
3.1.4 Linear algebra	17
3.2 Preliminaries from real analysis	19
3.2.1 Continuous function spaces	20
3.2.2 Some fundamental theorems	21
3.3 Hilbert function spaces	24
3.3.1 Lebesgue spaces	24
3.3.2 Sobolev spaces	25
3.3.3 The space $H(\mathbf{grad})$	26
3.3.4 The space $\mathbf{H}(\mathbf{curl})$	28
3.3.5 The space $\mathbf{H}(\mathbf{div})$	29

3.3.6	The abstract space $\mathbf{H}(\mathbf{B})$	30
3.4	Weak formulations of boundary vector–value problems	32
3.4.1	A regularized formulation in $\mathbf{H}_0(\mathbf{B})$	34
3.4.2	A weak formulation of three–dimensional linear magnetostatics	36
3.4.3	A weak formulation of two–dimensional linear magnetostatics	37
4	Finite element method	39
4.1	The concept of the method	39
4.1.1	Galerkin approximation	39
4.1.2	Finite element method	42
4.1.3	Discretization of the domain	42
4.1.4	Space of finite elements	43
4.1.5	Finite element discretization of the weak formulation	45
4.2	Assembling finite elements	46
4.2.1	Reference element	46
4.2.2	<i>BDB</i> integrators	48
4.2.3	The algorithm	49
4.3	Approximation properties	50
4.3.1	Approximation of the domain by polyhedra	51
4.3.2	A–priori error estimate	52
4.3.3	Regular discretizations	52
4.3.4	Convergence of the finite element method	54
4.4	Finite elements for magnetostatics	57
4.4.1	Linear Lagrange elements on triangles	58
4.4.2	Linear Nédélec elements on tetrahedra	62
5	Abstract optimal shape design problem	67
5.1	A fundamental theorem	67
5.2	Continuous setting	68
5.2.1	Admissible shapes	68
5.2.2	Multistate problem	69
5.2.3	Shape optimization problem	71
5.3	Regularized setting	72
5.4	Discretized setting	74
5.4.1	Discretized set of admissible shapes	74
5.4.2	Discretized multistate problem	75
5.4.3	Discretized optimization problem	80
6	Numerical methods for shape optimization	83
6.1	The discretized optimization problem revisited	83
6.1.1	Constraint function	84
6.1.2	Design–to–shape mapping	84
6.1.3	Shape–to–mesh mapping	84
6.1.4	Multistate problem	85
6.1.5	Cost functional	86
6.1.6	Smoothness of the cost functional	87
6.2	Newton–type optimization methods	89
6.2.1	Quadratic programming subproblem	90

6.2.2	Sequential quadratic programming	91
6.3	The first-order sensitivity analysis methods	92
6.3.1	Sensitivities of the cost and constraint functions	93
6.3.2	State sensitivity	94
6.3.3	Semianalytical methods	96
6.3.4	Adjoint method	98
6.3.5	An object-oriented software library	98
6.3.6	A note on using the automatic differentiation	100
6.4	Multilevel optimization approach	101
7	An application and numerical experiments	105
7.1	A physical problem	105
7.2	Three-dimensional mathematical setting	108
7.2.1	Geometries of the electromagnets	108
7.2.2	Set of admissible shapes	109
7.2.3	Continuous multistate problem	112
7.2.4	Continuous shape optimization problem	113
7.2.5	Regularization and finite element discretization	115
7.3	Two-dimensional mathematical setting	117
7.4	Numerical results	119
7.4.1	Testing the multilevel approach	120
7.4.2	Testing the adjoint method	121
7.5	Manufacture and measurements	124
8	Conclusion	127
	Bibliography	129
	Curriculum vitae	143

Chapter 1

Introduction

Nowadays dynamic progress in computer technology has made powerful computers to become cheap. This has been influencing the development of numerical methods. Many both commercial and academic simulation software tools are available for a large variety of problems. Computer simulations replaced prototyping. A usual picture is that developers in a company are modelling a new product on a computer, doing some calculations, and thinking what parameters and how to shift to achieve better properties of the product. Still increasing standard of technologies brings together experts from different areas. Developers' work is now much more interdisciplinary. It involves

- experts in the area of main interest, e.g., engineers, physicists, medics, economists, etc.,
- theoretical mathematicians who introduce correct theories that can be used for mathematical modelling,
- numerical analysts who design efficient numerical methods and analyze their properties, e.g., speedup, convergence rate, etc.,
- and computer scientists who effectively implement the methods on a proper platform.

The people who are experienced in more areas are especially welcome to coordinate the design process.

As far as the direct simulation is fast enough, it is straightforward to automatize also the synthesis (design) process. To this end, a developer has to exactly formulate

- the objective criterion saying what design is better,
- the design parameters that can be changed including their possible limit values,
- and some additional constraining criteria that the product must satisfy.

The objective criterion (the optimization goal) might be the minimal weight, the maximal output power, the minimal cost, the minimal loss, etc. The design parameters are for example size of the product, microstructure of the used material, or shape of the product. We might additionally require that the product must not exceed a given volume, weight, or that it must be robust, e.g., stiff enough. Once we know these exactly, we have formulated an *optimization problem* that can be solved automatically.

1.1 General aspects of optimization

We can optimize very disparate systems, for instance, maximize profit of a market, minimize pollution of a forest, minimize petrol consumption when driving a car, control a robot in an optimal way, etc. Optimal shape design is only a small area within the general optimization context. Besides, we can mention optimal control, optimal sizing, thickness optimization, topology optimization, optimization in graphs, and so further. Each class of optimization problems has a structure of its own. However, a direct simulation – the *state problem* – is always involved and the inverse process (the synthesis) plays around with some parameters and resolve the direct problem until the system behaves in a required way.

1.1.1 Optimization problems: Classification and connections

Optimization can be seen in a wider context of *inverse problems*, in which we know behaviour of a system, usually from physical measurements, and under this knowledge we are looking for structure of the system and/or for distribution of sources. A typical inverse problem is the computer tomography in medicine. An introductory textbook to this field is given by KIRSCH [109]. Inverse problems are known to be ill-posed which has to be treated by regularization techniques, see ENGL, HANKE, AND NEUBAUER [58]. Some connections between optimization and inverse problems are presented by NEITTAANMÄKI, RUDNICKI, AND SAVINI [144].

Here, we are especially interested in *structural optimization*, where we change the structure of an object, which is interacted in a physical field, in order to achieve required behaviour. The structure means either material properties, topology, or shape of the object boundary or interfaces. Various issues of structural optimization are covered in BANICHUK [17], BENDSØE [21], CHERKAEV [43], KALAMKAROV [105], OLHOFF AND TAYLOR [152], PEDERSEN [154], ROZVANY [173, 174, 175], SAVE AND PRAGER [180, 181], XIE AND STEVEN [213]. Applications in electromagnetism are given by HOPPE, PETROVA, AND SCHULZ [97], in plasticity by YUGE AND KIKUCHI [216], and, for instance, in ergonomics by RASMUSSEN ET AL. [165]. An optimal design of microstructures is presented by JACOBSEN, OLHOFF, AND RØNHOLT [100].

If we are interested in the topology design – it is usually the question where to put holes – we speak about *topology optimization*. The basic literature is BENDSØE [20, 21], BENDSØE AND SIGMUND [22], BORRVALL [25]. Some applications in electromagnetism are presented by HOPPE, PETROVA, AND SCHULZ [96], YOO AND KIKUCHI [214]. More theoretical issues are given by STADLER [198] or by SIGMUND AND PETERSSON [190].

In the design process, the second step after topology optimization is *shape optimization*, where we tune the shape of the boundary or interfaces. The basic literature on shape optimization is given by BEGIS AND GLOWINSKI [19], MURAT AND SIMON [140], PIRONNEAU [159], HASLINGER AND NEITTAANMÄKI [85], HASLINGER AND MÄKINEN [83], SOKOLOWSKI AND ZOLELIO [196], BÖRNER [24], DELFOUR AND ZOLELIO [54], KAWOHL ET AL. [107], MOHAMMADI AND PIRONNEAU [135]. Besides the basic textbooks, one can find a lot of theoretical analysis in BUCUR AND ZOLELIO [35], PEICHL AND RING [155, 156], PETERSSON AND HASLINGER [158], PETERSSON [157]. Papers focused on applications in electromagnetism are, for example, DI BARBA ET AL. [18], BRANDSTÄTTER ET AL. [30], LUKÁŠ [123], MARROCCO AND PIRONNEAU [132], TAKAHASHI [206].

It turns out that there is much in common in topology and shape optimization. Recently there have appeared several papers in this context, like CEA ET AL. [40], RIETZ AND PETERSSON [171], TANG AND CHANG [207].

1.1.2 Optimization methods

Another point of view to optimization is from the side of numerical mathematics. We can classify optimization problems with respect to what algorithm is used. There is a class of *evolutionary algorithms*, cf. XIE AND STEVEN [213], typical examples of which are genetic algorithms, which search for the global optimum. However, the number of evaluations of the objective functional is exponential to the number of design variables. This is due to the fact that the whole design space has to be randomly explored. On the other hand, there are *Newton-like algorithms*, which look for a local optimum. These use the first-, eventually the second-order derivatives to approximate the objective functional locally by a quadratic function. The local algorithms are much faster comparing to the global ones and in this thesis we will concern with them only. Many algorithmical issues of local optimization are covered in NOCEDAL AND WRIGHT [148], FLETCHER [61], DENNIS AND SCHNABEL [55], GILL, MURRAY, AND WRIGHT [66], GROSSMANN AND TERNO [72], CEA [39], HESTENSEN [88, 89], HAGER, HEARN, AND PARDALOS [81], POLAK [161], MÜHLHUBER [139], BOGGS AND TOLLE [23], CONN, GOULD, AND TOINT [49]. However, there are also optimization problems, whose cost functional is not differentiable, not even twice differentiable. This is the case of nonsmooth optimization, see e.g., CLARKE [48], MÄKELÄ AND NEITTAANMÄKI [130], HASLINGER, MIETTINEN, AND PANATIOTOPOULOS [84]. Let us also mention multicriterial optimization, which tries to include more aspects with respect to which the design should be optimal, see OLHOFF [151].

There are several interesting optimization techniques that have appeared just recently. In the papers by BURGER AND MÜHLHUBER [37, 38] they solve simultaneously for both the design and state variables, i.e., they minimize at the same time the cost functional as well as the quadratic energy functional of the direct problem. Another challenging issue in optimization is adaptivity. A hierarchical approach in shape optimization is used by LUKÁŠ [123, 128]. The works of RAMM, MAUTE, AND SCHWARZ [164], SCHLEUPEN, MAUTE, AND RAMM [185] make even use of the FE-adaptivity in both the topology and shape optimization. Using multilevel approach for solving nonlinear ill-posed problems is presented in SCHERZER [182].

The Newton-like optimization methods suffer from the computational costs and from the fact that they are searching for local optima. It is partly overcome by the homogenization method, the study of which has just been started. It aims at describing macroscopic behaviour of materials with heterogeneous microstructures. For the literature see ALLAIRE [5], ALLAIRE ET AL. [7], CIORANESCU AND DONATO [47]. The method is very much connected to structural (both shape and topology) optimization, which is studied in ALLAIRE [6], SUZUKI AND KIKUCHI [202], YOO AND KIKUCHI [214], YUGE, IWAI, AND KIKUCHI [215]. However, this method is well-suited only for some cost functionals and the linear elasticity. Another new interesting approach is the level-set method, see SETHIAN AND WIEGMANN [188], ALLAIRE, JOUVE, AND TOADER [8]. It determines the set of admissible designs implicitly by a level-set function and uses the shape or topology derivative with respect to this implicit scheme. The level-set method was already applied in the field of inverse problems, see BURGER [36]. Nevertheless, the method involves a time explicit scheme, which takes many iterations. An overcome can be done by a coupling with Newton methods.

1.1.3 Iterative methods for linear systems of equations

The main computational effort is related with solution of the state problem. Fast solution iterative methods have been especially developed for linear systems with sparse symmetric positive definite matrices. Such a system can be stated as a quadratic minimization problem. For those, since 50

years ago, the development of *conjugate gradients methods* has been running. The conjugate gradients methods are looking for minimum of the quadratic functional in the directions that are conjugated by the energy scalar product related to the system matrix. The research was initiated by HESTENSEN AND STIEFEL [90] and from then an extensive literature to this topic has been written, see ZOUTENDIJK [221], GOLUB AND VAN LOAN [69], AXELSSON [13], SAAD [179].

Nowadays, the key point is the construction of proper preconditioners. The ones that turned out to be the best are based on *multigrid techniques*. They construct a hierarchy of finite element discretizations such that they first minimize the low frequencies (eigenvalues) of the residual error on a coarse discretization, which is very fast, and then the higher frequencies on a finer one, which is again fast, as the low frequencies are not any longer present. The hierarchy can be constructed either with respect to the computational grid (geometrical multigrids) or with respect to the structure of the system matrix (algebraic multigrids). Various topics on multigrid techniques are presented in HACKBUSCH [78], BRAMBLE [28], BRAMBLE, PASCIAK, AND XU [29], HIPTMAIR [91], JUNG AND LANGER [102], REITZINGER [169], HAASE AND LANGER [74], HAASE ET AL. [75, 76]. Applications in electromagnetism can be found in SCHINNERL ET AL. [183, 184]. A software package based on algebraic multigrids was done by REITZINGER [168].

1.1.4 Commercial versus academic software tools

Basically, we distinguish between commercial and academic software. Commercial software tools, see THOMAS, ZHOU, AND SCHRAMM [209] for a review, are developed to provide a large functionality in a user-friendly way. They have to really attract as large audience as possible in order to survive in the commercial market. They try to be robust, automatic, and sexy. They benefit from a deep engineering experience. From the matter of fact, commercial software tools are much more suitable for immediate applications in the industry than the academic ones, because the developers are much more closer to the industrial users. However, from a lack of knowledge they cannot provide the latest scientific computational methods and the solution time is often rather slow. Whenever the user needs more functionality, he/she has to wait until a new release is done. Typical commercial software packages for both analysis and design are ANSYS [1] or FLUENT [2].

On the other hand, scientific computing tools are developed with respect to an a priori known scientific goal. At the very beginning they do not need to attract a large audience, as they are supported by research grants. They do not need to be user-friendly, as the researchers that use them know very well what is going on and can remove some errors themselves. Their main advantage is that the scientific computing tools implement the up-to-date knowledge and they use fast solution methods that have just appeared in the world research. However, they cannot be directly applied in the industry, since they do not treat complicated real-life geometries, they are not as user-friendly and as robust as the commercial ones. Some scientific computing tools for the analysis or optimization are presented in KUHN, LANGER, AND SCHÖBERL [117], SILVA AND BITTERCOUNT [191], RASMUSSEN ET AL. [166], PARKINSON AND BALLING [153]. An example of more educational software system is in TSCHERNIAK AND SIGMUND [210]. A typical commercial software directed to the academy is MATLAB [208].

Until recently, one could hardly work in both the industry and academy, as their objectives were rather different. The nowadays trend seems to be towards interdisciplinary work. Industrial partners are invited to talk at scientific symposia, and many companies invest to further education of their staff. The gap between the industry and academy gets smaller, thus, the difference between the commercial and scientific software does so. The commercial software should take more into account the latest research progress and the scientists developing research software packages should put more effort into the documentation, user-interface, and better coordination of the

development. The more communication between the industry and academy there is, the more improvement can be done.

1.2 Optimal shape design

In this thesis we treat with optimal shape design problems. These are very well-structured, in a consequence of which we can design a very efficient solution method taking the structure into account. The direct problem within the shape optimization is a partial differential equation (PDE). There is still a number of PDEs that can be considered. Imagine the following examples: a flying aeroplane, a loaded bridge, an electromagnet pumped by direct electric currents (DC), or a hydraulic press acting on a piece of steel. Here, the PDEs are very diverse, namely, air fluid can be modelled by a hyperbolic PDE, flight of the aeroplane by a parabolic PDE, load of the bridge or the DC electromagnet by elliptic PDEs, and the hydraulic press is modelled as a contact problem, which is nonsmooth. If we concern nonlinear constitutive relations, then the PDEs are even more complicated to solve. Hence, the solution method should be suited for the type of PDE that we are concerned with. In this thesis we will deal with shape optimization governed by elliptic linear PDEs.

Since we will employ Newton algorithms for smooth optimization, the crucial point is the *sensitivity analysis*, which is the evaluation of gradients of the cost and constraint functionals with respect to the design variables. One can either derive a Fréchet derivative from the continuous setting of the optimization problem, see SOKOŁOWSKI AND ZOLESIO [196], or discretize the continuous problem first and then use an algebraic approach, see HASLINGER AND NEITTAANMÄKI [85]. We prefer the second approach. In SOKOŁOWSKI AND ZOCHOWSKI [195] a connection between topological and shape sensitivity analysis is presented.

Let us consider a shape optimization problem governed by an elliptic PDE on a bounded computational domain. The most common solution approach is the following: At the very beginning, given an initial shape design, we decompose the computational domain into polygonal (or polyhedral) convex elements, cf. GEORGE [65]. Then we discretize a weak formulation of the elliptic PDE by the finite element method (FEM). We get a sparse positive definite system matrix and a right-hand side vector. We employ a fast iterative method to solve the system with a sufficient precision. Then we calculate the cost (optimization) functional and we can start play around with the shape. Some design variables describe the design boundary (or interface). Changes of the design variables are mapped onto displacements of the nodes lying on the design boundary (or interface), e.g., by means of Bézier parameterization, cf. FARIN [59]. Displacements of nodes along the design boundary influence displacements of the remaining nodes in the discretization grid. Finally, the displaced discretization grid influences the system matrix, and eventually the right-hand side vector. Thus, the grid influences the solution of the PDE and so it also influences the cost functional. The main effort in the sensitivity analysis lies on an efficient evaluation of the gradient of the solution to the direct simulation problem. This is usually done by the adjoint method, cf. HAUG, CHOI, AND KOMKOV [86].

Besides the finite element analysis, there are also boundary element methods (BEM). They discretize the boundary integral form of the PDE. These techniques are not as spread as the finite elements. The fundamental principles are covered in BANERJEE AND BUTTERFIELD [16], BREBBIA [31], CHEN AND ZHOU [41]. Using BEM in optimal shape design is presented in CHEN, ZHOU, AND MCLEAN [42], KITA AND TANIE [110], or in SIMON [193]. Applications in electromagnetism are given in HIPTMAIR AND SCHWAB [94], and in KALTENBACHER ET AL. [106].

1.3 Computational electromagnetism

Due to some historical aspects, the finite element method was firstly reviewed and applied by mechanical engineers, cf. ZIENKIEWICZ [217]. Then the fluid dynamics community started to use the method. Electrical engineers and scientists started to apply the method little bit later, nevertheless, mainly in the last two decades the finite element analysis has been becoming still more popular in the electromagnetic community, too. An important point was introducing a new class of finite elements by NÉDÉLEC [142, 143]. A lot of theoretical work was done by ADAM ET AL. [3], AM-ROUCHE ET AL. [9], COSTABEL AND DAUGE [50], HIPTMAIR [92, 93], KAČŮR ET AL. [104], MONK [136, 137, 138], NEITTAANMÄKI AND SARANEN [146, 147]. The fundamental textbooks were published by ARORA [11], BOSSAVIT [26], GIRAULT AND RAVIART [67], IDA AND BAS-TOS [99], KOST [112], KŘÍŽEK AND NEITTAANMÄKI [115], MAYERGOYZ [133], SILVESTER AND FERRARI [192], STEELE [199], VANRIENEN [211].

1.4 Structure of the thesis

The rest of the thesis is structured as follows. In Chapter 2, we describe Maxwell's equations for 3-dimensional time-dependent electromagnetic fields. By neglecting some phenomena we respectively arrive at 3-dimensional (3d) and 2-dimensional (2d) linear magnetostatic cases.

At the beginning of Chapter 3, we recall fundamental issues of the linear functional analysis. We describe the Sobolev spaces H^1 , $\mathbf{H}(\text{div})$, and $\mathbf{H}(\text{curl})$. Then we start to develop an abstract theory, which can cover a wide class of boundary value problems. Under four assumptions we introduce an abstract function space $\mathbf{H}(\mathbf{B})$ for some abstract elliptic first-order vector-value linear differential operator \mathbf{B} . We basically assume the abstract space to be dense in C^∞ , to fulfill the trace theorem, Green's theorem, and Friedrichs-like inequality. Further, we formulate an abstract linear elliptic second-order boundary vector-value problem (BVP) with the homogeneous Dirichlet boundary condition. We derive its weak formulation. In the cases when the kernel of the operator \mathbf{B} is not trivial, the bilinear form is not $\mathbf{H}(\mathbf{B})$ -elliptic and the weak formulation is not suited for the finite element discretization. Therefore, we introduce a regularized weak formulation and prove the convergence of the regularized solutions to the true one in the seminorm. At the end of Chapter 3, we apply the theory to both 2d and 3d linear magnetostatics while all the previously introduced assumptions are verified.

In Chapter 4, we first recall the general concept of the finite element method. Then we deal with algorithmical aspects and derive an efficient assembling algorithm, which approximates our abstract BVP. Further, under some assumptions we prove convergence of the approximate solutions to the true weak solution of the BVP. The proof is mainly based on first Strang's lemma and on the Lebesgue dominated convergence theorem. In this approximation theory there is also involved an inner approximation of the original Lipschitz boundary by polygonal (or polyhedral) ones. At the end, we present Lagrange nodal and Nédélec edge finite elements, which are respectively used for 2d and 3d magnetostatics. For these two types of elements we verify all the assumptions of the introduced convergence theory.

In Chapter 5, we introduce a continuous setting of a shape optimization problem, which is governed by the abstract BVP. The shape controls an interface between two materials while the computational domain is fixed. We spend some effort in proving the continuity of the state solution with respect to the shape. We suppose the cost functional to be continuously dependent on the state solution. The set of admissible shapes is compact by definition. Hence, the existence of an optimal solution can be proved. We further employ the regularization of the state problem

and prove the corresponding convergence result. Finally, we discretize the state problem and, consequently, the optimization problem by means of the finite element method. We prove the convergence of the discretized optimized shapes to a continuous optimal one. The convergence theory uses very standard tools of the functional and finite element analysis and it was inspired by the monograph of HASLINGER AND NEITTAANMÄKI [85]. Nevertheless, one of its assumptions, namely the continuity of the mapping between the shape nodes and the remaining grid nodes, is difficult to assure in practice. For non-academic problems with complex geometries and for fine discretizations one can hardly find such a continuous shape-to-mesh mapping, as for large changes of the design shape some disturbed or even flipped elements can appear and the geometry has to be re-meshed. This brings the discontinuity of the cost functional into the business. In Conclusion, possible outcomes are discussed.

In Chapter 6, we revisit our abstract optimization problem from the computational point of view. We analyze the structure of the cost functional and present it as a compound mapping, consisting of several smooth submappings. Therefore, we can prove the smoothness of the cost functional, which justify us to use algorithms of the Newton type afterwards. We briefly mention all the ingredients of the sequential quadratic programming method. Then, we derive an efficient method for the first-order sensitivity analysis, including its implementation in Matlab, which is enclosed on the CD. This is actually the heart of the whole thesis. Finally, we introduce a multilevel optimization algorithm, which is well-designed to be adaptive with respect to a posteriori the error analysis of the underlying finite element discretization of the state problem. We refer to the very recent papers by SCHLEUPEN, MAUTE, AND RAMM [185], RAMM, MAUTE, AND SCHWARZ [164], in which the finite element adaptivity is already used for calculating error of the approximation of the cost functional.

At the beginning of Chapter 7, we present an application, which has arisen from the research on magneto-optic effects. Our aim is to find optimal shapes of two electromagnets in order to minimize inhomogeneities of the magnetic field in a certain area. The electromagnets have rather complex 3-dimensional geometries. We formulate the cost functional, the set of admissible shapes, and the state problem such that, simultaneously, we verify the related theoretical assumptions. We further pose the corresponding reduced 2-dimensional settings of the problems. Then, both 2d and 3d numerical results are given. We discuss the speedup of the used adjoint method comparing to the numerical differentiation, and the speedup of the multilevel approach with respect to the standard approach. Finally, an optimized shape was manufactured and we are provided with physical measurements of the magnetic fields for both the original and optimized electromagnets. We present the improvements of the magnetic field in terms of the cost functional.

In Conclusion, we summarize the results of this thesis and give directions of the further research.

Chapter 2

Mathematical modelling in magnetostatics

In this chapter, we will start from Maxwell's equations in a general time-dependent 3-dimensional (3d) setting, we will pass through the time-harmonic case, and we will end up with the 3d magnetostatic boundary value problem. Neglecting the magnetic phenomena in a given direction, we will arrive at the 2-dimensional (2d) magnetostatic boundary value problem. Throughout this chapter, we will formally describe the physical phenomena, rather than introduce all the necessary assumptions on the smoothness of the domain or the differentiability of the physical quantities. Mathematically correct settings will be introduced in Chapter 3.

For the theory of electromagnetism we refer to FEYNAM, LEIGHTON, AND SANDS [60], HAUS AND MELCHER [87], SOLYMAR [197], and STRATTON [201]. The monographs focused more on numerical modelling are given by KŘÍŽEK AND NEITTAANMÄKI [115], BOSSAVIT [26], VAN RIENEN [211], IDA AND BASTOS [99], KOST [112], MAYERGOYZ [133], STEELE [199].

2.1 Maxwell's equations

The physical phenomena of the time-dependent 3-dimensional electromagnetic field are described by Maxwell's equations

$$\left. \begin{aligned} \operatorname{curl}(\mathcal{H}) &= \mathcal{J} + \sigma\mathcal{E} + \frac{\partial\mathcal{D}}{\partial t} \\ \operatorname{curl}(\mathcal{E}) &= -\frac{\partial\mathcal{B}}{\partial t} \\ \operatorname{div}(\mathcal{D}) &= \rho \\ \operatorname{div}(\mathcal{B}) &= 0 \end{aligned} \right\}, \quad (2.1)$$

together with the constitutive relations

$$\mathcal{D} = \varepsilon\mathcal{E} \text{ and } \mathcal{B} = \mu\mathcal{H}, \quad (2.2)$$

where \mathcal{E} denotes the electric field (electric intensity), \mathcal{D} is the electric flux density, $\varepsilon > 0$ is the permittivity, \mathcal{J} is the external electric current density, $\sigma > 0$ is the electric conductivity, $\rho \geq 0$ is the charge density, \mathcal{H} is the magnetic field, \mathcal{B} is the magnetic flux density, $\mu > 0$ is the permeability, $t \geq 0$ is the time and the differential operators are defined as follows:

$$\operatorname{div}(\mathbf{v}) := \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{\partial v_3}{\partial x_3},$$

$$\mathbf{curl}(\mathbf{v}) := \left(\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}, \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1}, \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \right),$$

where $\mathbf{v} = (v_1, v_2, v_3)$ is a vector function and $\mathbf{x} = (x_1, x_2, x_3)$ are point coordinates.

Now we introduce 3d time-harmonic linear Maxwell's equations. First, we restrict our consideration into a fixed bounded domain, assuming that the fields vanish outside the domain. Let $\Omega \subset \mathbb{R}^3$ be a nonempty domain the boundary $\partial\Omega$ of which is smooth enough. Further, let $\omega \geq 0$ be an angular frequency and assume that both the electric and magnetic field are time-harmonic

$$\begin{aligned} \mathcal{E}(\mathbf{x}, t) &:= \operatorname{Re}(\mathbf{E}(\mathbf{x})e^{i\omega t}), \\ \mathcal{H}(\mathbf{x}, t) &:= \operatorname{Re}(\mathbf{H}(\mathbf{x})e^{i\omega t}), \end{aligned}$$

where $\mathbf{E} := \mathbf{E}(\mathbf{x})$ and $\mathbf{H} := \mathbf{H}(\mathbf{x})$ are complex-valued vector functions, i denotes the imaginary unit, and $\operatorname{Re}(\mathbf{v}) := (\operatorname{Re}(v_1), \operatorname{Re}(v_2), \operatorname{Re}(v_3))$ is the component-wise real part of the vector $\mathbf{v} := (v_1, v_2, v_3)$. Moreover, we assume the charge density and Maxwell's current to be zeros

$$\rho = 0 \text{ and } \frac{\partial \mathcal{D}}{\partial t} = 0.$$

We also assume that the constitutive relations (2.2) are time-independent, real-valued and linear

$$\varepsilon := \varepsilon(\mathbf{x}) \text{ and } \mu := \mu(\mathbf{x}),$$

rather than $\varepsilon := \varepsilon(\mathbf{x}, \mathcal{E})$ and $\mu := \mu(\mathbf{x}, \mathcal{H})$ in the nonlinear case. Finally, we assume the external current density \mathcal{J} and the conductivity σ to be time-independent and real-valued

$$\mathcal{J} := \mathbf{J}(\mathbf{x}) \text{ and } \sigma := \sigma(\mathbf{x}).$$

Now, Maxwell's equations (2.1) can be rewritten as follows:

$$\left. \begin{aligned} \mathbf{curl}(\mathbf{H}) &= \mathbf{J} + \sigma\mathbf{E} \\ \mathbf{curl}(\mathbf{E}) &= -i\omega\mathbf{B} \\ \operatorname{div}(\mathbf{D}) &= 0 \\ \operatorname{div}(\mathbf{B}) &= 0 \end{aligned} \right\} \text{ in } \Omega, \quad (2.3)$$

where, according to (2.2), $\mathbf{D} = \varepsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$. We prescribe that the electric field vanishes on the boundary

$$\mathbf{n} \times \mathbf{E} = \mathbf{0} \text{ on } \partial\Omega, \quad (2.4)$$

where \mathbf{n} denotes the outer unit normal to $\partial\Omega$ and where, given vectors $\mathbf{u} := (u_1, u_2, u_3)$ and $\mathbf{v} := (v_1, v_2, v_3)$,

$$\mathbf{u} \times \mathbf{v} := (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)$$

is the vector cross product.

2.2 Three-dimensional linear magnetostatics

We introduce the magnetic vector potential \mathbf{u} by

$$\mathbf{curl}(\mathbf{u}) = \mathbf{B}.$$

The first two equations in (2.3) now read as follows:

$$\left. \begin{aligned} \operatorname{curl} \left(\frac{1}{\mu} \operatorname{curl}(\mathbf{u}) \right) &= \mathbf{J} + \sigma \mathbf{E} \\ \operatorname{curl}(\mathbf{E}) &= -i\omega \operatorname{curl}(\mathbf{u}) \end{aligned} \right\} \text{ in } \Omega,$$

the third equation becomes

$$\operatorname{div}(-i\omega \varepsilon \mathbf{u}) = 0 \text{ in } \Omega$$

and last Maxwell's equation is automatically fulfilled, since the vector identity $\operatorname{div}(\operatorname{curl}(\mathbf{u})) = 0$ holds. We consider the time-independent case of (2.3). Taking $\omega := 0$ and neglecting the electric field, we arrive at the following magnetostatic boundary value problem

$$\left. \begin{aligned} \operatorname{curl} \left(\frac{1}{\mu} \operatorname{curl}(\mathbf{u}) \right) &= \mathbf{J} \quad \text{in } \Omega \\ \mathbf{n} \times \mathbf{u} &= \mathbf{0} \quad \text{on } \partial\Omega \end{aligned} \right\}. \quad (2.5)$$

2.3 Two-dimensional linear magnetostatics

Let us assume that the magnetic field given by (2.5) does not significantly depend on the x_3 -coordinate. This is often the case when $\mathbf{J}(\mathbf{x}) = (0, 0, J(x_1, x_2))$ and $\mu(\mathbf{x}) = \mu(x_1, x_2)$ in a large enough neighbourhood of the zero-plane $\mathcal{Z} := \{\mathbf{x} \in \mathbb{R}^3 \mid x_3 = 0\}$. We are interested in an approximate solution of (2.5) in this neighbourhood. So, let us assume that

$$\mathbf{J}(\mathbf{x}) := (0, 0, J(x_1, x_2)), \quad \mu(\mathbf{x}) := \mu(x_1, x_2), \quad \text{and } \mathbf{u}(\mathbf{x}) := (0, 0, u(x_1, x_2)).$$

Using the latter, the problem (2.5) reduces to the following

$$\left. \begin{aligned} -\operatorname{div} \left(\frac{1}{\mu} \operatorname{grad}(u) \right) &= J \quad \text{in } \Omega_{2d} \\ u &= 0 \quad \text{on } \partial\Omega_{2d} \end{aligned} \right\}, \quad (2.6)$$

where

$$\Omega_{2d} := \{\mathbf{x}' = (x_1, x_2) \in \mathbb{R}^2 \mid (x_1, x_2, 0) \in \Omega\}$$

represents a cross section of Ω in the sense $\Omega_{2d} \times \{0\} = \Omega \cap \mathcal{Z}$, and where the differential operator **grad** is defined as follows:

$$\operatorname{grad}(u) := \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2} \right),$$

where u is a scalar function. It is easy to see that the magnetic flux density is then given by

$$\mathbf{B} = \left(\frac{\partial u}{\partial x_2}, -\frac{\partial u}{\partial x_1}, 0 \right),$$

where u solves (2.6).

Chapter 3

Abstract boundary vector–value problems

In this chapter, we will recall the necessary mathematics used for weak formulations in magnetostatics. We will begin with the continuous function spaces and introduce some Sobolev spaces together with the corresponding trace, Green’s theorems, and Friedrichs’–like inequalities. Then, we will formally describe the concept of a weak formulation for an abstract elliptic linear boundary vector–value problem. Particularly, we will illustrate the concept on both the 3d and 2d linear magnetostatic problems.

There is already an exhaustive literature on weak formulations of electromagnetic problems, see KŘÍŽEK AND NEITTAANMÄKI [115], AMROUCHE ET AL. [9], ADAM ET AL. [3], BOSSAVIT [26] VAN RIENEN [211], KAČUR, NEČAS, POLÁK, AND SOUČEK [104], MONK [136, 137, 138], HIPTMAIR [93], NEITTAANMÄKI AND SARANEN [146, 147], STEELE [199], SILVESTER AND FERRARI [192]. The monograph by GIRAULT AND RAVIART [67] and the paper by HIPTMAIR [92] inspired us to build an abstract theoretical framework for the weak formulations and consequent finite element discretizations of the linear elliptic boundary vector–value problems.

3.1 Preliminaries from linear functional analysis

This section recalls basic definitions from functional analysis which will be frequently used in the sequel. Most definitions as well as notation are due to KŘÍŽEK AND NEITTAANMÄKI [115, p. 12]. Let us mention the monographs by NEČAS [141], ODEN AND DEMKOWICZ [149], RUDIN [177], COURANT AND HILBERT [51], SHOWALTER [189], FRANČŮ [64], or by REKTORYS [170].

3.1.1 Normed linear vector spaces

The nonempty set V with the operations $+$: $V \times V \mapsto V$ and \cdot : $\mathbb{R} \times V \mapsto V$ is called a *linear vector space* over reals if for any $u, v, w \in V$ and any $\alpha, \beta \in \mathbb{R}$ the following axioms are satisfied

$$\begin{aligned} (u + v) + w &= u + (v + w), & u + v &= v + u, & \exists z \in V : u + z &= v, \\ \alpha(u + v) &= \alpha u + \alpha v, & (\alpha + \beta)u &= \alpha u + \beta u, & \alpha(\beta u) &= (\alpha\beta)u, \\ 1u &= u. \end{aligned}$$

Among others, the axioms imply that the following hold

$$\exists 0 \in V \forall u \in V : 0 + u = u, \quad \forall u \in V \exists (-u) \in V : u + (-u) = 0.$$

We define the operation $- : V \times V \mapsto V$ by

$$u - v := u + (-v), \quad u, v \in V.$$

The subset $U \subset V$ is called a *subspace* of V if it is also a linear vector space with respect to the operations \cdot and $+$.

Let V be a linear vector space. The mapping $\|\cdot\|_V : V \mapsto \mathbb{R}$ is called a *norm* if for any $u, v \in V$ and any $\alpha \in \mathbb{R}$ the relations

$$\|u + v\|_V \leq \|u\|_V + \|v\|_V, \quad \|\alpha u\|_V = |\alpha| \|u\|_V, \quad \|u\|_V \neq 0 \text{ if } u \neq 0 \quad (3.1)$$

hold. The space V equipped with a norm is called a *normed linear vector space*.

Let V be a normed linear vector space. The sequence $\{u_n\}_{n=1}^{\infty} \subset V$ is said to be *convergent* if there exists $u \in V$ such that

$$\|u_n - u\|_V \rightarrow 0, \text{ as } n \rightarrow \infty.$$

We denote it by $u_n \rightarrow u$ in V .

Let $M \subset V$ be a subset of a normed linear vector space V . The subset M is said to be *closed* if for any convergent sequence $\{u_n\}_{n=1}^{\infty} \subset M$ the following is true

$$u_n \rightarrow u \text{ in } V \Rightarrow u \in M.$$

The subset M is said to be *dense* in V if the condition

$$\forall u \in V \exists \{u_n\}_{n=1}^{\infty} \subset M : u_n \rightarrow u \text{ in } V$$

is satisfied. We denote it by

$$V = \overline{M} \text{ in the norm } \|\cdot\|_V.$$

Let V be a normed linear vector space and $U \subset V$ be a subspace. The space

$$V/U := \{[u] \subset V \mid u \in V \text{ and } \forall v \in U : u + v \in [u]\}$$

is called a *quotient space*. The space V/U equipped with the norm

$$\|[u]\|_{V/U} := \inf_{v \in U} \|u + v\|_V$$

forms a normed linear vector space. Moreover, if U is a closed subspace, then the infimum is realized on U and it becomes the minimum.

3.1.2 Linear operators

Let U, V be normed linear vector spaces. Then the mapping $L : U \mapsto V$ is called a *linear operator* if for any $u, v \in U$ and any $\alpha \in \mathbb{R}$ the following relations

$$L(u + v) = L(u) + L(v), \quad L(\alpha u) = \alpha L(u)$$

hold. The linear operator $L : U \mapsto V$ is *continuous* if the following is satisfied

$$\exists C > 0 \forall u \in U : \|L(u)\|_V \leq C \|u\|_U.$$

The set

$$\text{Ker}(L) := \{u \in U \mid L(u) = 0\} \quad (3.2)$$

is called the *kernel* of the operator L and it is a closed subspace of U . The linear operator $L^{-1} : V \mapsto U$ is called the *inverse* to L if

$$\forall u \in U \forall v \in V : L(u) = v \Leftrightarrow L^{-1}(v) = u.$$

The mapping $f : U \mapsto \mathbb{R}$ is called a *functional*. The space of continuous linear functionals that are defined on a normed linear vector space U is called a *dual space* and it is denoted by U' . The mapping $\langle \cdot, \cdot \rangle : U' \times U \mapsto \mathbb{R}$ defined by

$$\langle f, u \rangle := f(u), \quad f \in U', \quad u \in U,$$

is called a *duality pairing*. The following

$$\|f\|_{U'} := \sup_{\substack{u \in U \\ \|u\|_U=1}} |\langle f, u \rangle|$$

is a norm. The space U' equipped with $\|\cdot\|_{U'}$ and with the following operations

$$\langle f + g, u \rangle := \langle f, u \rangle + \langle g, u \rangle, \quad \langle \alpha f, u \rangle := \alpha \langle f, u \rangle \quad f, g \in U', \quad \alpha \in \mathbb{R}, \quad u \in U,$$

forms a normed linear vector space.

3.1.3 Hilbert spaces

The normed linear space V is called a *Banach space* if for any *Cauchy sequence* $\{u_n\}_{n=1}^{\infty} \subset V$, i.e.,

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall m, n \in \mathbb{N} : m, n \geq n_0 \Rightarrow \|u_m - u_n\|_V \leq \epsilon,$$

the following holds

$$\exists u \in V : u_n \rightarrow u \text{ in } V.$$

Let H be a linear vector space. The mapping $(\cdot, \cdot)_H : H \times H \mapsto \mathbb{R}$ which satisfies for any $u, v, w \in H$ and any $\alpha \in \mathbb{R}$ the following conditions

$$\begin{aligned} (\alpha(u + v), w)_H &= \alpha(u, w)_H + \alpha(v, w)_H, & (u, v)_H &= (v, u)_H, \\ (u, u)_H &\geq 0, & (u, u)_H &\neq 0 \text{ if } u \neq 0 \end{aligned}$$

is called a *scalar product*. The norm defined by

$$\|u\|_H := \sqrt{(u, u)_H}$$

is called the *induced norm*. Moreover, if the space H with the scalar product and the induced norm is a Banach space, then it is called a *Hilbert space*. The following *Cauchy–Schwarz inequality* holds:

$$|(u, v)_H| \leq \|u\|_H \|v\|_H \quad (3.3)$$

for any $u, v \in H$. Let $U \subset H$ be a closed subspace of H , then it is a Hilbert space, too.

Theorem 3.1. (*Riesz theorem*) Let H be a Hilbert space. Then for any $f \in H'$ there exists exactly one element $u \in H$ such that

$$\forall v \in H : (v, u)_H = f(v). \quad (3.4)$$

Moreover,

$$\|u\|_H = \|f\|_{H'}. \quad (3.5)$$

Proof. See ODEN AND DEOMKOWICZ [149, p. 557]. \square

Let H be a Hilbert space. The mapping $a(\cdot, \cdot) : H \times H \mapsto \mathbb{R}$ is called a *bilinear form* if for any fixed $u \in H$ both the mappings $a(\cdot, u)$ and $a(u, \cdot)$ are linear functionals. The bilinear form is said to be *continuous* on H if there exists a positive constant C_1 such that

$$\forall u, v \in H : |a(u, v)| \leq C_1 \|u\|_H \|v\|_H.$$

The bilinear form is called *H-elliptic* if there exists a positive constant C_2 such that

$$\forall v \in H : |a(v, v)| \geq C_2 \|v\|_H^2 \quad (3.6)$$

Lemma 3.1. (*Lax–Milgram lemma*) Let H be a Hilbert space and let $a(\cdot, \cdot)$ be a continuous bilinear form on H which is H -elliptic with the constant C_2 . Then for any $f \in H'$ there exists exactly one element $u \in V$ such that

$$\forall v \in H : a(v, u) = f(v). \quad (3.7)$$

Moreover,

$$\|u\|_H \leq \frac{1}{C_2} \|f\|_{H'}.$$

Proof. See NEČAS [141, p. 38]. \square

Lemma 3.2. Let the assumptions of Lemma 3.1 be satisfied and let the bilinear form be, in addition, symmetric on H , i.e.,

$$\forall u, v \in H : a(u, v) = a(v, u).$$

Then (3.7) is equivalent to: Find $u \in H$ such that

$$J(u) = \min_{v \in H} J(v),$$

where J is a quadratic functional given by

$$J(v) := \frac{1}{2} a(v, v) - f(v), \quad v \in H.$$

Proof. See KRÍŽEK AND NEITTAANMÄKI [115, p. 14]. \square

The normed linear vector spaces U and V are said to be *isomorphically isometric* if there exists a one-to-one linear operator $L : U \mapsto V$ such that

$$\forall u \in U : \|L(u)\|_V = \|u\|_U.$$

The operator L is called an *isomorphism*.

Let H be a Hilbert space and $U \subset H$ be a closed subspace. The space U^\perp defined by

$$U^\perp := \{u \in H \mid \forall v \in U : (u, v)_H = 0\} \quad (3.8)$$

is called the *complementary space* to U and the *orthogonal decomposition*

$$H = U \oplus U^\perp$$

holds, which means that

$$\forall u \in H \exists v \in U \exists w \in U^\perp : u = v + w.$$

Let H be a Hilbert space. We say that the set $E := \{e_i \in H \mid i \in \mathbb{N}\}$ forms a *base* of the space H if the following two assumptions are fulfilled

- (i) $\forall u \in H \forall i \in \mathbb{N} \exists \alpha_i \in \mathbb{R} : u = \sum_{i=1}^{\infty} \alpha_i e_i$,
- (ii) $\forall i \in \mathbb{N} \forall \alpha_i \in \mathbb{R} : \sum_{i=1}^{\infty} \alpha_i e_i = 0 \Rightarrow \alpha_i = 0$.

The vectors e_i are the *base vectors* and the real numbers α_i are the *coordinates* of the vector u in the base E . If the base consists of only a finite number of base vectors, we say that H is *finite-dimensional*, otherwise, H is *infinite-dimensional*.

3.1.4 Linear algebra

Linear algebra, cf. GOLUB AND VAN LOAN [69] or DOSTÁL [57], is a special case of the linear functional analysis, where we work with finite-dimensional Hilbert spaces – the *Euclidean spaces*. By the *Euclidean space* \mathbb{R}^n , $n \in \mathbb{N}$, we mean the Hilbert space \mathbb{R}^n equipped with the scalar product

$$(\mathbf{u}, \mathbf{v}) := \mathbf{u} \cdot \mathbf{v}, \quad \mathbf{u} \cdot \mathbf{v} := \sum_{i=1}^n u_i v_i,$$

where $\mathbf{u} := (u_1, \dots, u_n) \in \mathbb{R}^n$, $\mathbf{v} := (v_1, \dots, v_n) \in \mathbb{R}^n$ stand for column vectors. Then the set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ forms the *Euclidean base*, where all the entries of the *Euclidean base vector* $\mathbf{e}_i \in \mathbb{R}^n$ are zeros except for the i -th entry which is one.

Let $\mathcal{A} := (\mathcal{A}_1, \dots, \mathcal{A}_m) : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a linear vector operator acting between two Euclidean spaces. Then we can represent \mathcal{A} by the following *matrix*

$$\mathbf{A} := \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}, \quad \text{where } a_{i,j} := \mathcal{A}_i(\mathbf{e}_j) \text{ for } i = 1, \dots, m, j = 1, \dots, n.$$

We will also denote the matrix by $\mathbf{A} := (a_{i,j}) \equiv (a_{i,j})_{i,j} \in \mathbb{R}^{m \times n}$. From the linearity of \mathcal{A} it follows that for a vector $\mathbf{u} := (u_1, u_2, \dots, u_n) \in \mathbb{R}^n$

$$\mathcal{A}(\mathbf{u}) = \mathbf{A} \cdot \mathbf{u}, \quad \text{where } \mathbf{A} \cdot \mathbf{u} := \begin{pmatrix} \sum_{j=1}^n a_{1,j} u_j \\ \sum_{j=1}^n a_{2,j} u_j \\ \vdots \\ \sum_{j=1}^n a_{m,j} u_j \end{pmatrix}.$$

We define the *matrix norm* $\|\cdot\| : \mathbb{R}^{m \times n} \mapsto \mathbb{R}$ by

$$\|\mathbf{A}\| := \max_{i,j} |a_{i,j}|. \tag{3.9}$$

By the matrix $\mathbf{A}^T := (a_{i,j}^T) \in \mathbb{R}^{n \times m}$ we denote the *transpose matrix* to the matrix \mathbf{A} , where the entries of \mathbf{A}^T are as follows:

$$a_{i,j}^T := a_{j,i} \quad \text{for } i = 1, \dots, n, \quad j = 1, \dots, m.$$

The mapping $\mathcal{A}^T : \mathbb{R}^m \mapsto \mathbb{R}^n$, which is represented by the matrix \mathbf{A}^T , is called the *transpose mapping* to the mapping \mathcal{A} .

Let $\mathcal{A} := (\mathcal{A}_1, \dots, \mathcal{A}_m) : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $\mathcal{B} := (\mathcal{B}_1, \dots, \mathcal{B}_p) : \mathbb{R}^m \mapsto \mathbb{R}^p$ be linear vector operators, represented by the matrices $\mathbf{A} := (a_{i,j}) \in \mathbb{R}^{m \times n}$ and $\mathbf{B} := (b_{i,j}) \in \mathbb{R}^{p \times m}$, respectively. Then it can be easily proven that the *compound mapping* $\mathcal{C} := \mathcal{B} \circ \mathcal{A} : \mathbb{R}^n \mapsto \mathbb{R}^p$, defined by

$$[\mathcal{B} \circ \mathcal{A}](\mathbf{u}) := \mathcal{B}(\mathcal{A}(\mathbf{u})), \quad \mathbf{u} \in \mathbb{R}^n,$$

is represented by the matrix $\mathbf{C} := (c_{i,j}) \in \mathbb{R}^{p \times n}$, where

$$c_{i,j} := \sum_{k=1}^m b_{i,k} a_{k,j} \quad \text{for } i = 1, \dots, p, \quad j = 1, \dots, n.$$

The linear algebra provides a powerful tool for solving linear operator equations. Given a linear mapping $\mathcal{A} : \mathbb{R}^n \mapsto \mathbb{R}^m$ and a vector $\mathbf{f} := (f_1, \dots, f_m) \in \mathbb{R}^m$, the *linear operator equation*

$$\mathcal{A}(\mathbf{u}) = \mathbf{f}, \quad (3.10)$$

solved for $\mathbf{u} \in \mathbb{R}^n$, can be equivalently written as a *system of linear algebraic equations*, the matrix form of which is

$$\mathbf{A} \cdot \mathbf{u} = \mathbf{f}, \quad (3.11)$$

where the matrix $\mathbf{A} := (a_{i,j}) \in \mathbb{R}^{m \times n}$ represents the linear operator \mathcal{A} . Moreover, if $m = n$ and if there exists the inverse operator $\mathcal{A}^{-1} : \mathbb{R}^n \mapsto \mathbb{R}^n$, then the solution to the linear operator equation (3.10) is represented by

$$\mathbf{u} = \mathcal{A}^{-1}(\mathbf{f}).$$

The latter can be again written in terms of matrices. To this end, we introduce a multilinear form $\det(\mathbf{A})$, called the *determinant* of the matrix \mathbf{A} , which is recursively defined by

$$\det(\mathbf{A}) := \begin{cases} a_{1,1} & , n = 1 \\ \sum_{j=1}^n (-1)^{j+1} \det(\mathbf{A}_{1,j}) & , n \geq 2 \end{cases}, \quad (3.12)$$

where the matrix $\mathbf{A}_{i,j} \in \mathbb{R}^{(n-1) \times (n-1)}$ is made from the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ by excluding its i -th row and j -th column. Further, to the matrix \mathbf{A} we associate the *adjoint matrix* $\tilde{\mathbf{A}} := (\tilde{a}_{i,j}) \in \mathbb{R}^{n \times n}$ by

$$\tilde{a}_{i,j} := (-1)^{i+j} \det(\mathbf{A}_{j,i}). \quad (3.13)$$

Lemma 3.3. *Let $\mathcal{A} : \mathbb{R}^n \mapsto \mathbb{R}^n$, $n \in \mathbb{N}$, be a linear operator, which is represented by the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then there exists the inverse linear operator $\mathcal{A}^{-1} : \mathbb{R}^n \mapsto \mathbb{R}^n$ if and only if $\det(\mathbf{A}) \neq 0$. The corresponding inverse matrix is then as follows:*

$$\mathbf{A}^{-1} := \frac{1}{\det(\mathbf{A})} \tilde{\mathbf{A}} \quad (3.14)$$

and it is such that

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{I},$$

where $\mathbf{I} := [\mathbf{e}_1, \dots, \mathbf{e}_n] \in \mathbb{R}^{n \times n}$, where $\mathbf{e}_i \in \mathbb{R}^n$ denotes an Euclidean base vector.

Proof. See GOLUB AND VAN LOAN [69]. \square

Let a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ be given. If there does not exist the inverse matrix $\mathbf{A}^{-1} \in \mathbb{R}^{n \times n}$, then \mathbf{A} is said to be a *singular* matrix. Otherwise, \mathbf{A} is *nonsingular* and then the solution to the system of linear algebraic equations (3.11) reads as follows:

$$\mathbf{u} = \frac{1}{\det(\mathbf{A})} \tilde{\mathbf{A}} \cdot \mathbf{f}. \quad (3.15)$$

Note that (3.15) is extremely inappropriate for practical calculations of \mathbf{u} , since the computation of both $\det(\mathbf{A})$ and $\tilde{\mathbf{A}}$ is very time-consuming. We will rather use (3.15) for analysis only.

Given a nonsingular matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, the transposition and the inversion are mutually commutative and we abbreviate them as follows:

$$\mathbf{A}^{-T} := (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}.$$

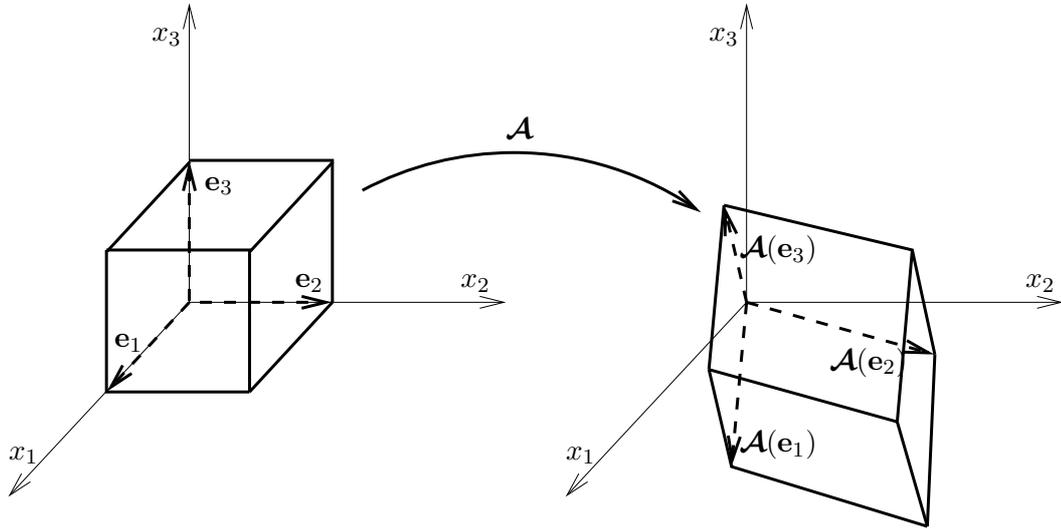


Figure 3.1: Linear transformation of a unit cube

The determinant $\det(\mathbf{A})$ has a clear geometric meaning, which is depicted in Fig. 3.1. If we write the matrix \mathbf{A} columnwise as $\mathbf{A} := (\mathbf{a}_1, \dots, \mathbf{a}_n)$, then the vectors $\mathbf{a}_i := (a_{1,i}, \dots, a_{n,i}) = \mathcal{A}(\mathbf{e}_i)$ are the images of the Euclidean base vectors \mathbf{e}_i in the mapping \mathcal{A} . The vectors \mathbf{e}_i determine an n -th dimensional unit cube, while the value $|\det(\mathbf{A})|$ is the n -th dimensional volume of the image of this cube after the transformation \mathcal{A} . This determinant property is for example used, when a substitution is employed in some n -th dimensional, e.g., volume integration.

3.2 Preliminaries from real analysis

A *domain* is due to HASLINGER AND NEITTAANMÄKI [85, p. 2] a bounded, open, and connected set $\Omega \subset \mathbb{R}^m$, $m \in \mathbb{N}$. The symbol $\bar{\Omega}$ stands for the closure of Ω , $\partial\Omega$ is the boundary of Ω , and \mathbf{n} denotes the unit outer normal vector to the boundary $\partial\Omega$. As only two- or three-dimensional domains are meaningful for optimal shape design, we employ the following:

Assumption 3.1. *In all what follows we will assume that Ω denotes a nonempty, bounded, open, and connected subset of \mathbb{R}^m , where $m \in \{2, 3\}$.*

Nevertheless, all the results up to Chapter 5 are valid for any $m \in \mathbb{N}$.

3.2.1 Continuous function spaces

This section is due to HASLINGER AND NEITTAANMÄKI [85, p. 2–4].

A sequence $\{u_n\}_{n=1}^{\infty}$ of real-valued functions defined in Ω is *uniformly bounded* in Ω if there exists a constant $C > 0$ such that

$$\forall n \in \mathbb{N} \forall \mathbf{x} \in \Omega : |u_n(\mathbf{x})| \leq C.$$

Let \mathcal{F} be a collection of functions $u : \Omega \rightarrow \mathbb{R}$. We say that the functions belonging to \mathcal{F} and the set \mathcal{F} itself are *equicontinuous at $\mathbf{x} \in \Omega$* if

$$\forall \varepsilon > 0 \exists \delta(\varepsilon, \mathbf{x}) > 0 \forall u \in \mathcal{F} : \|\mathbf{y} - \mathbf{x}\| < \delta(\varepsilon, \mathbf{x}) \Rightarrow |u(\mathbf{y}) - u(\mathbf{x})| \leq \varepsilon.$$

Functions are *equicontinuous in Ω* if they are equicontinuous at any $\mathbf{x} \in \Omega$.

Let $\{u_n\}_{n=1}^{\infty}$ be a sequence of functions and let u be a function, all defined over $\bar{\Omega}$. We say that u_n *uniformly converges* to u if

$$\max_{\mathbf{x} \in \bar{\Omega}} \{|u_n(\mathbf{x}) - u(\mathbf{x})|\} \rightarrow 0, \text{ as } n \rightarrow \infty, \quad (3.16)$$

and we denote this convergence by $u_n \rightrightarrows u$ in $\bar{\Omega}$, as $n \rightarrow \infty$.

Theorem 3.2. (*Ascoli–Arzelà*) *Let $\{u_n\}_{n=1}^{\infty}$ be a set of uniformly bounded equicontinuous functions in $\bar{\Omega}$, $u_n : \Omega \mapsto \mathbb{R}$. Then there exists a subsequence $\{u_{n_k}\}_{k=1}^{\infty} \subset \{u_n\}_{n=1}^{\infty}$ and a function u (continuous in $\bar{\Omega}$) such that $u_{n_k} \rightrightarrows u$ in $\bar{\Omega}$, as $k \rightarrow \infty$.*

Proof. See ODEN AND DEMKOWICZ [149, p. 365]. □

Let $k \in \mathbb{N} \cup \{0\}$. The symbol $C^k(\bar{\Omega})$ denotes the space of continuous real-valued functions that are *differentiable up to the order k* . In particular, we denote the space of continuous functions by

$$C(\bar{\Omega}) := C^0(\bar{\Omega}),$$

which, being equipped with the norm

$$\|u\|_{C(\bar{\Omega})} := \max_{\mathbf{x} \in \bar{\Omega}} |u(\mathbf{x})|,$$

forms a Banach space.

Lemma 3.4. *Let $\omega \subset \mathbb{R}^l$ and $\Omega \subset \mathbb{R}^m$ be domains, where $l, m \in \mathbb{N}$. Let $\mathbf{u} := (u_1, \dots, u_m) \in [C^k(\bar{\omega})]^m$ and $\mathbf{v} := (v_1, \dots, v_n) \in [C^k(\bar{\Omega})]^n$, $n \in \mathbb{N}$, be vector functions continuously differentiable up to the order $k \in \mathbb{N}$, and let*

$$\forall \mathbf{x} \in \bar{\omega} : \mathbf{u}(\mathbf{x}) \in \bar{\Omega}.$$

Then $\mathbf{v} \circ \mathbf{u} \in [C^k(\bar{\omega})]^n$, where for $\mathbf{x} \in \bar{\omega}$ the function $\mathbf{v} \circ \mathbf{u}$ is defined by

$$(\mathbf{v} \circ \mathbf{u})(\mathbf{x}) := \mathbf{v}(\mathbf{u}(\mathbf{x})).$$

Moreover, partial derivatives of the compound function are as follows:

$$\frac{\partial (\mathbf{v} \circ \mathbf{u})_i(\mathbf{x})}{\partial x_j} = \sum_{k=1}^m \frac{\partial v_i(\mathbf{y})}{\partial y_k} \frac{\partial u_k(\mathbf{x})}{\partial x_j}, \quad i = 1, \dots, n, \quad j = 1, \dots, l, \quad (3.17)$$

where $\mathbf{x} := (x_1, \dots, x_l) \in \bar{\omega}$ and $\mathbf{y} := (y_1, \dots, y_m) := (u_1(\mathbf{x}), \dots, u_m(\mathbf{x})) \in \bar{\Omega}$.

Proof. See RUDIN [178, p. 86]. □

Further, we introduce the space of *infinitely differentiable functions* by

$$C^\infty(\bar{\Omega}) := \bigcap_{k=1}^{\infty} C^k(\bar{\Omega})$$

and the space of *infinitely differentiable functions with a compact support* by

$$C_0^\infty(\Omega) := \{v \in C^\infty(\bar{\Omega}) \mid \text{supp } v \subset \Omega\},$$

where

$$\text{supp } v := \overline{\{\mathbf{x} \in \Omega \mid v(\mathbf{x}) \neq 0\}}.$$

We introduce the space of *Lipschitz continuous functions* by

$$C^{0,1}(\bar{\Omega}) := \{u \in C(\bar{\Omega}) \mid \exists C > 0 \forall \mathbf{x}, \mathbf{y} \in \bar{\Omega} : |u(\mathbf{x}) - u(\mathbf{y})| \leq C \|\mathbf{x} - \mathbf{y}\|\},$$

where $C > 0$ is a *Lipschitz constant*.

Now we define a class of domains of a more practical use. The following definition is due to KŘÍŽEK AND NEITTAANMÄKI [115, p. 17] or HASLINGER AND NEITTAANMÄKI [85, p. 4].

Definition 3.1. A nonempty domain $\Omega \subset \mathbb{R}^m$ is said to have a *Lipschitz continuous boundary* if for any $\mathbf{z} \in \partial\Omega$ there exists a neighbourhood $U := U(\mathbf{z})$ such that the set $U \cap \Omega$ can be expressed, in some Cartesian coordinate system (x_1, \dots, x_m) , by the inequality $x_m < F(x_1, \dots, x_{m-1})$, where F is a Lipschitz continuous function.

The symbol \mathcal{L} denotes the set of all domains with Lipschitz continuous boundaries.

3.2.2 Some fundamental theorems

Here, we refer to the classical textbooks by RUDIN [176, 178]. Let the vector $\alpha := (\alpha_1, \dots, \alpha_m)$ denotes a *multi-index*, where $\alpha_1, \dots, \alpha_m$ are non-negative integers, and let $|\alpha| := \alpha_1 + \dots + \alpha_m$ be an *order of the multi-index*. Then for any $u \in C^k(\bar{\Omega})$ and any $\mathbf{x} \in \Omega$ we define the α -th *classical derivative*, $\alpha \leq k$, at the point \mathbf{x} as follows:

$$D^\alpha u(\mathbf{x}) := \begin{cases} \frac{\partial^{|\alpha|} u(\mathbf{x})}{\partial x_1^{\alpha_1} \dots \partial x_m^{\alpha_m}} & , |\alpha| \in \mathbb{N} \\ u(\mathbf{x}) & , \alpha = (0, \dots, 0) \end{cases}.$$

The following classical theorems of real analysis are due to RUDIN [178].

Theorem 3.3. (*Taylor's theorem*) Let Ω be an open subset of \mathbb{R}^m , $m \in \mathbb{N}$, and let $u \in C^k(\bar{\Omega})$. Let further $\mathbf{x} := (x_1, \dots, x_m) \in \Omega$ and let $\mathbf{z} \in \mathbb{R}^m$ be such that

$$\forall t \in \mathbb{R} : 0 \leq t \leq 1 \Rightarrow \mathbf{x} + t\mathbf{z} \in \bar{\Omega}.$$

Then

$$u(\mathbf{x} + \mathbf{z}) = \sum_{i=1}^{k-1} \frac{1}{i!} \sum_{0 \leq |\alpha| \leq i} D^\alpha u(\mathbf{x}) \prod_{j=1}^m (x_j)^{\alpha_j} + r(\mathbf{z}),$$

where $\alpha := (\alpha_1, \dots, \alpha_m)$ denotes a multi-index,

$$i! := \begin{cases} \prod_{j=1}^k j & , i \in \mathbb{N} \\ 1 & , i = 0 \end{cases}$$

denotes the factorial of $i \in \mathbb{N} \cup \{0\}$, and where the remainder function $r : \Omega \mapsto \mathbb{R}$ satisfies

$$\lim_{\mathbf{z} \rightarrow 0} \frac{r(\mathbf{z})}{|\mathbf{z}|^{k-1}} = 0.$$

Proof. See RUDIN [178, Exercise 30]. □

Theorem 3.4. (Green's theorem) Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$, and let $u, v \in C^1(\overline{\Omega})$. Then, the relation

$$\int_{\Omega} \frac{\partial u}{\partial x_i} v \, d\mathbf{x} + \int_{\Omega} u \frac{\partial v}{\partial x_i} \, d\mathbf{x} = \int_{\partial\Omega} u v n_i \, ds \quad \text{for } i = 1, \dots, m$$

holds, where $\mathbf{n} := (n_1, \dots, n_m)$ denotes the outer unit normal to $\partial\Omega$.

Proof. See Exercise 1.22.3 in ODEN AND DEMKOWICZ [149, p. 120] for the case $m = 2$ and $\partial\Omega$ being a (closed) C^1 curve. □

Corollary 3.1. Let the assumptions on Ω and u hold and let $\mathbf{v} \in [C^1(\overline{\Omega})]^m$. Then the following is satisfied

$$\int_{\Omega} \mathbf{grad}(u) \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} u \operatorname{div}(\mathbf{v}) \, d\mathbf{x} = \int_{\partial\Omega} (u\mathbf{n}) \cdot \mathbf{v} \, ds,$$

where the differential operators \mathbf{grad} and div are respectively defined as follows:

$$\mathbf{grad}(u) := \left(\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_m} \right), \quad u \in C^1(\overline{\Omega}),$$

$$\operatorname{div}(\mathbf{v}) := \sum_{i=1}^m \frac{\partial v_i}{\partial x_i}, \quad \mathbf{v} := (v_1, \dots, v_m) \in [C^1(\overline{\Omega})]^m.$$

Proof. Denote $\mathbf{v} := (v_1, \dots, v_m)$, then, for each $i = 1, \dots, m$ Green's formula

$$\int_{\Omega} \frac{\partial u}{\partial x_i} v_i \, d\mathbf{x} + \int_{\Omega} u \frac{\partial v_i}{\partial x_i} \, d\mathbf{x} = \int_{\partial\Omega} (u n_i) v_i \, ds$$

holds. Summing up the latter for the index $i = 1, \dots, m$, we get the assertion. □

Corollary 3.2. Let $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$, and let $\mathbf{u}, \mathbf{v} \in [C^1(\overline{\Omega})]^3$. Then the following is satisfied

$$\int_{\Omega} \mathbf{curl}(\mathbf{u}) \cdot \mathbf{v} \, d\mathbf{x} - \int_{\Omega} \mathbf{u} \cdot \mathbf{curl}(\mathbf{v}) \, d\mathbf{x} = \int_{\partial\Omega} (\mathbf{n} \times \mathbf{u}) \cdot \mathbf{v} \, ds, \quad (3.18)$$

where for $\mathbf{u} := (u_1, u_2, u_3) \in [C^1(\Omega)]^3$ the differential operators \mathbf{curl} is defined by

$$\mathbf{curl}(\mathbf{u}) := \left(\frac{\partial u_3}{\partial x_2} - \frac{\partial u_2}{\partial x_3}, \frac{\partial u_1}{\partial x_3} - \frac{\partial u_3}{\partial x_1}, \frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2} \right), \quad (3.19)$$

and the cross product is as follows:

$$\mathbf{n} \times \mathbf{u} := (n_2u_3 - n_3u_2, n_3u_1 - n_1u_3, n_1u_2 - n_2u_1), \quad (3.20)$$

where $\mathbf{n} := (n_1, n_2, n_3)$ denotes the outer unit normal vector to $\partial\Omega$.

Proof. Using the definitions (3.19), (3.20), and Theorem 3.4, we can easily see that the relation (3.18) holds. \square

We will extend Theorem 3.4 for a general linear differential operator of the first order. To this end we define the following.

Definition 3.2. Suppose $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. Let $\mathbf{B} : [C^1(\overline{\Omega})]^{\nu_1} \mapsto [C(\overline{\Omega})]^{\nu_2}$, $\nu_1, \nu_2 \in \mathbb{N}$, be a linear differential operator of the first order defined by

$$\mathbf{B}(\mathbf{u}) := (B_1(\mathbf{u}), \dots, B_{\nu_2}(\mathbf{u})), \text{ where } B_i(\mathbf{u}) := \sum_{k=1}^{\nu_1} \sum_{l=1}^m b_i^{(k,l)} \frac{\partial u_k}{\partial x_l}, \quad (3.21)$$

where $b_i^{(k,l)} \in \mathbb{R}$ for $i = 1, \dots, \nu_2$, and where $\mathbf{u} := (u_1, \dots, u_{\nu_1}) \in [C^1(\overline{\Omega})]^{\nu_1}$. We define the adjoint operator $\mathbf{B}^* : [C^1(\overline{\Omega})]^{\nu_2} \mapsto [C(\overline{\Omega})]^{\nu_1}$ to the operator \mathbf{B} by

$$\mathbf{B}^*(\mathbf{v}) := (B_1^*(\mathbf{v}), \dots, B_{\nu_1}^*(\mathbf{v})), \text{ where } B_k^*(\mathbf{v}) := \sum_{i=1}^{\nu_2} \sum_{l=1}^m b_i^{(k,l)} \frac{\partial v_i}{\partial x_l} \text{ for } k = 1, \dots, \nu_1, \quad (3.22)$$

and where $\mathbf{v} := (v_1, \dots, v_{\nu_2}) \in [C^1(\overline{\Omega})]^{\nu_2}$.

Moreover, we define the trace operator $\gamma : [C(\overline{\Omega})]^{\nu_1} \mapsto [C(\partial\Omega)]^{\nu_2}$ associated to \mathbf{B} by

$$\gamma(\mathbf{u}) := (\gamma_1(\mathbf{u}), \dots, \gamma_{\nu_2}(\mathbf{u})), \text{ where } \gamma_i(\mathbf{u}) := \sum_{k=1}^{\nu_1} \sum_{l=1}^m b_i^{(k,l)} u_k|_{\partial\Omega} n_l \text{ for } i = 1, \dots, \nu_2, \quad (3.23)$$

where $\mathbf{n} := (n_1, \dots, n_m)$ is the outer unit normal to $\partial\Omega$.

The following is a consequence of Theorem 3.4.

Corollary 3.3. Let the assumptions and notation of the previous definition are fulfilled. Let $\mathbf{u} := (u_1, \dots, u_{\nu_1}) \in [C^1(\overline{\Omega})]^{\nu_1}$ and $\mathbf{v} := (v_1, \dots, v_{\nu_2}) \in [C^1(\overline{\Omega})]^{\nu_2}$. Then the following is satisfied

$$\int_{\Omega} \mathbf{B}(\mathbf{u}) \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \mathbf{u} \cdot \mathbf{B}^*(\mathbf{v}) \, d\mathbf{x} = \int_{\partial\Omega} \gamma(\mathbf{u}) \cdot \mathbf{v} \, ds. \quad (3.24)$$

Proof. Let $\mathbf{u} \in [C^1(\overline{\Omega})]^{\nu_1}$ and $\mathbf{v} \in [C^1(\overline{\Omega})]^{\nu_2}$ be arbitrary. Using the previous definition, we write down the left-hand side of (3.24)

$$\begin{aligned} \int_{\Omega} \mathbf{B}(\mathbf{u}) \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \mathbf{u} \cdot \mathbf{B}^*(\mathbf{v}) \, d\mathbf{x} &= \sum_{k=1}^{\nu_1} \sum_{i=1}^{\nu_2} \sum_{l=1}^m b_i^{(k,l)} \left(\int_{\Omega} \frac{\partial u_k}{\partial x_l} v_i \, d\mathbf{x} + \int_{\Omega} u_k \frac{\partial v_i}{\partial x_l} \, d\mathbf{x} \right) = \\ &= \sum_{k=1}^{\nu_1} \sum_{i=1}^{\nu_2} \sum_{l=1}^m b_i^{(k,l)} \int_{\partial\Omega} u_k v_i n_l \, ds = \int_{\partial\Omega} \gamma(\mathbf{u}) \cdot \mathbf{v} \, ds, \end{aligned}$$

where we used Theorem 3.4. \square

Theorem 3.5. (Stokes' theorem) Let $\Omega \subset \mathbb{R}^3$ be an open set and let $\mathbf{u} := (u_1, u_2, u_3) \in [C^1(\bar{\omega})]^3$, where ω is a 2-dimensional surface with the boundary $\partial\omega$ being a piecewise C^1 curve. Then the following is satisfied

$$\int_{\omega} \operatorname{curl}(\mathbf{u}) \cdot \mathbf{n} \, d\mathbf{x} = \int_{\partial\omega} \mathbf{u} \times \mathbf{n} \, ds,$$

where $\mathbf{n} := (n_1, n_2, n_3)$ denotes the outer unit normal to $\partial\omega$.

Proof. See RUDIN [178, p. 287] □

3.3 Hilbert function spaces

3.3.1 Lebesgue spaces

Let $\Omega \in \mathcal{L}$. In the sequel all the integrals will be understood in the Lebesgue's sense, cf. LUKEŠ AND MALÝ [129]. We introduce the *Lebesgue spaces* of real-valued functions

$$L^p(\Omega) := \left\{ u : \Omega \rightarrow \mathbb{R} \mid \int_{\Omega} |u|^p \, d\mathbf{x} < +\infty \right\}, \quad p \in [1, \infty),$$

equipped with the norm

$$\|u\|_{L^p(\Omega)} := \left(\int_{\Omega} |u|^p \, d\mathbf{x} \right)^{1/p}, \quad u \in L^p(\Omega).$$

Let $p, q \in (1, \infty)$ be adjoint by

$$\frac{1}{p} + \frac{1}{q} = 1$$

and let $u \in L^p(\Omega)$, $v \in L^q(\Omega)$, then the following *Hölder inequality* holds

$$\left| \int_{\Omega} uv \, d\mathbf{x} \right| = \left(\int_{\Omega} |u|^p \, d\mathbf{x} \right)^{1/p} \left(\int_{\Omega} |v|^q \, d\mathbf{x} \right)^{1/q}. \quad (3.25)$$

The Lebesgue space of measurable essentially bounded functions is defined by

$$L^\infty(\Omega) := \left\{ u : \Omega \rightarrow \mathbb{R} \mid \operatorname{ess\,sup}_{x \in \Omega} |u(x)| < +\infty \right\}$$

and it is equipped with the norm

$$\|u\|_{L^\infty(\Omega)} := \operatorname{ess\,sup}_{x \in \Omega} |u(x)|, \quad u \in L^\infty(\Omega).$$

We say that the set Ω is *measurable* in the Lebesgue sense if the following Lebesgue integral exists

$$\operatorname{meas}(\Omega) := \int_{\Omega} d\mathbf{x},$$

and we call it the *measure* of Ω .

Let $\Omega \in \mathcal{L}$. We say that the function $u \in L^1(\Omega)$ is defined *almost everywhere* (a.e.) if it is defined for each $\mathbf{x} \in \Omega \setminus \omega$, where the subset $\omega \subset \Omega$ is such that $\operatorname{meas}(\omega) = 0$. The notion *almost everywhere* is understood similarly in different contexts, e.g., the sequence $\{u_n\}_{n=1}^\infty \subset L^1(\Omega)$ is said to *converge almost everywhere* to $u : \Omega \rightarrow \mathbb{R}$ if for any $\omega \subset \Omega$ such that $\operatorname{meas}(\omega) = 0$ the following holds

$$u_n(x) \rightarrow u(x) \text{ in } \Omega \setminus \omega.$$

Theorem 3.6. (*Lebesgue dominated convergence theorem*) Let $\{u_n\}_{n=1}^{\infty} \subset L^1(\Omega)$ be a sequence of functions measurable in the Lebesgue's sense. Let $u_n \rightarrow u$ almost everywhere in Ω , where $u : \Omega \rightarrow \mathbb{R}$ is a function. If there exists a function $v \in L^1(\Omega)$ such that $|u_n| \leq v$ almost everywhere in Ω for all $n \in \mathbb{N}$, then $u \in L^1(\Omega)$ and

$$\int_{\Omega} u \, d\mathbf{x} = \lim_{n \rightarrow \infty} \int_{\Omega} u_n \, d\mathbf{x}.$$

Proof. See LUKEŠ AND MALÝ [129, p. 26]. \square

3.3.2 Sobolev spaces

There is a lot of references on this topic. Let us mention the monographs by SOBOLEV [194], NEČAS [141], ADAMS [4], KUFNER, JOHN, AND FUČÍK [116], MAZYA [134], or the paper by DOKTOR [56].

The function $z \in L^2(\Omega)$ is said to be the α -th *generalized derivative* of the function $u \in L^2(\Omega)$ if the following is satisfied

$$\forall v \in C_0^{\infty}(\Omega) : \int_{\Omega} zv \, d\mathbf{x} = (-1)^{|\alpha|} \int_{\Omega} u D^{\alpha}v \, d\mathbf{x}. \quad (3.26)$$

We can easily see that for any $u \in C^k(\overline{\Omega})$, $k \in \mathbb{N} \cup \{0\}$, for a multi-index α such that $|\alpha| \leq k$, and for $z := D^{\alpha}u \in C(\overline{\Omega})$, which is the α -th classical derivative of u , the relation (3.26) holds in virtue of Theorem 3.4. Therefore, we can extend the symbol $D^{\alpha}u$ and we denote the α -th generalized derivative still by $D^{\alpha}u := z$.

Now, for $k \in \mathbb{N} \cup \{0\}$ we define the *Sobolev spaces* as follows:

$$H^k(\Omega) := \{u \in L^2(\Omega) \mid \forall \alpha : |\alpha| \leq k \Rightarrow \exists D^{\alpha}u \in L^2(\Omega)\}.$$

The latter, equipped with the scalar product

$$(u, v)_{k, \Omega} := \sum_{|\alpha| \leq k} \int_{\Omega} D^{\alpha}u D^{\alpha}v \, d\mathbf{x}, \quad u, v \in H^k(\Omega),$$

forms a Hilbert space with the following induced norm and seminorm

$$\|u\|_{k, \Omega} := \sqrt{(u, u)_{k, \Omega}}, \quad |u|_{k, \Omega} := \sqrt{\sum_{|\alpha|=k} \int_{\Omega} |D^{\alpha}u|^2 \, d\mathbf{x}}, \quad u \in H^k(\Omega),$$

respectively. The Sobolev spaces of vector functions $[H^k(\Omega)]^n$, $n \in \mathbb{N}$, equipped with the scalar product

$$(\mathbf{u}, \mathbf{v})_{n, k, \Omega} := \sum_{i=1}^n (u_i, v_i)_{k, \Omega}, \quad \mathbf{u}, \mathbf{v} \in [H^k(\Omega)]^n,$$

where $\mathbf{u} := (u_1, \dots, u_n)$ and $\mathbf{v} := (v_1, \dots, v_n)$ are Hilbert spaces, too.

We will make use of some properties of Sobolev spaces. The following theorem gives us an insight how functions behave along the boundary $\partial\Omega$.

Theorem 3.7. (*Trace theorem*) Let $\Omega \in \mathcal{L}$. Then there exists exactly one linear continuous operator $\gamma : H^1(\Omega) \mapsto L^2(\partial\Omega)$ such that

$$\forall u \in C^{\infty}(\overline{\Omega}) : \gamma(u) = u|_{\partial\Omega}.$$

Proof. See KUFNER, JOHN, AND FUČÍK [116, p. 318] or NEČAS [141, p. 15]. \square

The function $\gamma(u)$ is called the *trace* of u . The trace theorem enables us to define the space

$$H_0^1(\Omega) := \{u \in H^1(\Omega) \mid \gamma(u) = 0\}.$$

Finally, we denote the space of traces by

$$H^{1/2}(\partial\Omega) := \{v \in L^2(\partial\Omega) \mid \exists u \in H^1(\Omega) : \gamma(u) = v\}$$

and its dual space by $H^{-1/2}(\partial\Omega)$.

The spaces $[C^\infty(\overline{\Omega})]^n$ and $[C_0^\infty(\Omega)]^n$, respectively, are dense in $[H^1(\Omega)]^n$ and $[H_0^1(\Omega)]^n$, i.e.,

$$[H^1(\Omega)]^n = \overline{[C^\infty(\overline{\Omega})]^n} \text{ and } [H_0^1(\Omega)]^n = \overline{[C_0^\infty(\Omega)]^n} \text{ in the norm } \|\cdot\|_{n,1,\Omega}. \quad (3.27)$$

The next theorem extends Theorem 3.4.

Theorem 3.8. (*Green's theorem*) Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$, and let $u, v \in H^1(\Omega)$. Then the relation

$$\int_{\Omega} \frac{\partial u}{\partial x_i} v \, d\mathbf{x} + \int_{\Omega} u \frac{\partial v}{\partial x_i} \, d\mathbf{x} = \int_{\partial\Omega} \gamma(u) \gamma(v) n_i \, ds \quad \text{for } i = 1, \dots, m$$

holds, where $\mathbf{n} := (n_1, \dots, n_m)$ denotes the outer unit normal to $\partial\Omega$.

Proof. See NEČAS [141, p. 29]. \square

Note that, avoiding some additional effort, yet we have not defined either the boundary integral or the space $L^2(\partial\Omega)$, for which we refer to KUFNER, JOHN, AND FUČÍK [116] or NEČAS [141].

The last theorem, of which we will make use later when analyzing the ellipticity of differential operators in $H^1(\Omega)$, is due to HASLINGER AND NEITTAANMÄKI [85, p. 9] or KŘÍŽEK AND NEITTAANMÄKI [115, p. 26].

Theorem 3.9. (*Friedrichs' inequality*) Let $\Omega \in \mathcal{L}$. Then there exists a positive constant $C_3 \equiv C_3(\Omega)$ such that

$$\forall u \in H_0^1(\Omega) : \|u\|_{1,\Omega} \leq C_3 |u|_{1,\Omega}.$$

Proof. See NEČAS [141, p. 30]. \square

3.3.3 The space $H(\mathbf{grad})$

Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. In the previous section we extended the notion of the partial derivative to the generalized case. Now we extend the differential operator $\mathbf{grad} : C^1(\overline{\Omega}) \mapsto [C(\overline{\Omega})]^m$ onto a subspace of $L^2(\Omega)$. The function $\mathbf{z} \in [L^2(\Omega)]^m$ is said to be the *generalized gradient* of $u \in L^2(\Omega)$ if the following is satisfied

$$\forall \mathbf{v} \in [C_0^\infty(\Omega)]^m : \int_{\Omega} u \operatorname{div}(\mathbf{v}) \, d\mathbf{x} = - \int_{\Omega} \mathbf{z} \cdot \mathbf{v} \, d\mathbf{x},$$

and we denote the generalized gradient by $\mathbf{grad}(u) := \mathbf{z}$. In particular, from (3.26) it is clear that

$$\forall u \in H^1(\Omega) : \mathbf{grad}(u) = \left(\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_m} \right),$$

where the partial derivatives are the generalized ones. We define the space

$$H(\mathbf{grad}; \Omega) := \{u \in L^2(\Omega) \mid \exists \mathbf{z} \in [L^2(\Omega)]^m : \mathbf{z} = \mathbf{grad}(u)\}.$$

The latter together with the scalar product

$$(u, v)_{\mathbf{grad}, \Omega} := \int_{\Omega} uv \, d\mathbf{x} + \int_{\Omega} \mathbf{grad}(u) \cdot \mathbf{grad}(v) \, d\mathbf{x}, \quad u, v \in H(\mathbf{grad}; \Omega),$$

forms a Hilbert space. We introduce the following induced norm and seminorm

$$\|u\|_{\mathbf{grad}, \Omega} := \sqrt{(u, u)_{\mathbf{grad}, \Omega}}, \quad |u|_{\mathbf{grad}, \Omega} := \sqrt{\int_{\Omega} \|\mathbf{grad}(u)\|^2 \, d\mathbf{x}}, \quad u \in H(\mathbf{grad}; \Omega),$$

respectively. Clearly

$$\|u\|_{\mathbf{grad}, \Omega}^2 = \|u\|_{0, \Omega}^2 + |u|_{\mathbf{grad}, \Omega}^2, \quad u \in H(\mathbf{grad}; \Omega),$$

holds. From the definition of $H^1(\Omega)$ it is obvious that $H(\mathbf{grad}; \Omega) = H^1(\Omega)$, $(u, v)_{\mathbf{grad}, \Omega} = (u, v)_{1, \Omega}$, $\|u\|_{\mathbf{grad}, \Omega} = \|u\|_{1, \Omega}$, and $|u|_{\mathbf{grad}, \Omega} = |u|_{1, \Omega}$. Therefore, Theorem 3.7 holds and we can define the space

$$H_0(\mathbf{grad}; \Omega) := \{u \in H(\mathbf{grad}; \Omega) \mid \gamma(u) = \mathbf{0}\},$$

where the trace operator

$$\gamma(u) := \gamma(u)\mathbf{n},$$

where $\mathbf{n} := (n_1, \dots, n_m)$ denotes the outer unit normal to $\partial\Omega$ and γ is due to Theorem 3.8. Obviously, $H_0(\mathbf{grad}; \Omega) = H_0^1(\Omega)$ holds and $H_0(\mathbf{grad}; \Omega)/\text{Ker}(\mathbf{grad}; \Omega)$ is equal to $H_0^1(\Omega)$, since $\text{Ker}(\mathbf{grad}; \Omega) = \{0\}$, where due to (3.2)

$$\text{Ker}(\mathbf{grad}; \Omega) := \{u \in H_0(\mathbf{grad}; \Omega) \mid \mathbf{grad}(u) = \mathbf{0}\}. \quad (3.28)$$

Theorem 3.10. (Green's theorem in $H(\mathbf{grad})$) Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$, and let $u \in H(\mathbf{grad}; \Omega)$, $\mathbf{v} := (v_1, \dots, v_m) \in [H^1(\Omega)]^m$. Then the relation

$$\int_{\Omega} \mathbf{grad}(u) \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} u \, \text{div}(\mathbf{v}) \, d\mathbf{x} = \int_{\Omega} \gamma(u) \cdot \mathbf{v} \, d\mathbf{x}$$

holds, where the differential operator div is extended onto $[H^1(\Omega)]^m$ as follows:

$$\text{div}(\mathbf{v}) := \sum_{i=1}^m \frac{\partial v_i}{\partial x_i}, \quad \mathbf{v} := (v_1, \dots, v_m) \in [H^1(\Omega)]^m.$$

Proof. We use Theorem 3.8 and similar arguments as in the proof of Corollary 3.1. \square

Finally, in Theorem 3.9 we replace the symbols $\|u\|_{1, \Omega}$ and $|u|_{1, \Omega}$ by the symbols $\|u\|_{\mathbf{grad}, \Omega}$ and $|u|_{\mathbf{grad}, \Omega}$, respectively, and the theorem holds with the same constant C_3 .

3.3.4 The space $\mathbf{H}(\mathbf{curl})$

Let $\Omega \subset \mathbb{R}^3$ and $\Omega \in \mathcal{L}$. Now, like in the cases of D^α and \mathbf{grad} , we extend the differential operator $\mathbf{curl} : [C^1(\overline{\Omega})]^3 \mapsto [C(\overline{\Omega})]^3$ onto a subspace of $[L^2(\Omega)]^3$. The function $\mathbf{z} \in [L^2(\Omega)]^3$ is said to be the *generalized rotation* of $\mathbf{u} \in [L^2(\Omega)]^3$ if the following is satisfied

$$\forall \mathbf{v} \in [C_0^\infty(\Omega)]^3 : \int_{\Omega} \mathbf{u} \cdot \mathbf{curl}(\mathbf{v}) \, d\mathbf{x} = \int_{\Omega} \mathbf{z} \cdot \mathbf{v} \, d\mathbf{x},$$

and we denote the generalized rotation by $\mathbf{curl}(\mathbf{u}) := \mathbf{z}$. We define the space

$$\mathbf{H}(\mathbf{curl}; \Omega) := \left\{ \mathbf{u} \in [L^2(\Omega)]^3 \mid \exists \mathbf{z} \in [L^2(\Omega)]^3 : \mathbf{z} = \mathbf{curl}(\mathbf{u}) \right\},$$

which, together with the scalar product

$$(\mathbf{u}, \mathbf{v})_{\mathbf{curl}, \Omega} := \int_{\Omega} \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \mathbf{curl}(\mathbf{u}) \cdot \mathbf{curl}(\mathbf{v}) \, d\mathbf{x}, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{curl}; \Omega),$$

forms a Hilbert space. We introduce the induced norm and seminorm

$$\|\mathbf{u}\|_{\mathbf{curl}, \Omega} := \sqrt{(\mathbf{u}, \mathbf{u})_{\mathbf{curl}, \Omega}}, \quad |\mathbf{u}|_{\mathbf{curl}, \Omega} := \sqrt{\int_{\Omega} \|\mathbf{curl}(\mathbf{u})\|^2 \, d\mathbf{x}}, \quad \mathbf{u} \in \mathbf{H}(\mathbf{curl}; \Omega),$$

respectively, and the following is satisfied

$$\|\mathbf{u}\|_{\mathbf{curl}, \Omega}^2 = \|\mathbf{u}\|_{3,0,\Omega}^2 + |\mathbf{u}|_{\mathbf{curl}, \Omega}^2, \quad \mathbf{u} \in \mathbf{H}(\mathbf{curl}; \Omega).$$

The following two theorems are due to GIRAULT AND RAVIART [67, p. 34].

Theorem 3.11. (*Trace theorem in $\mathbf{H}(\mathbf{curl})$*) Let $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$. Then there exists exactly one linear continuous operator $\gamma : \mathbf{H}(\mathbf{curl}; \Omega) \mapsto [H^{-1/2}(\partial\Omega)]^3$ such that

$$\forall \mathbf{u} \in [C^\infty(\overline{\Omega})]^3 : \gamma(\mathbf{u}) = \mathbf{n} \times \mathbf{u}|_{\partial\Omega},$$

where \mathbf{n} is the outer unit normal to $\partial\Omega$.

Proof. See GIRAULT AND RAVIART [67, p. 34]. □

Theorem 3.11 enables us to define the space

$$\mathbf{H}_0(\mathbf{curl}; \Omega) := \{ \mathbf{u} \in \mathbf{H}(\mathbf{curl}; \Omega) \mid \gamma(\mathbf{u}) = \mathbf{0} \}.$$

Then, due to (3.2),

$$\mathbf{Ker}(\mathbf{curl}; \Omega) := \{ \mathbf{u} \in \mathbf{H}_0(\mathbf{curl}; \Omega) \mid \mathbf{curl}(\mathbf{u}) = \mathbf{0} \}.$$

By GIRAULT AND RAVIART [67, Corollary 2.9], the space $\mathbf{Ker}(\mathbf{curl}; \Omega)$ is equal to the space

$$\mathbf{H}_{0,0}(\mathbf{curl}; \Omega) := \left\{ \mathbf{u} \in [L^2(\Omega)]^3 \mid \exists p \in H_0^1(\Omega) : \mathbf{u} = \mathbf{grad}(p) \right\} \quad (3.29)$$

and, by HIPTMAIR [91, p. 94–95], the quotient space $\mathbf{H}_0(\mathbf{curl}; \Omega)/\mathbf{Ker}(\mathbf{curl}; \Omega)$ is isomorphically isometric to

$$\mathbf{H}_{0,\perp}(\mathbf{curl}; \Omega) := \left\{ \mathbf{u} \in \mathbf{H}_0(\mathbf{curl}; \Omega) \mid \forall v \in H_0^1(\Omega) : \int_{\Omega} \mathbf{u} \cdot \mathbf{grad}(v) \, d\mathbf{x} = 0 \right\}, \quad (3.30)$$

and, moreover, the following orthogonal decomposition holds

$$\mathbf{H}_0(\mathbf{curl}; \Omega) = \mathbf{H}_{0,\perp}(\mathbf{curl}; \Omega) \oplus \mathbf{H}_{0,0}(\mathbf{curl}; \Omega).$$

The spaces $[C^\infty(\overline{\Omega})]^3$ and $[C_0^\infty(\Omega)]^3$ are dense in $\mathbf{H}(\mathbf{curl}; \Omega)$ and $\mathbf{H}_0(\mathbf{curl}; \Omega)$, respectively, i.e.,

$$\mathbf{H}(\mathbf{curl}; \Omega) = \overline{[C^\infty(\overline{\Omega})]^3} \text{ and } \mathbf{H}_0(\mathbf{curl}; \Omega) = \overline{[C_0^\infty(\Omega)]^3} \text{ in the norm } \|\cdot\|_{\mathbf{curl},\Omega}. \quad (3.31)$$

Theorem 3.12. (*Green's theorem in $\mathbf{H}(\mathbf{curl})$*) Let $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$, and let $\mathbf{u} \in \mathbf{H}(\mathbf{curl}; \Omega)$, $\mathbf{v} \in [H^1(\Omega)]^3$. Then the relation

$$\int_{\Omega} \mathbf{curl}(\mathbf{u}) \cdot \mathbf{v} \, d\mathbf{x} - \int_{\Omega} \mathbf{u} \cdot \mathbf{curl}(\mathbf{v}) \, d\mathbf{x} = \langle \gamma(\mathbf{u}), \mathbf{v} \rangle_{\partial\Omega}$$

holds, where $\langle \gamma(\mathbf{u}), \mathbf{v} \rangle_{\partial\Omega}$ denotes the duality pairing between $[H^{-1/2}(\partial\Omega)]^3$ and $[H^{1/2}(\partial\Omega)]^3$.

Proof. See GIRAULT AND RAVIART [67, p. 34]. \square

The last theorem is a Friedrichs'-like inequality and it will be useful for analyzing the ellipticity of differential operators defined on $\mathbf{H}(\mathbf{curl}; \Omega)$.

Theorem 3.13. (*Friedrichs' inequality in $\mathbf{H}(\mathbf{curl})$*) Let $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$. Then there exists a positive constant $C_4 \equiv C_4(\Omega)$ such that

$$\forall \mathbf{u} \in \mathbf{H}_{0,\perp}(\mathbf{curl}; \Omega) : \|\mathbf{u}\|_{\mathbf{curl},\Omega} \leq C_4 |\mathbf{u}|_{\mathbf{curl},\Omega}.$$

Proof. See HIPTMAIR [91, p. 96]. \square

3.3.5 The space $\mathbf{H}(\text{div})$

Let $\Omega \subset \mathbb{R}^3$ and $\Omega \in \mathcal{L}$. We extend the differential operator $\text{div} : [C^1(\overline{\Omega})]^3 \mapsto C(\overline{\Omega})$ onto a subspace of $[L^2(\Omega)]^3$. The function $z \in L^2(\Omega)$ is said to be the *generalized divergence* of $\mathbf{u} \in [L^2(\Omega)]^3$ if the following is satisfied

$$\forall v \in C_0^\infty(\Omega) : \int_{\Omega} \mathbf{u} \cdot \mathbf{grad}(v) \, d\mathbf{x} = - \int_{\Omega} zv \, d\mathbf{x},$$

and we denote the generalized divergence by $\text{div}(\mathbf{u}) := z$. We define the space

$$\mathbf{H}(\text{div}; \Omega) := \left\{ \mathbf{u} \in [L^2(\Omega)]^3 \mid \exists z \in L^2(\Omega) : z = \text{div}(\mathbf{u}) \right\},$$

which, together with the scalar product

$$(\mathbf{u}, \mathbf{v})_{\text{div},\Omega} := \int_{\Omega} \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \text{div}(\mathbf{u}) \text{div}(\mathbf{v}) \, d\mathbf{x}, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\text{div}; \Omega),$$

forms a Hilbert space. We introduce the induced norm and seminorm

$$\|\mathbf{u}\|_{\text{div},\Omega} := \sqrt{(\mathbf{u}, \mathbf{u})_{\text{div},\Omega}}, \quad |\mathbf{u}|_{\text{div},\Omega} := \sqrt{\int_{\Omega} (\text{div}(\mathbf{u}))^2 \, d\mathbf{x}}, \quad \mathbf{u} \in \mathbf{H}(\text{div}; \Omega),$$

respectively, and the following is satisfied

$$\|\mathbf{u}\|_{\text{div},\Omega}^2 = \|\mathbf{u}\|_{3,0,\Omega}^2 + |\mathbf{u}|_{\text{div},\Omega}^2, \quad \mathbf{u} \in \mathbf{H}(\text{div}; \Omega).$$

The following two theorems are due to GIRAULT AND RAVIART [67, p. 27–28].

Theorem 3.14. (Trace theorem in $\mathbf{H}(\text{div})$) Let $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$. Then there exists exactly one linear continuous operator $\gamma : \mathbf{H}(\text{div}; \Omega) \mapsto H^{-1/2}(\partial\Omega)$ such that

$$\forall \mathbf{u} \in [C^\infty(\overline{\Omega})]^3 : \gamma(\mathbf{u}) = \mathbf{n} \cdot \mathbf{u}|_{\partial\Omega},$$

where \mathbf{n} is the outer unit normal to $\partial\Omega$.

Proof. See GIRAULT AND RAVIART [67, p. 27]. □

Theorem 3.14 enables us to define the spaces

$$\mathbf{H}_0(\text{div}; \Omega) := \{\mathbf{u} \in \mathbf{H}(\text{div}; \Omega) \mid \gamma(\mathbf{u}) = 0\},$$

$$\mathbf{Ker}(\text{div}; \Omega) := \{\mathbf{u} \in \mathbf{H}_0(\text{div}; \Omega) \mid \text{div}(\mathbf{u}) = 0\}.$$

The spaces $[C^\infty(\overline{\Omega})]^3$ and $[C_0^\infty(\Omega)]^3$, respectively, are dense in $\mathbf{H}(\text{div}; \Omega)$ and $\mathbf{H}_0(\text{div}; \Omega)$, i.e.,

$$\mathbf{H}(\text{div}; \Omega) = \overline{[C^\infty(\overline{\Omega})]^3} \text{ and } \mathbf{H}_0(\text{div}; \Omega) = \overline{[C_0^\infty(\Omega)]^3} \text{ in the norm } \|\cdot\|_{\text{div}, \Omega}.$$

Theorem 3.15. (Green's theorem in $\mathbf{H}(\text{div})$) Let $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$, and let $\mathbf{u} \in \mathbf{H}(\text{div}; \Omega)$, $v \in H^1(\Omega)$. Then the relation

$$\int_{\Omega} \text{div}(\mathbf{u}) v \, d\mathbf{x} + \int_{\Omega} \mathbf{u} \cdot \mathbf{grad}(v) \, d\mathbf{x} = \langle \gamma(\mathbf{u}), v \rangle_{\partial\Omega}$$

holds, where $\langle \gamma(\mathbf{u}), v \rangle_{\partial\Omega}$ denotes the duality pairing between $\gamma(\mathbf{u}) \in H^{-1/2}(\partial\Omega)$ and $v \in H^{1/2}(\partial\Omega)$.

Proof. See GIRAULT AND RAVIART [67, p. 28]. □

3.3.6 The abstract space $\mathbf{H}(\mathbf{B})$

In the previous subsections we could observe a similar structure, which we will formally summarize now.

Suppose $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. Let $\mathbf{B} : [C^1(\overline{\Omega})]^{\nu_1} \mapsto [C(\overline{\Omega})]^{\nu_2}$ be a linear differential operator of the first order defined by (3.21), where $\nu_1, \nu_2 \in \mathbb{N}$. Let the adjoint operator $\mathbf{B}^* : [C^1(\overline{\Omega})]^{\nu_2} \mapsto [C(\overline{\Omega})]^{\nu_1}$ be defined by (3.22). Then, the Green's formula (3.24) holds. We extend the differential operator \mathbf{B} onto a subspace of $[L^2(\Omega)]^{\nu_1}$. The function $\mathbf{z} \in [L^2(\Omega)]^{\nu_2}$ is said to be the *generalized first-order linear differential operator* \mathbf{B} of $\mathbf{u} \in [L^2(\Omega)]^{\nu_1}$ if the following is satisfied

$$\forall \mathbf{v} \in [C_0^\infty(\Omega)]^{\nu_2} : \int_{\Omega} \mathbf{u} \cdot \mathbf{B}^*(\mathbf{v}) \, d\mathbf{x} = - \int_{\Omega} \mathbf{z} \cdot \mathbf{v} \, d\mathbf{x}, \quad (3.32)$$

and we denote the generalized operator by $\mathbf{B}(\mathbf{u}) := \mathbf{z}$. We define the space

$$\mathbf{H}(\mathbf{B}; \Omega) := \{\mathbf{u} \in [L^2(\Omega)]^{\nu_1} \mid \exists \mathbf{z} \in [L^2(\Omega)]^{\nu_2} : \mathbf{z} = \mathbf{B}(\mathbf{u})\},$$

which is obviously a linear space. Further, we define the bilinear form

$$(\mathbf{u}, \mathbf{v})_{\mathbf{B}, \Omega} := \int_{\Omega} \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \mathbf{B}(\mathbf{u}) \cdot \mathbf{B}(\mathbf{v}) \, d\mathbf{x}, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega), \quad (3.33)$$

which can be shown to be a scalar product on $\mathbf{H}(\mathbf{B}; \Omega)$. The space $\mathbf{H}(\mathbf{B}; \Omega)$, together with this scalar product, forms a Hilbert space. The induced norm and seminorm are as follows:

$$\|\mathbf{u}\|_{\mathbf{B}, \Omega} := \sqrt{(\mathbf{u}, \mathbf{u})_{\mathbf{B}, \Omega}}, \quad |\mathbf{u}|_{\mathbf{B}, \Omega} := \sqrt{\int_{\Omega} \|\mathbf{B}(\mathbf{u})\|^2 dx}, \quad \mathbf{u} \in \mathbf{H}(\mathbf{B}; \Omega),$$

and the following is satisfied

$$\|\mathbf{u}\|_{\mathbf{B}, \Omega}^2 = \|\mathbf{u}\|_{\nu_1, 0, \Omega}^2 + |\mathbf{u}|_{\mathbf{B}, \Omega}^2, \quad \mathbf{u} \in \mathbf{H}(\mathbf{B}; \Omega). \quad (3.34)$$

We assume that the following trace property holds.

Assumption 3.2. Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. We assume that the trace operator $\gamma : [C^\infty(\overline{\Omega})]^{\nu_1} \mapsto [C(\partial\Omega)]^{\nu_2}$ defined by (3.23) can be uniquely extended by continuity to the operator, still denoted by, $\gamma : \mathbf{H}(\mathbf{B}; \Omega) \mapsto [H^{-1/2}(\partial\Omega)]^{\nu_2}$ such that on $[C^\infty(\overline{\Omega})]^{\nu_1}$ the relation (3.23) holds.

Now, we define the spaces

$$\mathbf{H}_0(\mathbf{B}; \Omega) := \{\mathbf{u} \in \mathbf{H}(\mathbf{B}; \Omega) \mid \gamma(\mathbf{u}) = \mathbf{0}\},$$

$$\mathbf{Ker}(\mathbf{B}; \Omega) := \{\mathbf{u} \in \mathbf{H}_0(\mathbf{B}; \Omega) \mid \mathbf{B}(\mathbf{u}) = \mathbf{0}\}. \quad (3.35)$$

Assumption 3.3. Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. We assume that $[C^\infty(\overline{\Omega})]^{\nu_1}$ and $[C_0^\infty(\Omega)]^{\nu_1}$, respectively, are dense in $\mathbf{H}(\mathbf{B}; \Omega)$ and $\mathbf{H}_0(\mathbf{B}; \Omega)$, i.e.,

$$\mathbf{H}(\mathbf{B}; \Omega) = \overline{[C^\infty(\overline{\Omega})]^{\nu_1}} \text{ and } \mathbf{H}_0(\mathbf{B}; \Omega) = \overline{[C_0^\infty(\Omega)]^{\nu_1}} \text{ in the norm } \|\cdot\|_{\mathbf{B}, \Omega}.$$

The following lemma gives a space which will be useful for the finite element approximation.

Lemma 3.5. *The space*

$$\mathbf{H}_{0, \perp}(\mathbf{B}; \Omega) := \{\mathbf{u} \in \mathbf{H}_0(\mathbf{B}; \Omega) \mid \forall \varphi \in \mathbf{Ker}(\mathbf{B}; \Omega) : (\mathbf{u}, \varphi)_{\nu_1, 0, \Omega} = 0\} \quad (3.36)$$

is isomorphically isometric to $\mathbf{H}_0(\mathbf{B}; \Omega)/\mathbf{Ker}(\mathbf{B}; \Omega)$ and the orthogonal decomposition

$$\mathbf{H}_0(\mathbf{B}; \Omega) = \mathbf{H}_{0, \perp}(\mathbf{B}; \Omega) \oplus \mathbf{Ker}(\mathbf{B}; \Omega) \quad (3.37)$$

holds.

Proof. Here, we use exactly the same technique as presented in HIPTMAIR [91, p. 94–95].

Let us recall the norm in the quotient space $\mathbf{H}_0(\mathbf{B}; \Omega)/\mathbf{Ker}(\mathbf{B}; \Omega)$

$$\|[\mathbf{v}]\|_{\mathbf{H}_0(\mathbf{B}; \Omega)/\mathbf{Ker}(\mathbf{B}; \Omega)} := \min_{\mathbf{w} \in \mathbf{Ker}(\mathbf{B}; \Omega)} \|\mathbf{v} + \mathbf{w}\|_{\mathbf{B}, \Omega}, \quad \mathbf{v} \in \mathbf{H}_0(\mathbf{B}; \Omega).$$

We look for a subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$ that consists of the minimizers $\mathbf{v} + \mathbf{w}(\mathbf{v})$ determined as follows:

$$\|\mathbf{v} + \mathbf{w}(\mathbf{v})\|_{\mathbf{B}, \Omega}^2 = \|\mathbf{v}\|_{\mathbf{B}, \Omega}^2 + \min_{\mathbf{w}(\mathbf{v}) \in \mathbf{Ker}(\mathbf{B}; \Omega)} \left\{ 2(\mathbf{v}, \mathbf{w}(\mathbf{v}))_{\nu_1, 0, \Omega} + \|\mathbf{w}(\mathbf{v})\|_{\nu_1, 0, \Omega}^2 \right\}.$$

Due to Lemma 3.2, we arrive at the variational problem

$$\left. \begin{aligned} \text{Find } \mathbf{w}(\mathbf{v}) \in \mathbf{Ker}(\mathbf{B}; \Omega) : \\ (\varphi, \mathbf{w}(\mathbf{v}))_{\nu_1, 0, \Omega} = -(\mathbf{v}, \varphi)_{\nu_1, 0, \Omega} \quad \forall \varphi \in \mathbf{Ker}(\mathbf{B}; \Omega) \end{aligned} \right\}$$

and, since $(\cdot, \cdot)_{\nu_1, 0, \Omega}$ is a scalar product on $\mathbf{Ker}(\mathbf{B}; \Omega)$, by Theorem 3.1 $\mathbf{w}(\mathbf{v})$ is unique. Therefore, the minimizer $\mathbf{u} := \mathbf{v} + \mathbf{w}(\mathbf{v}) \in \mathbf{H}_0(\mathbf{B}; \Omega)$ is uniquely characterized by

$$\forall \varphi \in \mathbf{Ker}(\mathbf{B}; \Omega) : (\mathbf{u}, \varphi)_{\nu_1, 0, \Omega} = (\mathbf{u}, \varphi)_{\mathbf{B}, \Omega} = 0. \quad (3.38)$$

The space $\mathbf{H}_{0, \perp}(\mathbf{B}; \Omega)$, see (3.36), which consists of such minimizers is a closed subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$ and, due to (3.38) and (3.8),

$$\mathbf{H}_{0, \perp}(\mathbf{B}; \Omega) = \mathbf{Ker}(\mathbf{B}; \Omega)^\perp,$$

which completes the proof. \square

Further, we assume that the following Green's formula holds.

Assumption 3.4. Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$, and let $\mathbf{u} \in \mathbf{H}(\mathbf{B}; \Omega)$, $\mathbf{v} \in [H^1(\Omega)]^{\nu_2}$. We assume that the relation

$$\int_{\Omega} \mathbf{B}(\mathbf{u}) \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} \mathbf{u} \cdot \mathbf{B}^*(\mathbf{v}) \, d\mathbf{x} = \langle \gamma(\mathbf{u}), \mathbf{v} \rangle_{\partial\Omega}$$

holds, in which the duality pairing between $[H^{-1/2}(\partial\Omega)]^{\nu_2}$ and $[H^{1/2}(\partial\Omega)]^{\nu_2}$ is denoted by $\langle \gamma(\mathbf{u}), \mathbf{v} \rangle_{\partial\Omega}$.

Finally, we will need the ellipticity. To this end, we assume that the following Friedrichs'–like inequality holds.

Assumption 3.5. Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. We assume that there exists a positive constant $C_5 \equiv C_5(\Omega)$ such that

$$\forall \mathbf{u} \in \mathbf{H}_{0, \perp}(\mathbf{B}; \Omega) : \|\mathbf{u}\|_{\mathbf{B}, \Omega} \leq C_5 |\mathbf{u}|_{\mathbf{B}, \Omega}.$$

At the end, we summarize how the abstract operators \mathbf{B} , \mathbf{B}^* , and γ read in the spaces introduced above.

m	ν_1	ν_2	$\mathbf{H}(\mathbf{B}; \Omega)$	\mathbf{B}	\mathbf{B}^*	$\gamma(\mathbf{v})$
$\in \mathbb{N}$	1	m	$H(\mathbf{grad}; \Omega)$	\mathbf{grad}	\mathbf{div}	$\mathbf{v}\mathbf{n}$
$\in \mathbb{N}$	m	1	$\mathbf{H}(\mathbf{div}; \Omega)$	\mathbf{div}	\mathbf{grad}	$\mathbf{n} \cdot \mathbf{v}$
3	3	3	$\mathbf{H}(\mathbf{curl}; \Omega)$	\mathbf{curl}	$-\mathbf{curl}$	$\mathbf{n} \times \mathbf{v}$

Table 3.1: Operators in Hilbert function spaces

3.4 Weak formulations of boundary vector–value problems

Let us refer to some more literature, where the authors deal with weak settings of boundary value problems, see RITZ [172], AUBIN [12], WASHIZU [212], GROSSMANN AND ROSS [71], JOHNSON [101], SHOWALTER [189], or HLAVÁČEK, HASLINGER, NEČAS, AND LOVÍŠEK [95].

Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. Like in KŘÍŽEK AND NEITTAANMÄKI [115, p. 27], we consider a boundary value problem, the *strong formulation* of which reads as follows: Find $\mathbf{u} \in [C^2(\overline{\Omega})]^{\nu_1}$, $\nu_1 \in \mathbb{N}$, such that

$$\left. \begin{aligned} -\mathbf{B}^*(\mathbf{D} \cdot \mathbf{B}(\mathbf{u})) &= \mathbf{f} \text{ in } \Omega \\ \gamma(\mathbf{u}) &= \mathbf{0} \text{ on } \partial\Omega \end{aligned} \right\}, \quad (S)$$

where \mathbf{B} , \mathbf{B}^* , and γ are defined by Definition 3.2, and where $\mathbf{D} \in [C^1(\overline{\Omega})]^{\nu_2 \times \nu_2}$, $\nu_2 \in \mathbb{N}$, is a uniformly positive definite real-valued matrix, i.e., there exists a constant $C_6 > 0$ such that

$$\forall x \in \overline{\Omega} \forall \mathbf{v} \in \mathbb{R}^{\nu_2} : \mathbf{v} \cdot (\mathbf{D}(x) \cdot \mathbf{v}) \geq C_6 \|\mathbf{v}\|^2, \quad (3.39)$$

and where $\mathbf{f} \in [C(\overline{\Omega})]^{\nu_1}$. The function $\mathbf{u} \in [C^2(\overline{\Omega})]^{\nu_1}$ is called the *classical solution* to (S).

Now, we introduce a weak setting of (S), which will enable us to weaken the assumptions on the differentiability of the data in (S) and to deal with problems of more practical purposes. Let us take into account the extensions of definitions of \mathbf{B} , \mathbf{B}^* , and γ from Section 3.3.6, as well as Assumptions 3.2–3.5 and Lemma 3.5 introduced there. We define the continuous bilinear form $a : \mathbf{H}(\mathbf{B}; \Omega) \times \mathbf{H}(\mathbf{B}; \Omega) \mapsto \mathbb{R}$ and the continuous linear functional $f : \mathbf{H}(\mathbf{B}; \Omega) \mapsto \mathbb{R}$ by

$$a(\mathbf{v}, \mathbf{u}) := \int_{\Omega} \mathbf{B}(\mathbf{v}) \cdot (\mathbf{D} \cdot \mathbf{B}(\mathbf{u})) \, d\mathbf{x}, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega), \quad (3.40)$$

$$f(\mathbf{v}) := \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x}, \quad \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega), \quad (3.41)$$

respectively, where $\mathbf{f} \in [L^2(\Omega)]^{\nu_1}$, $\mathbf{D} := (d_{i,j})$ is a matrix the entries of which $d_{i,j} \in L^\infty(\Omega)$, $i, j = 1, \dots, \nu_2$, and the condition (3.39) holds almost everywhere (a.e.) in Ω . A *weak formulation* of the problem (S) reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{u} \in \mathbf{H}_0(\mathbf{B}; \Omega): \\ a(\mathbf{v}, \mathbf{u}) = f(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_0(\mathbf{B}; \Omega) \end{array} \right\}. \quad (3.42)$$

Just by applying Corollary 3.3, we can see that the classical solution $\mathbf{u} \in [C^2(\overline{\Omega})]^{\nu_1}$ of the problem (S) is also a solution to (3.42). However, the problem (3.42) admits more general and physically still reasonable data.

We can observe that if $\mathbf{u} \in \mathbf{H}_0(\mathbf{B}; \Omega)$ is a solution to (3.42), then for any $\mathbf{p} \in \mathbf{Ker}(\mathbf{B}; \Omega)$ the function $\mathbf{u} + \mathbf{p}$ is a solution, too. This indicates a Neumann-like problem. Therefore, we restrict our consideration onto the quotient space $\mathbf{H}_0(\mathbf{B}; \Omega)/\mathbf{Ker}(\mathbf{B}; \Omega)$, which is by Lemma 3.5 isomorphically isometric to $\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$, being a subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$. In this case, we have to introduce a compatibility condition on the right-hand side \mathbf{f}

$$\forall \mathbf{p} \in \mathbf{Ker}(\mathbf{B}; \Omega) : \int_{\Omega} \mathbf{f} \cdot \mathbf{p} \, d\mathbf{x} = 0. \quad (3.43)$$

The correct weak formulation of (S) reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{u} \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega): \\ a(\mathbf{v}, \mathbf{u}) = f(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega) \end{array} \right\}, \quad (W)$$

where \mathbf{f} satisfies (3.43). It can be easily verified that a solution \mathbf{u} to (W) also solves the problem (3.42). On the other hand, the next theorem shows that (W) has a unique solution, unlike the problem (3.42).

Theorem 3.16. *There exists exactly one solution $\mathbf{u} \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ to the problem (W). Moreover, there exists a positive constant C_7 such that*

$$\|\mathbf{u}\|_{\mathbf{B}, \Omega} \leq C_7 \|\mathbf{f}\|_{\nu_1, 0, \Omega}. \quad (3.44)$$

Proof. We will check the assumptions of Lemma 3.1 and the assertion then follows.

Since $\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ is a closed subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$, it is also a Hilbert space equipped with the scalar product (3.33). The form a is clearly bilinear. Concerning the matrix \mathbf{D} , we denote

$$d := \max_{i,j} \left\{ \operatorname{ess\,sup}_{x \in \Omega} |d_{i,j}(x)| \right\} \quad (3.45)$$

and, using (3.3) and (3.34), we prove the continuity of a as follows:

$$|a(\mathbf{v}, \mathbf{u})| \leq d \left| \int_{\Omega} \mathbf{B}(\mathbf{v}) \cdot \mathbf{B}(\mathbf{u}) \, dx \right| \leq d \|\mathbf{B}(\mathbf{v})\|_{\nu_2,0,\Omega} \|\mathbf{B}(\mathbf{u})\|_{\nu_2,0,\Omega} \leq d \|\mathbf{v}\|_{\mathbf{B},\Omega} \|\mathbf{u}\|_{\mathbf{B},\Omega},$$

where $\mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega)$, and where we used that $\mathbf{D} \in [L^\infty(\Omega)]^{\nu_2 \times \nu_2}$, (3.3), and (3.34), respectively. The $\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ -ellipticity of a follows from

$$a(\mathbf{v}, \mathbf{v}) = \int_{\Omega} \mathbf{B}(\mathbf{v}) \cdot (\mathbf{D} \cdot \mathbf{B}(\mathbf{v})) \, dx \geq C_6 \int_{\Omega} \|\mathbf{B}(\mathbf{v})\|^2 \, dx \geq \frac{C_6}{C_5^2} \|\mathbf{v}\|_{\mathbf{B},\Omega}^2, \quad \mathbf{v} \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega), \quad (3.46)$$

where (3.39) and Assumption 3.5 were used, respectively. Finally, f is obviously a linear functional on $\mathbf{H}(\mathbf{B}; \Omega)$ and it is continuous thereon, too, since

$$|f(\mathbf{v})| = \left| \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx \right| \leq \|\mathbf{f}\|_{\nu_1,0,\Omega} \|\mathbf{v}\|_{\nu_1,0,\Omega} \leq \|\mathbf{f}\|_{\nu_1,0,\Omega} \|\mathbf{v}\|_{\mathbf{B},\Omega}, \quad \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega),$$

where we used (3.3) and (3.34), respectively. The assertion now follows from Lemma 3.1, where the $\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ -ellipticity constant is

$$C_7 := \frac{C_6}{C_5^2}.$$

□

3.4.1 A regularized formulation in $\mathbf{H}_0(\mathbf{B})$

In many cases, as in both 2d and 3d magnetostatics, we look for $\mathbf{B}(\mathbf{u})$ rather than for \mathbf{u} , a solution to (W) . Since there is the additional condition (3.38) in the definition of the space $\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$, it would be difficult to approximate the space by the finite element method. Therefore, we will introduce a regularized weak formulation in the original space $\mathbf{H}_0(\mathbf{B}; \Omega)$ such that its solution tends towards the solution $\mathbf{u} \in \mathbf{H}_0(\mathbf{B}; \Omega)$ of the problem (W) , but in the seminorm $|\cdot|_{\mathbf{B},\Omega}$ only.

Let $\varepsilon > 0$ be a regularization parameter. We introduce the following bilinear form

$$a_\varepsilon(\mathbf{v}, \mathbf{u}) := a(\mathbf{v}, \mathbf{u}) + \varepsilon \int_{\Omega} \mathbf{v} \cdot \mathbf{u} \, dx, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega), \quad (3.47)$$

where a is given by (3.40). The regularized weak formulation reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{u}_\varepsilon \in \mathbf{H}_0(\mathbf{B}; \Omega): \\ a_\varepsilon(\mathbf{v}, \mathbf{u}_\varepsilon) = f(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_0(\mathbf{B}; \Omega) \end{array} \right\}, \quad (W_\varepsilon)$$

where f is given by (3.41) such that (3.43) holds.

Theorem 3.17. *For each $\varepsilon > 0$ there exists a unique solution $\mathbf{u}_\varepsilon \in \mathbf{H}_0(\mathbf{B}; \Omega)$ to the problem (W_ε) . Moreover, there exists a positive constant $C_8(\varepsilon)$ such that*

$$\|\mathbf{u}_\varepsilon\|_{\mathbf{B},\Omega} \leq C_8(\varepsilon) \|\mathbf{f}\|_{\nu_1,0,\Omega}.$$

Proof. The proof is fairly the same as the one of Theorem 3.16. The continuity of a is proven as follows:

$$|a_\varepsilon(\mathbf{v}, \mathbf{u})| \leq d \left| \int_{\Omega} \mathbf{B}(\mathbf{v}) \cdot \mathbf{B}(\mathbf{u}) \, d\mathbf{x} \right| + \varepsilon \left| \int_{\Omega} \mathbf{v} \cdot \mathbf{u} \, d\mathbf{x} \right| \leq \max\{d, \varepsilon\} \|\mathbf{v}\|_{\mathbf{B}, \Omega} \|\mathbf{u}\|_{\mathbf{B}, \Omega}, \quad (3.48)$$

where $\mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega)$ and where d is given by (3.45). The $\mathbf{H}_0(\mathbf{B}; \Omega)$ -ellipticity of a_ε follows from

$$a_\varepsilon(\mathbf{v}, \mathbf{v}) \geq C_6 \int_{\Omega} \|\mathbf{B}(\mathbf{v})\|^2 \, d\mathbf{x} + \varepsilon \int_{\Omega} |\mathbf{v}|^2 \, d\mathbf{x} \geq \min\{C_6, \varepsilon\} \|\mathbf{v}\|_{\mathbf{B}, \Omega}^2, \quad \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega),$$

where C_6 is given by (3.39). Therefore, the $\mathbf{H}_0(\mathbf{B}; \Omega)$ -ellipticity constant is

$$C_8(\varepsilon) := \min\{C_6, \varepsilon\}. \quad (3.49)$$

□

Theorem 3.18. *The following holds:*

$$\mathbf{B}(\mathbf{u}_\varepsilon) \rightarrow \mathbf{B}(\mathbf{u}) \text{ in } [L^2(\Omega)]^{\nu_2}, \text{ as } \varepsilon \rightarrow 0_+, \quad (3.50)$$

where $\mathbf{u}_\varepsilon \in \mathbf{H}_0(\mathbf{B}; \Omega)$ are the solutions to (W_ε) and $\mathbf{u} \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ is the solution to (W) .

Proof. Let $\varepsilon > 0$ be arbitrary. Using (3.39) and the definitions of (W) and (W_ε) , we have

$$\begin{aligned} \|\mathbf{B}(\mathbf{u}_\varepsilon) - \mathbf{B}(\mathbf{u})\|_{\nu_2, 0, \Omega}^2 &= |\mathbf{u}_\varepsilon - \mathbf{u}|_{\mathbf{B}, \Omega}^2 = \int_{\Omega} \|\mathbf{B}(\mathbf{u}_\varepsilon) - \mathbf{B}(\mathbf{u})\|^2 \, d\mathbf{x} \leq \\ &\leq \frac{1}{C_6} \int_{\Omega} \mathbf{B}(\mathbf{u}_\varepsilon - \mathbf{u}) \cdot (\mathbf{D} \cdot \mathbf{B}(\mathbf{u}_\varepsilon - \mathbf{u})) \, d\mathbf{x} = \frac{1}{C_6} a(\mathbf{u}_\varepsilon - \mathbf{u}, \mathbf{u}_\varepsilon - \mathbf{u}) \leq \\ &\leq \frac{1}{C_6} a_\varepsilon(\mathbf{u}_\varepsilon - \mathbf{u}, \mathbf{u}_\varepsilon - \mathbf{u}) = \frac{1}{C_6} (f(\mathbf{u}_\varepsilon - \mathbf{u}) - a_\varepsilon(\mathbf{u}_\varepsilon - \mathbf{u}, \mathbf{u})) = \\ &= \frac{1}{C_6} \left(f(\mathbf{u}_\varepsilon - \mathbf{u}) - a(\mathbf{u}_\varepsilon - \mathbf{u}, \mathbf{u}) - \varepsilon \int_{\Omega} (\mathbf{u}_\varepsilon - \mathbf{u}) \cdot \mathbf{u} \, d\mathbf{x} \right) \end{aligned} \quad (3.51)$$

Using the orthogonal decomposition (3.37), there exists $\mathbf{u}_{\varepsilon, \perp} \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ and there exists $\mathbf{u}_{\varepsilon, 0} \in \mathbf{Ker}(\mathbf{B}; \Omega)$ such that

$$\mathbf{u}_\varepsilon = \mathbf{u}_{\varepsilon, \perp} + \mathbf{u}_{\varepsilon, 0}.$$

Using the latter, the condition (3.43), (3.38) and (3.25), the estimate (3.51) reads

$$\begin{aligned} \|\mathbf{B}(\mathbf{u}_\varepsilon) - \mathbf{B}(\mathbf{u})\|_{\nu_2, 0, \Omega}^2 &= \|\mathbf{B}(\mathbf{u}_{\varepsilon, \perp}) - \mathbf{B}(\mathbf{u})\|_{\nu_2, 0, \Omega}^2 \leq \\ &\leq \frac{1}{C_6} \left(f(\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}) - a(\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}, \mathbf{u}) - \varepsilon \int_{\Omega} (\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}) \cdot \mathbf{u} \, d\mathbf{x} \right) \leq \\ &\leq \frac{\varepsilon}{C_6} \left| \int_{\Omega} (\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}) \cdot \mathbf{u} \, d\mathbf{x} \right| \leq \frac{\varepsilon}{C_6} \|\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}\|_{\mathbf{B}, \Omega} \|\mathbf{u}\|_{\nu_1, 0, \Omega}. \end{aligned}$$

Now we use Assumption 3.5

$$\|\mathbf{B}(\mathbf{u}_\varepsilon) - \mathbf{B}(\mathbf{u})\|_{\nu_2, 0, \Omega}^2 = |\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}|_{\mathbf{B}, \Omega}^2 \leq \varepsilon \frac{C_5}{C_6} |\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}|_{\mathbf{B}, \Omega} \|\mathbf{u}\|_{\nu_1, 0, \Omega}.$$

After dividing the latter by $|\mathbf{u}_{\varepsilon, \perp} - \mathbf{u}|_{\mathbf{B}, \Omega}$, the statement follows. □

3.4.2 A weak formulation of three-dimensional linear magnetostatics

Here we apply the results of the previous sections to the strong formulation (2.5).

Let $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$. Taking a look into Table 3.1, at (S), and (2.5), we specify the symbols

$$\mathbf{B} := \mathbf{curl}, \quad \mathbf{B}^* := -\mathbf{curl}, \quad \text{and} \quad \gamma(\mathbf{v}) := \mathbf{n} \times \mathbf{v}|_{\partial\Omega},$$

where $\mathbf{n} := (n_1, n_2, n_3)$ denotes the outer unit normal to $\partial\Omega$. Due to (2.5), we determine the symbols

$$\mathbf{D} := \frac{1}{\mu}, \quad \mathbf{f} := \mathbf{J},$$

where $\mu \in L^\infty(\Omega)$, $\mu > 0$ a.e. in Ω , $\mathbf{J} \in [L^2(\Omega)]^3$. The condition (3.39) is now equivalent to

$$\exists \mu_1 > 0 : \mu(x) \leq \mu_1 \text{ a.e. in } \Omega \quad (3.52)$$

in such a way that

$$C_6 := \frac{1}{\mu_1}.$$

Since $\mathbf{Ker}(\mathbf{curl}; \Omega)$ is equal to the space $\mathbf{H}_{0,0}(\mathbf{curl}; \Omega)$ defined by (3.29) the condition (3.43) reads as follows:

$$\forall p \in H_0^1(\Omega) : \int_{\Omega} \mathbf{J} \cdot \mathbf{grad}(p) \, dx = 0.$$

Finally, we specify the terms in (W). As we have seen in Section 3.3.4, the quotient space $\mathbf{H}_0(\mathbf{curl}; \Omega) / \mathbf{Ker}(\mathbf{curl}; \Omega)$ is isomorphically isometric to the space $\mathbf{H}_{0,\perp}(\mathbf{curl}; \Omega)$, which was defined by (3.30). The bilinear form (3.40) and the linear form (3.41) are, respectively, determined by

$$a(\mathbf{v}, \mathbf{u}) := \int_{\Omega} \mathbf{curl}(\mathbf{v}) \cdot \left(\frac{1}{\mu} \mathbf{curl}(\mathbf{u}) \right) \, dx, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{curl}; \Omega),$$

and by

$$f(\mathbf{v}) := \int_{\Omega} \mathbf{J} \cdot \mathbf{v} \, dx, \quad \mathbf{v} \in \mathbf{H}(\mathbf{curl}; \Omega).$$

We have specified all the assumptions on the abstract weak formulation (W) introduced in Section 3.4. Therefore, Theorem 3.16 holds with the $\mathbf{H}_{0,\perp}(\mathbf{curl}; \Omega)$ -ellipticity constant

$$C_7 := \frac{1}{\mu_0 C_4^2},$$

where C_4 is given by Theorem 3.13. In case of the formulation (W_ε) , we only recall the regularized bilinear form

$$a_\varepsilon(\mathbf{v}, \mathbf{u}) := \int_{\Omega} \mathbf{curl}(\mathbf{v}) \cdot \left(\frac{1}{\mu} \mathbf{curl}(\mathbf{u}) \right) \, dx + \varepsilon \int_{\Omega} \mathbf{v} \cdot \mathbf{u} \, dx, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{curl}; \Omega),$$

whose ellipticity constant is given by (3.49).

3.4.3 A weak formulation of two-dimensional linear magnetostatics

Now we apply the results of Section 3.4 to the strong formulation (2.6).

Let $\Omega \subset \mathbb{R}^2$, $\Omega \in \mathcal{L}$. We specify the symbols

$$\mathbf{B} := \mathbf{grad}, \quad \mathbf{B}^* := -\operatorname{div}, \quad \text{and} \quad \gamma(v) := v|_{\partial\Omega}\mathbf{n},$$

where $\mathbf{n} := (n_1, n_2)$ denotes the outer unit normal to $\partial\Omega$. Due to (2.6), we determine the symbols

$$\mathbf{D} := \frac{1}{\mu}, \quad \mathbf{f} := J, \tag{3.53}$$

where $\mu \in L^\infty(\Omega)$, $\mu > 0$ a.e. in Ω , $J \in L^2(\Omega)$. We again replace the condition (3.39) by (3.52). Since $\operatorname{Ker}(\mathbf{grad}; \Omega)$ defined by (3.28) is equal to the zero space $\{0\}$, the condition (3.43) always holds and does not need to be introduced in this case. Finally, we specify the terms in (W) . As we have seen in Section 3.3.3, the quotient space $H_0(\mathbf{grad}; \Omega)/\{0\}$ is equal to the space $H_0^1(\Omega)$. The bilinear form (3.40) and the linear functional (3.41) are, respectively, determined by

$$a(v, u) := \int_{\Omega} \mathbf{grad}(v) \cdot \left(\frac{1}{\mu} \mathbf{grad}(u) \right) d\mathbf{x}, \quad u, v \in H^1(\Omega),$$

and by

$$f(v) := \int_{\Omega} Jv d\mathbf{x}, \quad v \in H^1(\Omega).$$

Now, Theorem 3.16 holds with the $H_0^1(\Omega)$ -ellipticity constant

$$C_7 := \frac{1}{\mu_0 C_3^2},$$

where C_3 is given by Theorem 3.9. In this case, we do not need to introduce the regularized problem (W_ε) , since all the spaces are equal

$$H_{0,\perp}(\mathbf{grad}; \Omega) = H_0(\mathbf{grad}; \Omega) = H_0^1(\Omega).$$

Chapter 4

Finite element method

In this chapter, we will recall the basic techniques of the finite element method. First, we will present the ideas and concept of the method. Then, we will deal with algorithmic issues. Further, approximation properties will be proved. At the end, we will describe the finite elements used for the linear magnetostatics, namely Lagrange nodal elements on triangles and Nédélec edge elements on tetrahedra. We refer to KŘÍŽEK AND NEITTAANMÄKI [115, Chapter 4] for a detailed description of the method.

The large popularity of the method can be documented by the following unsorted list of literature: GIRAULT AND RAVIART [67], BRAESS [27], BOSSAVIT [26], STRANG AND FIX [200], CIARLET [45], ZLÁMAL [219], RAVIART AND THOMAS [167], HIPTMAIR [91, 93], ZIENKIEWICZ [217], ZIENKIEWICZ AND TAYLOR [218], HASLINGER, MIETTINEN, AND PANAGIOTOPULOS [84], BREZZI AND FORTIN [33], HUGHES [98], JUNG AND LANGER [103], BRENNER AND SCOTT [32], GROSSMANN AND ROSS [71], JOHNSON [101], KIKUCHI [108], ODEN AND REDDY [150], SCHWAB [187], SZABÓ AND BABUŠKA [205], BABUŠKA AND AZIZ [14, 15], GLOWINSKI [68], KŘÍŽEK AND NEITTAANMÄKI [114, 115], KŘÍŽEK [113], NEITTAANMÄKI AND SARANEN [147, 146], SILVESTER AND FERRARI [192], HACKBUSCH AND SAUTER [79, 80].

4.1 The concept of the method

We consider the regularized weak formulation (W_ε) of the abstract elliptic linear boundary vector-value problem (S), which was introduced in Section 3.4.1. The aim of this chapter is to develop a method which approximates the continuous regularized solution \mathbf{u}_ε of the problem (W_ε) by a sequence of some discretized solutions \mathbf{u}_ε^h , where $h > 0$ stands for a *discretization parameter*.

4.1.1 Galerkin approximation

Let $\mathbf{V}^h \subset \mathbf{H}_0(\mathbf{B}; \Omega)$ be a closed subspace. We introduce approximations of the bilinear form (3.47) and of the linear functional (3.41), respectively, by

$$a_\varepsilon^h(\mathbf{v}, \mathbf{w}) := \int_{\Omega} \mathbf{B}(\mathbf{v}) \cdot (\mathbf{D}^h \cdot \mathbf{B}(\mathbf{w})) \, dx + \varepsilon \int_{\Omega} \mathbf{v} \cdot \mathbf{w} \, dx, \quad \mathbf{v}, \mathbf{w} \in \mathbf{H}(\mathbf{B}; \Omega), \quad (4.1)$$

$$f^h(\mathbf{v}) := \int_{\Omega} \mathbf{f}^h \cdot \mathbf{v} \, dx, \quad \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega), \quad (4.2)$$

where $\mathbf{f}^h \in (L^2(\Omega))^{\nu_1}$ is such that

$$\forall h > 0 \forall \mathbf{p} \in \mathbf{Ker}(\mathbf{B}; \Omega) : \int_{\Omega} \mathbf{f}^h \cdot \mathbf{p} \, d\mathbf{x} = 0,$$

and where $\mathbf{D}^h := \left(d_{i,j}^h \right)_{i,j=1}^{\nu_2}$ is a matrix the entries of which $d_{i,j}^h \in L^\infty(\Omega)$ and such that

$$\forall h > 0 : \operatorname{ess\,sup}_{\mathbf{x} \in \Omega} \left| d_{i,j}^h(\mathbf{x}) \right| \leq \operatorname{ess\,sup}_{\mathbf{x} \in \Omega} |d_{i,j}(\mathbf{x})|, \quad i, j = 1, \dots, \nu_2. \quad (4.3)$$

Both \mathbf{f}^h and \mathbf{D}^h are assumed to be piecewise constant. Moreover, we suppose that

$$\forall h > 0 \forall \mathbf{v} \in \mathbb{R}^{\nu_2} : \mathbf{v} \cdot \left(\mathbf{D}^h(\mathbf{x}) \cdot \mathbf{v} \right) \geq C_6 \|\mathbf{v}\|^2 \quad \text{a.e. in } \Omega, \quad (4.4)$$

where $C_6 > 0$ is given by (3.39). We consider the following problem

$$\left. \begin{array}{l} \text{Find } \mathbf{u}_\varepsilon^h \in \mathbf{V}^h: \\ a_\varepsilon^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^h) = f^h(\mathbf{v}^h) \quad \forall \mathbf{v}^h \in \mathbf{V}^h \end{array} \right\}, \quad (W_\varepsilon^h)$$

which is called the *Galerkin approximation* to the problem (W_ε) .

Theorem 4.1. *For each $\varepsilon > 0$ and $h > 0$ there exists a unique solution $\mathbf{u}_\varepsilon^h \in \mathbf{V}^h$ to the problem (W_ε^h) . Moreover, there exists a positive constant $C_8(\varepsilon)$ such that*

$$\forall h > 0 : \|\mathbf{u}_\varepsilon^h\|_{\mathbf{B}, \Omega} \leq C_8(\varepsilon) \left\| \mathbf{f}^h \right\|_{\nu_1, 0, \Omega}.$$

Proof. The proof is fairly the same as the one of Theorem 3.17.

Let $\varepsilon > 0$ and $h > 0$ be arbitrary. By definition, \mathbf{V}^h is a closed subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$, therefore, it is also a Hilbert space. The form a_ε^h is obviously bilinear and f^h is a linear functional on \mathbf{V}^h . We have the continuity of a_ε^h

$$\left| a_\varepsilon^h(\mathbf{v}, \mathbf{w}) \right| \leq \max\{d^h, \varepsilon\} \|\mathbf{v}\|_{\mathbf{B}, \Omega} \|\mathbf{w}\|_{\mathbf{B}, \Omega}, \quad \mathbf{v}, \mathbf{w} \in \mathbf{H}(\mathbf{B}; \Omega),$$

where

$$d^h := \max_{i,j} \left\{ \operatorname{ess\,sup}_{\mathbf{x} \in \Omega} \left| d_{i,j}^h(\mathbf{x}) \right| \right\}.$$

Due to (4.4), we have the \mathbf{V}^h -ellipticity of a_ε^h

$$a_\varepsilon^h(\mathbf{v}, \mathbf{v}) \geq \min\{C_6, \varepsilon\} \|\mathbf{v}\|_{\mathbf{B}, \Omega}^2, \quad \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega),$$

independently of h . Finally, the continuity of f^h follows from

$$f^h(\mathbf{v}) \leq \left\| \mathbf{f}^h \right\|_{\nu_1, 0, \Omega} \|\mathbf{v}\|_{\mathbf{B}, \Omega}, \quad \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega).$$

Therefore, the assertion follows with the \mathbf{V}^h -ellipticity constant

$$C_8(\varepsilon) := \min\{C_6, \varepsilon\}. \quad (4.5)$$

□

The following lemma, cf. BRAESS [27], CIARLET [45], or STRANG AND FIX [200], says that we can study the approximation properties of \mathbf{u}_ε^h via the approximation of $\mathbf{H}_0(\mathbf{B}; \Omega)$ by its subspaces \mathbf{V}^h and via the approximation properties of the forms a_ε^h and f^h .

Lemma 4.1. (*1st Strang's lemma*) *Let $\varepsilon > 0$ and $h > 0$ be a regularization and discretization parameter, respectively. Then there exists a positive constant $C_9(\varepsilon)$, independent of h , such that*

$$\forall \mathbf{v}^h \in \mathbf{V}^h : \left\| \mathbf{u}_\varepsilon - \mathbf{u}_\varepsilon^h \right\|_{\mathbf{B}, \Omega} \leq C_9(\varepsilon) \left\{ \left\| \mathbf{u}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} + \frac{|a_\varepsilon(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h) - a_\varepsilon^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h)|}{\|\mathbf{u}_\varepsilon^h - \mathbf{v}^h\|_{\mathbf{B}, \Omega}} + \frac{|f(\mathbf{u}_\varepsilon^h - \mathbf{v}^h) - f^h(\mathbf{u}_\varepsilon^h - \mathbf{v}^h)|}{\|\mathbf{u}_\varepsilon^h - \mathbf{v}^h\|_{\mathbf{B}, \Omega}} \right\},$$

where $\mathbf{u}_\varepsilon \in \mathbf{H}_0(\mathbf{B}; \Omega)$ is a solution to (W_ε) and $\mathbf{u}_\varepsilon^h \in \mathbf{V}^h$ are solutions to (W_ε^h) .

Proof. Let $h > 0$ and $\mathbf{v}^h \in \mathbf{V}^h$ be arbitrary. By the triangle inequality,

$$\left\| \mathbf{u}_\varepsilon - \mathbf{u}_\varepsilon^h \right\|_{\mathbf{B}, \Omega} \leq \left\| \mathbf{u}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} + \left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega}. \quad (4.6)$$

To make the proof more readable, we introduce the symbol

$$\mathbf{w}_\varepsilon^h := \mathbf{u}_\varepsilon^h - \mathbf{v}^h. \quad (4.7)$$

Now, we use the \mathbf{V}^h -ellipticity of a_ε , the bilinearity of a_ε , a_ε^h , and the definitions of problems (W_ε) , (W_ε^h) , respectively, and we get

$$\begin{aligned} C_8(\varepsilon) \left\| \mathbf{w}_\varepsilon^h \right\|_{\mathbf{B}, \Omega}^2 &\leq a_\varepsilon^h(\mathbf{w}_\varepsilon^h, \mathbf{w}_\varepsilon^h) = a_\varepsilon(\mathbf{w}_\varepsilon^h, \mathbf{u}_\varepsilon - \mathbf{v}^h) - a_\varepsilon(\mathbf{w}_\varepsilon^h, \mathbf{u}_\varepsilon - \mathbf{v}^h) + a_\varepsilon^h(\mathbf{w}_\varepsilon^h, \mathbf{w}_\varepsilon^h) = \\ &= a_\varepsilon(\mathbf{w}_\varepsilon^h, \mathbf{u}_\varepsilon - \mathbf{v}^h) + (a_\varepsilon(\mathbf{w}_\varepsilon^h, \mathbf{v}^h) - a_\varepsilon^h(\mathbf{w}_\varepsilon^h, \mathbf{v}^h)) + \\ &\quad + (a_\varepsilon^h(\mathbf{w}_\varepsilon^h, \mathbf{u}_\varepsilon^h) - a_\varepsilon(\mathbf{w}_\varepsilon^h, \mathbf{u}_\varepsilon)) = \\ &= a_\varepsilon(\mathbf{w}_\varepsilon^h, \mathbf{u}_\varepsilon - \mathbf{v}^h) + (a_\varepsilon(\mathbf{w}_\varepsilon^h, \mathbf{v}^h) - a_\varepsilon^h(\mathbf{w}_\varepsilon^h, \mathbf{v}^h)) + \\ &\quad + (f^h(\mathbf{w}_\varepsilon^h) - f(\mathbf{w}_\varepsilon^h)). \end{aligned}$$

Dividing the latter by $\|\mathbf{w}_\varepsilon^h\|_{\mathbf{B}, \Omega}$ and using (3.48) yield

$$\begin{aligned} C_8(\varepsilon) \left\| \mathbf{w}_\varepsilon^h \right\|_{\mathbf{B}, \Omega} &\leq \max\{d, \varepsilon\} \left\| \mathbf{u}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} + \frac{|a_\varepsilon(\mathbf{v}^h, \mathbf{w}_\varepsilon^h) - a_\varepsilon^h(\mathbf{v}^h, \mathbf{w}_\varepsilon^h)|}{\|\mathbf{w}_\varepsilon^h\|_{\mathbf{B}, \Omega}} + \\ &\quad + \frac{|f(\mathbf{w}_\varepsilon^h) - f^h(\mathbf{w}_\varepsilon^h)|}{\|\mathbf{w}_\varepsilon^h\|_{\mathbf{B}, \Omega}}. \end{aligned} \quad (4.8)$$

Combining (4.5), (4.6), (4.7), and (4.8), the assertion is proved, where the constant is as follows:

$$C_9(\varepsilon) := \max \left\{ 1 + \frac{\max\{d, \varepsilon\}}{\min\{C_6, \varepsilon\}}, \frac{1}{\min\{C_6, \varepsilon\}} \right\}.$$

□

4.1.2 Finite element method

Let \mathbf{V}^h be of a finite dimension. Having a base $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of the space \mathbf{V}^h , the problem (W_ε^h) is equivalent to the following system of linear equations

$$\mathbf{A}_\varepsilon^n \cdot \mathbf{u}_\varepsilon^n = \mathbf{f}^n, \quad (4.9)$$

where the matrix $\mathbf{A}_\varepsilon^n \in \mathbb{R}^{n \times n}$ and the right-hand side vector $\mathbf{f}^n \in \mathbb{R}^n$ are as follows:

$$\mathbf{A}_\varepsilon^n := \left(a_\varepsilon^h(\mathbf{v}_i, \mathbf{v}_j) \right)_{i,j=1}^n, \quad \mathbf{f}^n := \left(f^h(\mathbf{v}_i) \right)_{i=1}^n, \quad (4.10)$$

respectively, and the solution vector $\mathbf{u}_\varepsilon^n := (u_{\varepsilon,1}^n, \dots, u_{\varepsilon,n}^n) \in \mathbb{R}^n$ corresponds to the approximate solution \mathbf{u}_ε^h in the following way

$$\mathbf{u}_\varepsilon^h = \sum_{i=1}^n u_{\varepsilon,i}^n \mathbf{v}_i.$$

The *finite element method* is a special case of the Galerkin method. The base $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of the space \mathbf{V}^h is chosen such that the matrix \mathbf{A}_ε^n is sparse. In this case the system (4.9) can be solved much faster and the matrix \mathbf{A}_ε^n takes less computer memory. The finite element method is determined as follows:

- The domain $\Omega \subset \mathbb{R}^m$ is decomposed into smaller convex subdomains, e.g., line segments for $m = 1$, triangles for $m = 2$, or tetrahedra for $m = 3$.
- The base $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is chosen as simple functions, e.g., polynomials. The space \mathbf{V}^h is called the *space of finite elements*.
- The basis functions $\mathbf{v}_1, \dots, \mathbf{v}_n$ have small supports, which make the matrix \mathbf{A}_ε^n to be sparse.

4.1.3 Discretization of the domain

We will employ polyhedral elements. To this end, we have to replace the original domain Ω by a polyhedral subdomain Ω^h . We subdivide Ω^h into a finite number of polyhedral subdomains $K_1, \dots, K_{n_{\Omega^h}}$, $n_{\Omega^h} \in \mathbb{N}$, i.e., into open and connected polyhedral subsets of Ω^h such that the following assumptions are satisfied:

- $\overline{\Omega^h} = \bigcup_{i=1}^{n_{\Omega^h}} \overline{K_i}$,
- $K_i \neq \emptyset$ for each $i = 1, \dots, n_{\Omega^h}$,
- $K_i \neq K_j \Rightarrow K_i \cap K_j = \emptyset$ for each $i, j = 1, \dots, n_{\Omega^h}$,
- any face of any K_i is either a subset of the boundary $\partial\Omega^h$ or a face of another element K_j , where $i, j = 1, \dots, n_{\Omega^h}$,
- each K_i has exactly $m + 1$ faces.

The last assumption means that in the cases of $m = 1$, $m = 2$, and $m = 3$ we deal with line segments, triangles, and tetrahedra, respectively. This assumption will provide us to introduce a reference element.

The set

$$\mathcal{T}^h := \{K_i \mid i = 1, \dots, n_{\Omega^h}\}$$

is called a *discretization* of Ω^h . In Fig. 4.1 we can see some discretizations, which were generated by Netgen, see SCHÖBERL [186]. The block column vector

$$\mathbf{x}^h := \left[\mathbf{x}_1^h, \dots, \mathbf{x}_{n_{\mathbf{x}^h}}^h \right] \in \mathbb{R}^{mn_{\mathbf{x}^h}}, \quad (4.11)$$

where $\mathbf{x}_i^h := (x_{i,1}^h, \dots, x_{i,m}^h) \in \overline{\Omega^h}$ for $i = 1, \dots, n_{\mathbf{x}^h}$, contains all the grid nodes \mathbf{x}_i^h of the discretization \mathcal{T}^h , where $n_{\mathbf{x}^h} \in \mathbb{N}$ stands for the number of the discretization nodes.

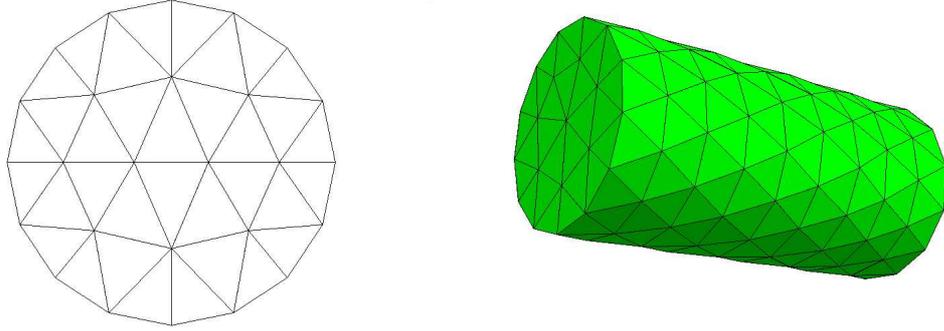


Figure 4.1: Discretization of a circle and cylinder

4.1.4 Space of finite elements

A *finite element* is a triple $e := (\overline{K}^e, \mathbf{P}^e, \Sigma^e)$, where $K^e \in \mathcal{T}^h$ is an *element domain*, $\mathbf{P}^e \subset [C^\infty(\overline{K}^e)]^{\nu_1}$ is an n_e -dimensional space of vector functions, the *finite element space* defined over \overline{K}^e , where $n_e \in \mathbb{N}$ is common for all the elements, and where $\Sigma^e := \{\sigma_1^e, \dots, \sigma_{n_e}^e\}$, $\sigma_i^e \in ([C(\overline{K}^e)]^{\nu_1})'$, is a set of n_e linearly independent continuous linear functionals. The functionals σ_i^e are called *local degrees of freedom*. The space \mathbf{P}^e usually consists of polynomials of a given order. We denote the set of finite elements by

$$E^h := \{e_i = (\overline{K}_i, \mathbf{P}^{e_i}, \Sigma^{e_i}) \mid i = 1, \dots, n_{\Omega^h}\},$$

where $K_i \equiv K^{e_i}$ stands for the same element domain.

We will introduce *global degrees of freedom*. Two adjacent elements e_i, e_j , i.e., $i \neq j$ and $\overline{K}_i \cap \overline{K}_j \neq \emptyset$, have in common some degrees of freedom. Therefore, the total number of degrees of freedom is $n < n_{\Omega^h} n_e$. We denote the set of global degrees of freedom by

$$\Sigma^h := \left\{ \sigma_i^h \in \left([C(\overline{\Omega^h})]^{\nu_1} \right)' \mid i = 1, \dots, n \right\}, \quad (4.12)$$

where the global degree of freedom σ_i^h corresponds to a local degree of freedom σ_j^e , $e \in E^h$, by means of a mapping $\mathcal{G}^e : \{1, \dots, n_e\} \mapsto \{1, \dots, n\}$ defined by

$$\mathcal{G}^e(i) = j \quad \text{if} \quad \sigma_j^h|_{([C(\overline{K}^e)]^{\nu_1})'} = \sigma_i^e, \quad i = 1, \dots, n_e, \quad j = 1, \dots, n.$$

Since $\dim(\mathbf{P}^e) = n_e$, $n_e \in \mathbb{N}$, and $\sigma_1^e, \dots, \sigma_{n_e}^e \in \Sigma^e$ are linearly independent, then there exists a basis $\{\boldsymbol{\xi}_1^e, \dots, \boldsymbol{\xi}_{n_e}^e\} \subset \mathbf{P}^e$ such that

$$\sigma_i^e(\boldsymbol{\xi}_j^e) = \delta_{i,j}, \quad i, j = 1, \dots, n_e, \quad \text{where } \delta_{i,j} := \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases}.$$

These base functions are called *shape functions*. In the same virtue, we introduce *global shape functions* $\boldsymbol{\xi}_1^h, \dots, \boldsymbol{\xi}_n^h : \overline{\Omega^h} \mapsto \mathbb{R}^{n_e}$ such that

$$\sigma_i^h(\boldsymbol{\xi}_j^h) = \delta_{i,j}, \quad i, j = 1, \dots, n. \quad (4.13)$$

The global shape functions correspond to the local ones as follows:

$$\boldsymbol{\xi}_{\mathcal{G}^e(i)}^h|_{\overline{K^e}} = \boldsymbol{\xi}_i^e, \quad e \in E^h, \quad i = 1, \dots, n_e.$$

The global shape functions form a basis for the following space

$$\mathbf{P}^h := \left\{ \mathbf{v}^h = \sum_{i=1}^n v_i \boldsymbol{\xi}_i^h \mid v_1, \dots, v_n \in \mathbb{R} \right\}. \quad (4.14)$$

Hence, the space \mathbf{P}^h consists of such functions that are elementwise in \mathbf{P}^e , i.e.,

$$\forall \mathbf{v}^h \in \mathbf{P}^h \quad \forall e \in E^h : \mathbf{v}^h|_{\overline{K^e}} \in \mathbf{P}^e$$

We need \mathbf{P}^h to be a subspace of $\mathbf{H}(\mathbf{B}; \Omega^h)$. This property is called the *conformity* of the finite elements. The following lemma gives a sufficient condition on the conformity.

Lemma 4.2. *Let $\mathbf{v}^h \in \mathbf{P}^h$. Then $\mathbf{v}^h \in \mathbf{H}(\mathbf{B}; \Omega^h)$ if for any two adjacent elements $e_i, e_j \in E^h$, $i \neq j$, with a common face $f_{i,j} := \overline{K_i} \cap \overline{K_j}$, the trace γ is continuous over $f_{i,j}$, i.e.,*

$$\gamma_{K_i}(\mathbf{v}^h|_{\overline{K_i}})|_{f_{i,j}} = -\gamma_{K_j}(\mathbf{v}^h|_{\overline{K_j}})|_{f_{i,j}}, \quad (4.15)$$

where γ is given by (3.23). Note that the minus sign appeared, since the outer unit normal vectors on $f_{i,j}$ satisfy $\mathbf{n}_{K_i} = -\mathbf{n}_{K_j}$.

Proof. Let $\mathbf{v}^h \in \mathbf{P}^h$ and let (4.15) holds. Clearly $\mathbf{v}^h \in [L^2(\Omega^h)]^{\nu_1}$. We will prove that $\mathbf{v}^h \in \mathbf{H}(\mathbf{B}; \Omega^h)$ by means of (3.32). We take $\mathbf{z}^h \in [L^2(\Omega^h)]^{\nu_2}$ such that

$$\mathbf{z}^h|_{K_i} := \mathbf{B}(\mathbf{v}^h|_{K_i}), \quad i = 1, \dots, n_{\Omega^h}.$$

Let $\mathbf{w}^h \in [C_0^\infty(\Omega^h)]^{\nu_2}$ be arbitrary. Then, Corollary 3.3 and (4.15) yield

$$\begin{aligned} \int_{\Omega^h} \mathbf{v}^h \cdot \mathbf{B}^*(\mathbf{w}^h) \, d\mathbf{x} &= \sum_{K_i \in \mathcal{T}^h} \int_{K_i} \mathbf{v}^h \cdot \mathbf{B}^*(\mathbf{w}^h) \, d\mathbf{x} = \\ &= \sum_{K_i \in \mathcal{T}^h} \left(- \int_{K_i} \mathbf{z}^h \cdot \mathbf{w}^h \, d\mathbf{x} + \int_{\partial K_i} \gamma_{K_i}(\mathbf{v}^h) \cdot \mathbf{w}^h \, ds \right) = \\ &= - \int_{\Omega^h} \mathbf{z}^h \cdot \mathbf{w}^h \, d\mathbf{x} + \sum_{f_{i,j}} \left(\int_{f_{i,j}} \gamma_{K_i}(\mathbf{v}^h) \cdot \mathbf{w}^h \, ds + \int_{f_{i,j}} \gamma_{K_j}(\mathbf{v}^h) \cdot \mathbf{w}^h \, ds \right) = \\ &= - \int_{\Omega^h} \mathbf{z}^h \cdot \mathbf{w}^h \, d\mathbf{x}, \end{aligned}$$

which completes the proof. \square

The next assumption ensures the conformity of the finite elements, i.e., $\mathbf{P}^h \subset \mathbf{H}(\mathbf{B}; \Omega^h)$. We refer to HIPTMAIR [92] for a unified way of the design of conforming finite elements.

Assumption 4.1. *Let $e \in E^h$ be an element and let $f \subset \partial K^e$ denote a face. We assume that the degrees of freedom connected to the face f are exactly the ones which determine the trace $\gamma_{K^e}(\mathbf{v}^e)|_f$, where $\mathbf{v}^e \in \mathbf{P}^e$.*

Further, we introduce the set of indices of those degrees of freedom which determine the trace γ_{Ω^h} . Due to (4.14), we can write an arbitrary $\mathbf{v}^h \in \mathbf{P}^h$ as

$$\mathbf{v}^h = \sum_{i=1}^n v_i^h \boldsymbol{\xi}_i^h,$$

where $v_i^h \in \mathbb{R}$ for $i = 1, \dots, n$ and where $\boldsymbol{\xi}_i^h$ denote the global shape functions. Then,

$$\gamma_{\Omega^h}(\mathbf{v}^h) = \sum_{i=1}^n v_i^h \gamma_{\Omega^h}(\boldsymbol{\xi}_i^h).$$

Therefore, the trace is determined by the following set of indices

$$\mathcal{I}_0^h := \left\{ i \in \{1, \dots, n\} \mid \gamma_{\Omega^h}(\boldsymbol{\xi}_i^h) \neq \mathbf{0} \right\}. \quad (4.16)$$

Then, the finite element space $\mathbf{V}^h \equiv \mathbf{H}_0(\mathbf{B}; \Omega^h)^h \subset \mathbf{H}_0(\mathbf{B}; \Omega^h)$ is defined by

$$\mathbf{H}_0(\mathbf{B}; \Omega^h)^h := \left\{ \mathbf{v}^h \in \mathbf{P}^h \mid \forall i \in \mathcal{I}_0^h : \sigma_i^h(\mathbf{v}^h) = 0 \right\}. \quad (4.17)$$

4.1.5 Finite element discretization of the weak formulation

Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$. We rewrite the problem (W_ε) as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{u}_\varepsilon(\Omega) \in \mathbf{H}_0(\mathbf{B}; \Omega): \\ a_\varepsilon(\mathbf{v}, \mathbf{u}_\varepsilon(\Omega)) = f(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_0(\mathbf{B}; \Omega) \end{array} \right\}. \quad (W_\varepsilon(\Omega))$$

As long as $\Omega \in \mathcal{L}$, the existence of the unique solution $\mathbf{u}_\varepsilon(\Omega)$ to $(W_\varepsilon(\Omega))$ is given by Theorem 3.17.

Further, let $h > 0$ be a *discretization parameter* and let $\Omega^h \subset \Omega$ be a nonempty polyhedral subdomain. Then, $\Omega^h \in \mathcal{L}$. Let \mathcal{T}^h be a discretization of Ω^h and let E^h be the corresponding set of finite elements. Concerning the bilinear form a_ε^h and the linear functional f^h given by (4.1) and (4.2), respectively, we assume that for each $e \in E^h$ there exist a constant matrix $\mathbf{D}^e \in \mathbb{R}^{\nu_2 \times \nu_2}$ and vector $\mathbf{f}^e \in \mathbf{R}^{\nu_1}$ such that

$$\forall \mathbf{x} \in K^e : \mathbf{D}^h(\mathbf{x}) = \mathbf{D}^e \text{ and } \mathbf{f}^h(\mathbf{x}) = \mathbf{f}^e. \quad (4.18)$$

The Galerkin approximation of the problem $(W_\varepsilon(\Omega^h))$ reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{u}_\varepsilon^h(\Omega^h) \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h : \\ a_\varepsilon^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^h(\Omega^h)) = f^h(\mathbf{v}^h) \quad \forall \mathbf{v}^h \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h \end{array} \right\}, \quad (W_\varepsilon^h(\Omega^h))$$

where a_ε^h , f^h , and $\mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ are respectively defined by (4.1), (4.2), and by (4.17). The existence and uniqueness of the solution $\mathbf{u}_\varepsilon^h(\Omega^h)$ to $(W_\varepsilon^h(\Omega^h))$ follows from Theorem 4.1.

4.2 Assembling finite elements

In the previous sections, we provided a discretization technique of the weak formulation ($W_\varepsilon(\Omega)$). The aim of this section is to build up an algorithm for an efficient assembling of the finite element matrix \mathbf{A}_ε^n and the right-hand side vector \mathbf{f}^n , both defined by (4.10).

Let us consider the global shape (base) functions $\boldsymbol{\xi}_1^h, \dots, \boldsymbol{\xi}_n^h \in \mathbf{P}^h$. They have small supports, since the shape function $\boldsymbol{\xi}_i^h$ is nonzero just for those neighbouring elements which have a degree of freedom σ_i^h in common, i.e.,

$$\text{supp } \boldsymbol{\xi}_i^h \subset \bigcup_{e \in E_i^h} \overline{K^e}, \quad i = 1, \dots, n, \quad (4.19)$$

where

$$E_i^h := \left\{ e \in E^h \mid \exists j \in \{1, \dots, n_e\} : \mathcal{G}^e(j) = i \right\}, \quad i = 1, \dots, n, \quad (4.20)$$

is the set of the elements neighbouring with e_i . Since a_ε^h is a bilinear form and f^h is a linear functional, we assemble the matrix \mathbf{A}_ε^n and the right-hand side vector \mathbf{f}^n , see (4.10), elementwise. Due to (4.19), each element contributes only by its n_e global degrees of freedom, i.e.,

$$(\mathbf{A}_\varepsilon^n)_{i,j} = \sum_{e \in E_i^h \cap E_j^h} \sum_{k,l=1}^{n_e} a_\varepsilon^e(\boldsymbol{\xi}_k^e, \boldsymbol{\xi}_l^e), \quad (\mathbf{f}^n)_i = \sum_{e \in E_i^h} \sum_{k=1}^{n_e} f^e(\boldsymbol{\xi}_k^e), \quad i, j = 1, \dots, n, \quad (4.21)$$

where the local contributions to the matrix and to the right-hand side vector are

$$a_\varepsilon^e(\boldsymbol{\xi}_k^e, \boldsymbol{\xi}_l^e) := \int_{K^e} \mathbf{B}(\boldsymbol{\xi}_k^e) \cdot (\mathbf{D}^e \cdot \mathbf{B}(\boldsymbol{\xi}_l^e)) \, d\mathbf{x} + \varepsilon \int_{K^e} \boldsymbol{\xi}_k^e \cdot \boldsymbol{\xi}_l^e \, d\mathbf{x}, \quad k, l = 1, \dots, n_e, \quad (4.22)$$

$$f^e(\boldsymbol{\xi}_k^e) := \int_{K^e} \mathbf{f}^e \cdot \boldsymbol{\xi}_k^e \, d\mathbf{x}, \quad k = 1, \dots, n_e, \quad (4.23)$$

respectively. The solution to the problem ($W_\varepsilon^h(\Omega^h)$) is then given by

$$\mathbf{u}_\varepsilon^h(\Omega^h) := \sum_{i=1}^n u_{\varepsilon,i}^n \boldsymbol{\xi}_i^h, \quad (4.24)$$

where $\mathbf{u}_\varepsilon^n := (u_{\varepsilon,1}^n, \dots, u_{\varepsilon,n}^n) \in \mathbb{R}^n$ denotes the solution to the linear system (4.9).

4.2.1 Reference element

As each element domain K^e is a polyhedron of $m+1$ faces, it can be uniquely described by the following block column vector, which consists of the $m+1$ corners

$$\mathbf{x}^e := [\mathbf{x}_1^e, \dots, \mathbf{x}_{m+1}^e] \in \mathbb{R}^{m(m+1)}, \quad (4.25)$$

where $\mathbf{x}_i^e := (x_{i,1}^e, \dots, x_{i,m}^e) \in \overline{K^e}$ for $i = 1, \dots, m+1$. To each element $e \in E^h$ we associate a mapping $\mathcal{H}^e : \{1, \dots, m+1\} \mapsto \{1, \dots, n_{\mathbf{x}^h}\}$ which maps the element nodal indices to the global ones as follows:

$$\mathcal{H}^e(i) = j \quad \text{if} \quad \mathbf{x}_i^e = \mathbf{x}_j^h, \quad i = 1, \dots, m+1, \quad j = 1, \dots, n_{\mathbf{x}^h}. \quad (4.26)$$

We further introduce a reference element $r := (\overline{K}^r, \mathbf{P}^r, \Sigma^r)$ such that the polyhedral domain K^r is determined by the following block column vector consisting of the reference corners

$$\widehat{\mathbf{x}}^r := [\widehat{\mathbf{x}}_1^r, \dots, \widehat{\mathbf{x}}_{m+1}^r] \in \mathbb{R}^{m(m+1)},$$

where $\widehat{\mathbf{x}}_i^r := (x_{i,1}^r, \dots, x_{i,m}^r) \in \overline{K}^r$, $i = 1, \dots, m+1$, and where $\dim(\mathbf{P}^r) = \dim(\mathbf{P}^e) = n_e$. To each element $e \in E^h$, we associate a one-to-one linear mapping $\mathcal{R}^e : \overline{K}^r \mapsto \overline{K}^e$ defined by

$$\mathbf{x} := \mathcal{R}^e(\widehat{\mathbf{x}}) := \mathbf{R}^e \cdot \widehat{\mathbf{x}} + \mathbf{r}^e, \quad \mathbf{x} \in \overline{K}^e, \widehat{\mathbf{x}} \in \overline{K}^r, \quad (4.27)$$

where $\mathbf{R}^e \in \mathbb{R}^{m \times m}$ is a nonsingular matrix and $\mathbf{r}^e \in \mathbb{R}^m$ is a vector both of which are uniquely determined by \mathbf{x}^e as follows:

$$\mathbf{R}^e \cdot \widehat{\mathbf{x}}_i^r + \mathbf{r}^e = \mathbf{x}_i^e, \quad i = 1, \dots, m+1. \quad (4.28)$$

Obviously, both \mathbf{R}^e and \mathbf{r}^e are continuously differentiable with respect to each coordinate of the corners \mathbf{x}^e .

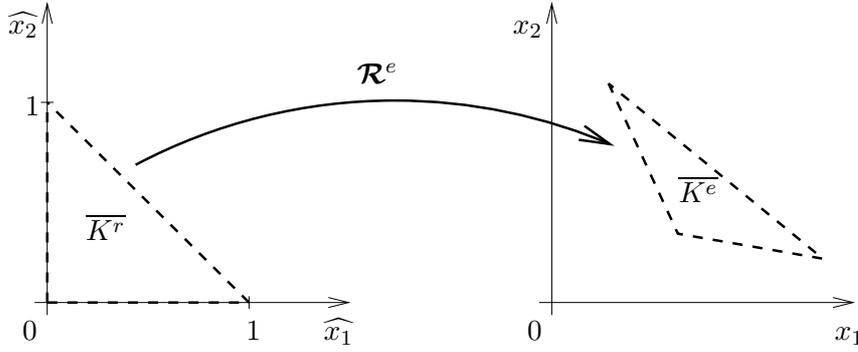


Figure 4.2: A transformation between the reference and an element domain

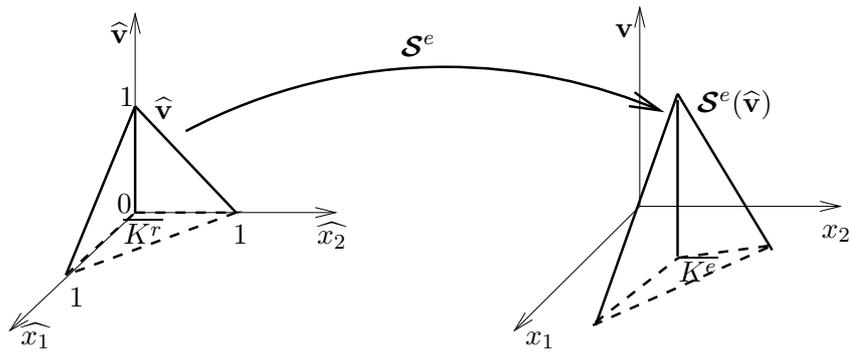


Figure 4.3: A transformation between the reference and an element shape function

Further, let us denote by $\widehat{\xi}_1^r, \dots, \widehat{\xi}_{n_e}^r$ the shape functions acting on the reference element r , i.e.,

$$\sigma_i^r(\widehat{\xi}_j^r) = \delta_{i,j}, \quad i, j = 1, \dots, n_e.$$

Assumption 4.2. We assume that there exist nonsingular matrices $\mathbf{S}^e \in \mathbb{R}^{\nu_1 \times \nu_1}$ and $\mathbf{S}_B^e \in \mathbb{R}^{\nu_2 \times \nu_2}$, both of which are continuously differentiable with respect to the corners \mathbf{x}^e , such that

$$\mathbf{S}^e \cdot \widehat{\boldsymbol{\xi}}_i^r(\widehat{\mathbf{x}}) = \boldsymbol{\xi}_i^e(\mathbf{x}), \quad i = 1, \dots, n_e,$$

and

$$\mathbf{S}_B^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_i^r(\widehat{\mathbf{x}})) = \mathbf{B}_{\mathbf{x}}(\boldsymbol{\xi}_i^e(\mathbf{x})), \quad i = 1, \dots, n_e,$$

where $\mathbf{x} := \mathcal{R}^e(\widehat{\mathbf{x}})$, and where $\mathbf{B}_{\mathbf{x}}$ and $\mathbf{B}_{\widehat{\mathbf{x}}}$, respectively, stand for the differential operator \mathbf{B} , defined by (3.21), with respect to the global coordinates \mathbf{x} and with respect to the reference coordinates $\widehat{\mathbf{x}}$. We define transformations $\mathcal{S}^e : \mathbf{P}^r \mapsto \mathbf{P}^e$ and $\mathcal{S}_B^e : [L^2(K^r)]^{\nu_2} \mapsto [L^2(K^e)]^{\nu_2}$ by

$$\mathcal{S}^e(\widehat{\mathbf{v}}(\widehat{\mathbf{x}})) := \mathbf{S}^e \cdot \widehat{\mathbf{v}}(\widehat{\mathbf{x}}) \text{ and } \mathcal{S}_B^e(\mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\mathbf{v}}(\widehat{\mathbf{x}}))) := \mathbf{S}_B^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\mathbf{v}}(\widehat{\mathbf{x}})),$$

where $\widehat{\mathbf{v}}(\widehat{\mathbf{x}}) \in \mathbf{P}^r$.

The linear transformations \mathcal{S}_B^e and \mathcal{S}^e are associated to the differential operator \mathbf{B} and to the identity operator, respectively. In general, the theory of differential forms, cf. HIPTMAIR [91], can be used in order to derive a canonical transformation, see HIPTMAIR [92], which is related to some differential operator and to some degrees of freedom.

4.2.2 BDB integrators

Making use of the reference element, the integration in (4.22) and (4.23) can be unified by the substitutions \mathcal{R}^e and \mathcal{S}_B^e in the first term of (4.22), and by the substitutions \mathcal{R}^e and \mathcal{S}^e in the second term of (4.22) and in (4.23), as follows:

$$\begin{aligned} a_\varepsilon^e(\boldsymbol{\xi}_k^e, \boldsymbol{\xi}_l^e) &= \int_{K^r} \left(\mathbf{S}_B^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_k^r) \right) \cdot \left(\mathbf{D}^e \cdot \left(\mathbf{S}_B^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_l^r) \right) \right) |\det(\mathbf{R}^e)| d\widehat{\mathbf{x}} + \\ &+ \varepsilon \int_{K^r} \left(\mathbf{S}^e \cdot \widehat{\boldsymbol{\xi}}_k^r \right) \cdot \left(\mathbf{S}^e \cdot \widehat{\boldsymbol{\xi}}_l^r \right) |\det(\mathbf{R}^e)| d\widehat{\mathbf{x}}, \end{aligned} \quad (4.29)$$

$$f^e(\boldsymbol{\xi}_k^e) = \int_{K^r} \mathbf{f}^e \cdot \left(\mathbf{S}^e \cdot \widehat{\boldsymbol{\xi}}_k^r \right) |\det(\mathbf{R}^e)| d\widehat{\mathbf{x}}. \quad (4.30)$$

Now we employ the Gaussian quadrature method, cf. RALSTON [163], CIARLET AND LI-ONS [46]. Having a sufficient number of Gaussian integration points, we can calculate the integrals exactly. Then, the matrix and the right-hand side vector in (4.10) are evaluated elementwise, where the contributions of the elements are

$$\begin{aligned} \mathbf{A}_\varepsilon^e &:= \sum_{i=1}^{n_G} w_i^G \mathbf{B}_B^e(\widehat{\mathbf{x}}_i^G)^T \cdot \mathbf{D}^e \cdot \mathbf{B}_B^e(\widehat{\mathbf{x}}_i^G) + \varepsilon \sum_{i=1}^{n_G} w_i^G \mathbf{B}^e(\widehat{\mathbf{x}}_i^G)^T \cdot \mathbf{I}^e \cdot \mathbf{B}^e(\widehat{\mathbf{x}}_i^G), \\ \mathbf{f}^e &:= \sum_{i=1}^{n_G} w_i^G \mathbf{B}^e(\widehat{\mathbf{x}}_i^G)^T \cdot \mathbf{g}^e, \end{aligned}$$

where $\widehat{\mathbf{x}}_1^G, \dots, \widehat{\mathbf{x}}_{n_G}^G \in \overline{K^r}$ are the Gaussian integration points, $w_1^G, \dots, w_{n_G}^G \in \mathbb{R}$ are the Gaussian integration weights, and where

$$\mathbf{B}_B^e(\widehat{\mathbf{x}}) := \mathbf{S}_B^e \cdot \left[\mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_1^r(\widehat{\mathbf{x}})), \dots, \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_{n_e}^r(\widehat{\mathbf{x}})) \right], \quad \mathbf{B}^e(\widehat{\mathbf{x}}) := \mathbf{S}^e \cdot \left[\widehat{\boldsymbol{\xi}}_1^r(\widehat{\mathbf{x}}), \dots, \widehat{\boldsymbol{\xi}}_{n_e}^r(\widehat{\mathbf{x}}) \right],$$

$$\begin{aligned} \mathbf{D}^e &:= |\det(\mathbf{R}^e)| \cdot \mathbf{D}^e, & \mathbf{I}^e &:= |\det(\mathbf{R}^e)| \cdot \mathbf{I}_{n_e}, \\ \mathbf{g}^e &:= |\det(\mathbf{R}^e)| \cdot \mathbf{f}^e, \end{aligned}$$

where $\mathbf{I}_{n_e} \in \mathbb{R}^{n_e \times n_e}$ is the unit matrix. Note that we will employ the lowest-order, i.e., linear, finite elements. In this case, since both \mathbf{D}^h and \mathbf{f}^h are elementwise constant, then all the integrands are linear over the element, thus, we will employ only one Gaussian point being the mass point of K^r , and the corresponding weight. These are respectively as follows:

$$\mathbf{x}_1^G := \begin{cases} \frac{1}{2} & , \text{ for } m = 1 \\ \left(\frac{1}{3}, \frac{1}{3}\right) & , \text{ for } m = 2, \\ \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right) & , \text{ for } m = 3 \end{cases}, \quad w_1^G := 1.$$

The name *BDB integrators* is due to the structure of the contributions to the bilinear form. The differential operator is involved in the matrix \mathbf{B}^e while the matrix \mathbf{D}^e or \mathbf{I}^e provide the material properties and geometrical parameters of the element domain K^e .

4.2.3 The algorithm

The structure of the BDB integrators offers an efficient implementation using the object-oriented technologies, cf. KUHN, LANGER, AND SCHÖBERL [117]. Algorithm 1 describes assembling the whole system matrix \mathbf{A}_ε^n and the right-hand side vector \mathbf{f}^n , which is called *postprocessing*. When assembling the matrix and the right-hand side vector, we have to omit those rows and columns whose corresponding degrees of freedom are connected to the boundary $\partial\Omega^h$ along which the zero trace is prescribed. This is done by setting all those rows and columns to zero except for the diagonal entries in the matrix \mathbf{A}_ε^n .

Algorithm 1 Finite element method: preprocessing

```

 $\mathbf{A}_\varepsilon^n := \mathbf{0}, \mathbf{f}^n := \mathbf{0}$ 
for  $i := 1, \dots, n_{\Omega^h}$  do
  Evaluate  $\mathbf{A}_\varepsilon^{e_i}, \mathbf{f}^{e_i}$ 
  for  $j := 1, \dots, n_e$  do
    for  $k := 1, \dots, n_e$  do
      if  $j = k$  or  $(\mathcal{G}^{e_i}(j) \notin \mathcal{I}_0^h \text{ and } \mathcal{G}^{e_i}(k) \notin \mathcal{I}_0^h)$  then
         $(\mathbf{A}_\varepsilon^n)_{\mathcal{G}^{e_i}(j), \mathcal{G}^{e_i}(k)} := (\mathbf{A}_\varepsilon^n)_{\mathcal{G}^{e_i}(j), \mathcal{G}^{e_i}(k)} + (\mathbf{A}_\varepsilon^{e_i})_{j,k}$ 
      end if
    end for
  if  $\mathcal{G}^{e_i}(j) \notin \mathcal{I}_0^h$  then
     $(\mathbf{f}^n)_{\mathcal{G}^{e_i}(j)} := (\mathbf{f}^n)_{\mathcal{G}^{e_i}(j)} + (\mathbf{f}^{e_i})_j$ 
  end if
  end for
end for

```

The approximate solution $\mathbf{u}_\varepsilon^h(\Omega^h)$ to the problem $(W_\varepsilon^h(\Omega^h))$, discretized by the finite element method, is given by (4.24) while we solve the system

$$\mathbf{A}_\varepsilon^n \cdot \mathbf{u}_\varepsilon^n = \mathbf{f}^n$$

for $\mathbf{u}_\varepsilon^n := (u_{\varepsilon,1}^n, \dots, u_{\varepsilon,n}^n) \in \mathbb{R}^n$. In fact, we rather look for $\mathbf{B}(\mathbf{u}_\varepsilon^h(\mathbf{x}))$, which is elementwise constant, since we have employed the lowest, i.e., the first-order finite elements only. Therefore,

we can describe $\mathbf{B}(\mathbf{u}_\varepsilon^h(\mathbf{x}))$ by the block column vector

$$\mathbf{B}_\varepsilon^n := \left[\mathbf{B}_\varepsilon^{n,e_1}, \dots, \mathbf{B}_\varepsilon^{n,e_{n_{\Omega^h}}} \right] \in \mathbb{R}^{\nu_2 n_{\Omega^h}}$$

such that

$$\mathbf{B}(\mathbf{u}_\varepsilon^h(\mathbf{x}))|_{K_i} = \mathbf{B}_\varepsilon^{n,e_i},$$

where

$$\mathbf{B}_\varepsilon^{n,e_i} := \sum_{j=1}^{n_e} u_{\varepsilon, \mathcal{G}^{e_i(j)}}^n \mathbf{S}_\mathbf{B}^{e_i} \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}})) \in \mathbb{R}^{\nu_2} \quad \text{for } i = 1, \dots, n_{\Omega^h}.$$

This procedure which assembles the vector \mathbf{B}_ε^n is called *postprocessing* and it is depicted in Algorithm 2.

Algorithm 2 Finite element method: postprocessing

```

Given  $\mathbf{u}_\varepsilon^n$ 
 $\mathbf{B}_\varepsilon^n := \mathbf{0}$ 
for  $i := 1, \dots, n_{\Omega^h}$  do
   $\mathbf{B}_\varepsilon^{n,e_i} := \mathbf{0}$ 
  for  $j := 1, \dots, n_e$  do
    Evaluate  $\mathbf{B}_\varepsilon^{n,e_i} := \mathbf{B}_\varepsilon^{n,e_i} + u_{\varepsilon, \mathcal{G}^{e_i(j)}}^n \mathbf{S}_\mathbf{B}^{e_i} \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}}))$ 
  end for
   $k := \nu_2(i - 1)$ 
  for  $j := 1, \dots, \nu_2$  do
     $[\mathbf{B}_\varepsilon^n]_{k+j} := [\mathbf{B}_\varepsilon^{n,e_i}]_j$ 
  end for
end for

```

4.3 Approximation properties

Now, we specify the meaning of the discretization parameter h in the formulation ($W_\varepsilon^h(\Omega^h)$). To each element domain K^e we associate an *element discretization parameter* h^e which is the maximum edge size of $\overline{K^e}$, i.e., the length of the line segment, the maximum side of the triangle, or the maximum edge of the tetrahedron in the cases of $m = 1$, $m = 2$, or $m = 3$, respectively. The (*global*) *discretization parameter* h is defined by

$$h := \max_{e \in E^h} h^e. \quad (4.31)$$

Convention 4.1. *In what follows, we will assume that there exists $\bar{h} > 0$, being, e.g., the minimum diameter of a sphere (or circle) containing Ω , such that any considered discretization parameter h fulfills*

$$h \leq \bar{h}. \quad (4.32)$$

The aim of this section is to prove a convergence, in some sense, of the approximate finite element solutions $\mathbf{u}_\varepsilon^h(\Omega^h)$ to the true solution $\mathbf{u}_\varepsilon(\Omega)$.

4.3.1 Approximation of the domain by polyhedra

We employ polyhedral elements. Therefore, we have to deal with an approximation of the original domain Ω by polyhedra and with a convergence of solutions over these polyhedra.

We introduce an extension operator. Let $h > 0$, $\nu \in \mathbb{N}$, and let $\Omega^h \subset \Omega$ be a nonempty polyhedral subdomain. We define a linear *extension operator* $\mathbf{X}_\nu^h : [L^2(\Omega^h)]^\nu \mapsto [L^2(\Omega)]^\nu$ by

$$\mathbf{X}_\nu^h(\mathbf{v}(\mathbf{x})) := \begin{cases} \mathbf{v}(\mathbf{x}) & , \mathbf{x} \in \Omega^h \\ \mathbf{0} & , \mathbf{x} \in \Omega \setminus \Omega^h \end{cases}, \quad \mathbf{v}(\mathbf{x}) \in [L^2(\Omega^h)]^\nu. \quad (4.33)$$

Lemma 4.3. *Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$ be a domain, let $\Omega^h \subset \Omega$ be its nonempty polyhedral subdomain, and let $\mathbf{v}^h \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$, where $\mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ is defined by (4.17). Then $\mathbf{X}_{\nu_1}^h(\mathbf{v}^h) \in \mathbf{H}_0(\mathbf{B}; \Omega)$ and the space*

$$\mathbf{X}_0(\mathbf{B}; \Omega; \Omega^h)^h := \left\{ \mathbf{X}_{\nu_1}^h(\mathbf{v}^h) \in \mathbf{H}_0(\mathbf{B}; \Omega) \mid \mathbf{v}^h \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h \right\} \quad (4.34)$$

is a closed subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$.

Proof. Let $\mathbf{v}^h \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ be arbitrary. We denote $\mathbf{b}^h := \mathbf{X}_{\nu_2}^h(\mathbf{B}(\mathbf{v}^h))$. Clearly, $\mathbf{b}^h \in [L^2(\Omega)]^{\nu_2}$. Let $\varphi \in [C_0^\infty(\Omega)]^{\nu_2}$, then $\varphi|_{\Omega^h} \in [H^1(\Omega^h)]^{\nu_2}$. Now the definition (4.33), Assumption 3.4, and $\gamma_{\Omega^h}(\mathbf{v}^h) = \mathbf{0}$, which is the trace along $\partial\Omega^h$, yield

$$\begin{aligned} \int_\Omega \mathbf{b}^h \cdot \varphi \, d\mathbf{x} &= \int_{\Omega^h} \mathbf{B}(\mathbf{v}^h) \cdot \varphi|_{\Omega^h} \, d\mathbf{x} = - \int_{\Omega^h} \mathbf{v}^h \cdot \mathbf{B}^*(\varphi|_{\Omega^h}) \, d\mathbf{x} + \langle \gamma_{\Omega^h}(\mathbf{v}^h), \varphi|_{\Omega^h} \rangle_{\partial\Omega^h} = \\ &= - \int_\Omega \mathbf{X}_{\nu_1}^h(\mathbf{v}^h) \cdot \mathbf{B}^*(\varphi) \, d\mathbf{x}. \end{aligned}$$

The latter implies that $\mathbf{b}^h = \mathbf{B}(\mathbf{X}_{\nu_1}^h(\mathbf{v}^h))$ and $\mathbf{X}_{\nu_1}^h(\mathbf{v}^h) \in \mathbf{H}(\mathbf{B}; \Omega)$. Further, let $\psi \in [H^1(\Omega)]^{\nu_2}$, then $\psi|_{\Omega^h} \in [H^1(\Omega^h)]^{\nu_2}$. By (4.33), by Assumption 3.4, and since $\gamma_{\Omega^h}(\mathbf{v}^h) = \mathbf{0}$, we get

$$\begin{aligned} \langle \gamma_\Omega(\mathbf{X}_{\nu_1}^h(\mathbf{v}^h)), \psi \rangle_{\partial\Omega} &= \int_\Omega \mathbf{B}(\mathbf{X}_{\nu_1}^h(\mathbf{v}^h)) \cdot \psi \, d\mathbf{x} + \int_\Omega \mathbf{X}_{\nu_1}^h(\mathbf{v}^h) \cdot \mathbf{B}^*(\psi) \, d\mathbf{x} = \\ &= \int_{\Omega^h} \mathbf{B}(\mathbf{v}^h) \cdot \psi|_{\Omega^h} \, d\mathbf{x} + \int_{\Omega^h} \mathbf{v}^h \cdot \mathbf{B}^*(\psi|_{\Omega^h}) \, d\mathbf{x} = \\ &= \langle \gamma_{\Omega^h}(\mathbf{v}^h), \psi|_{\Omega^h} \rangle_{\partial\Omega^h} = 0, \end{aligned}$$

where $\gamma_\Omega : \mathbf{H}(\mathbf{B}; \Omega) \mapsto [H^{-1/2}(\partial\Omega)]^{\nu_2}$ stands for the trace operator along $\partial\Omega$. Therefore, $\mathbf{X}_{\nu_1}^h(\mathbf{v}^h) \in \mathbf{H}_0(\mathbf{B}; \Omega)$. Since $\mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ is a finite-dimensional Hilbert space (of the dimension less than n) and $\mathbf{X}_{\nu_1}^h : \mathbf{H}_0(\mathbf{B}; \Omega^h)^h \mapsto \mathbf{H}_0(\mathbf{B}; \Omega)$ is a linear mapping, then the set $\mathbf{X}_0(\mathbf{B}; \Omega; \Omega^h)^h$, defined by (4.34), is obviously a closed subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$, hence, again a finite-dimensional Hilbert space. \square

Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$ be a domain and let $h > 0$ be a discretization parameter. We say that the class $\{\Omega^h\}_{h>0}$, $\Omega^h \subset \Omega$ approximates Ω from the inner if the following is satisfied

$$\forall \mathbf{x}^h \in \partial\Omega^h \exists \mathbf{x} \in \partial\Omega : \|\mathbf{x}^h - \mathbf{x}\| \leq h. \quad (4.35)$$

We denote this convergence by $\Omega^h \nearrow \Omega$, as $h \rightarrow 0_+$. Let us further introduce a function $\chi_{\Omega^h} : \Omega \mapsto \{0, 1\}$, which is called the *characteristic function* of Ω^h , by

$$\chi_{\Omega^h}(\mathbf{x}) := \begin{cases} 1 & , \mathbf{x} \in \Omega^h \\ 0 & , \mathbf{x} \in \Omega \setminus \Omega^h \end{cases}. \quad (4.36)$$

It is obvious that if $\Omega^h \nearrow \Omega$, then

$$\chi_{\Omega^h}(\mathbf{x}) \rightarrow 1 \text{ a.e. in } \Omega, \text{ as } h \rightarrow 0_+.$$

Assumption 4.3. *We assume that $\Omega^h \nearrow \Omega$.*

4.3.2 A–priori error estimate

We introduce an *interpolation operator* $\boldsymbol{\pi}^e : [C^\infty(\overline{K^e})]^{\nu_1} \mapsto \mathbf{P}^e$ such that

$$\sigma_i^e(\boldsymbol{\pi}^e(\mathbf{v})) = \sigma_i^e(\mathbf{v}), \quad i = 1, \dots, n_e,$$

holds for any $\mathbf{v} \in [C^\infty(\overline{K^e})]^{\nu_1}$. Further, we introduce a *global interpolation operator* $\boldsymbol{\pi}^h : [C^\infty(\overline{\Omega^h})]^{\nu_1} \mapsto \mathbf{P}^h$ such that for any $\mathbf{v} \in [C^\infty(\overline{\Omega^h})]^{\nu_1}$

$$\sigma_i^h(\boldsymbol{\pi}^h(\mathbf{v})) = \sigma_i^h(\mathbf{v}), \quad i = 1, \dots, n, \quad (4.37)$$

or we can introduce that equivalently by

$$\boldsymbol{\pi}^h(\mathbf{v})|_{K^e} := \boldsymbol{\pi}^e(\mathbf{v}|_{K^e}), \quad K^e \in \mathcal{T}^h, \quad (4.38)$$

where \mathbf{P}^h is due to (4.14). Moreover, another global interpolation operator $\boldsymbol{\pi}_0^h : [C^\infty(\overline{\Omega^h})]^{\nu_1} \mapsto \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ is introduced such that for any $\mathbf{v} \in [C^\infty(\overline{\Omega^h})]^{\nu_1}$

$$\sigma_i^h(\boldsymbol{\pi}_0^h(\mathbf{v})) := \begin{cases} \sigma_i^h(\mathbf{v}) & , i \notin \mathcal{I}_0^h \\ 0 & , i \in \mathcal{I}_0^h \end{cases},$$

where \mathcal{I}_0^h is defined by (4.16).

We suppose that the following *a–priori error estimate* holds. For more results on a–priori error estimates see CIARLET [45], KŘÍŽEK AND NEITTAANMÄKI [115], STRANG AND FIX [200].

Assumption 4.4. *We assume that there exists a positive constant $C_{10} \equiv C_{10}(K^e)$ such that*

$$\forall \mathbf{v} \in [H^2(K^e)]^{\nu_1} : \|\mathbf{v} - \boldsymbol{\pi}^e(\mathbf{v})\|_{\mathbf{B}, K^e} \leq C_{10} h^e \|\mathbf{v}\|_{\nu_1, 2, K^e}.$$

4.3.3 Regular discretizations

We suppose that \mathcal{T}^h are *regular discretizations* in the sense of the following three assumptions.

Assumption 4.5. *We assume that there exists a positive constant C_{11} such that*

$$\forall h > 0 \forall K^e \in \mathcal{T}^h : C_{10}(K^e) \leq C_{11}.$$

Assumption 4.6. We assume that for each $\mathbf{v} \in [C_0^\infty(\Omega)]^{\nu_1}$ there exist positive constants $C_{12} \equiv C_{12}(\mathbf{v})$ and $C_{13} \equiv C_{13}(\mathbf{v})$ such that

$$\forall h > 0 \forall K^e \in \mathcal{T}^h : K^e \subset \Delta^h \Omega^h \Rightarrow |\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| \|\mathbf{S}^e\| \leq C_{12} \text{ and } |\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| \|\mathbf{S}_B^e\| \leq C_{13},$$

where $j = 1, \dots, n_e$ and where

$$\Delta^h \Omega^h := \left\{ \mathbf{y} \in \Omega^h \mid \exists K^e \in \mathcal{T}^h \exists j \in \{1, \dots, n_e\} : \mathbf{y} \in \overline{K^e} \text{ and } \mathcal{G}^e(j) \in \mathcal{I}_0^h \right\}. \quad (4.39)$$

is the most outer layer of finite elements.

Assumption 4.7. We assume that for each $\mathbf{v} \in [C_0^\infty(\Omega)]^{\nu_1}$ there exists a positive constant $C_{14} \equiv C_{14}(\mathbf{v})$ such that

$$\forall h > 0 \forall K^e \in \mathcal{T}^h \forall \mathbf{x} \equiv \mathcal{R}^e(\widehat{\mathbf{x}}) \in \overline{K^e} : \left\| \mathbf{B}_x(\boldsymbol{\pi}^e(\mathbf{v}|_{\overline{K^e}}(\mathbf{x}))) \right\| = \left\| \sum_{i=1}^{n_e} \sigma_i^e(\mathbf{v}|_{\overline{K^e}}) \mathbf{S}_B^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_i^r(\widehat{\mathbf{x}})) \right\| \leq C_{14},$$

where $\|\cdot\|$ is the Euclidean norm.

Let us note that Assumption 4.5 is replaced, in case of $m = 2$, by the minimum angle condition, see ZLÁMAL [219, p. 397], or by the maximum angle condition, see KŘÍŽEK AND NEITTAANMÄKI [115, p. 67], and, in case of $m = 3$, by either the minimum or maximum angle conditions between the edges as well as between the faces, see KŘÍŽEK AND NEITTAANMÄKI [115, p. 83]. For the used kind of elements we will show that Assumptions 4.6 and 4.7 follow from the angle conditions.

Lemma 4.4. Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$ be a domain and let $\{\Omega^h\}_{h>0}$ be a class of its nonempty polyhedral subdomains such that $\Omega^h \nearrow \Omega$. Let further $\mathbf{v} \in [C_0^\infty(\Omega)]^{\nu_1}$. Then, under Assumption 4.6, the following convergence holds

$$\left\| \boldsymbol{\pi}^h(\mathbf{v}|_{\overline{\Omega^h}}) - \boldsymbol{\pi}_0^h(\mathbf{v}|_{\overline{\Omega^h}}) \right\|_{\mathbf{B}, \Omega^h} = \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \boldsymbol{\xi}_i^h \right\|_{\mathbf{B}, \Omega^h} \rightarrow 0, \text{ as } h \rightarrow 0_+.$$

Proof. The proof bases on Theorem 3.6. Let us write the square of the norm

$$\begin{aligned} \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \boldsymbol{\xi}_i^h \right\|_{\mathbf{B}, \Omega^h}^2 &= \int_{\Omega} \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \boldsymbol{\xi}_i^h(\mathbf{x}) \right\|^2 d\mathbf{x} + \\ &+ \int_{\Omega} \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \mathbf{B}_x(\boldsymbol{\xi}_i^h(\mathbf{x})) \right\|^2 d\mathbf{x}. \end{aligned} \quad (4.40)$$

Due to (4.19), both $\boldsymbol{\xi}_i^h$ and $\mathbf{B}_x(\boldsymbol{\xi}_i^h)$ have small supports. We take an arbitrary $\mathbf{x} \in \Omega$. Since $\Omega^h \nearrow \Omega$ and $\forall K^e \in \mathcal{T}^h : h^e \leq h$, then due to (4.35) there exists $h_0 := \min_{\mathbf{y} \in \partial\Omega} \|\mathbf{x} - \mathbf{y}\|/2$ such that

$$\forall h \leq h_0 : \mathbf{x} \in \Omega^h \setminus \overline{\Delta^h \Omega^h}$$

holds. Clearly, $\text{supp } \boldsymbol{\xi}_i^h \subset \Delta^h \Omega^h$ and for the above $\mathbf{x} \in \Omega^h \setminus \overline{\Delta^h \Omega^h}$ we get

$$\forall h \leq h_0 : \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \boldsymbol{\xi}_i^h(\mathbf{x}) \right\|^2 = 0 \text{ and } \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \mathbf{B}_x(\boldsymbol{\xi}_i^h(\mathbf{x})) \right\|^2 = 0, \quad (4.41)$$

where $\|\cdot\|$ denotes the Euclidean norm.

Now, we bound the integrands in (4.40). Given a fixed $\mathbf{x} \in \Omega$, it is either the case of $\mathbf{x} \in \Omega \setminus \overline{\Delta^h \Omega^h}$, then the integrands vanish, or the case of $\mathbf{x} \equiv \mathcal{R}^e(\widehat{\mathbf{x}}) \in \overline{K^e} \subset \overline{\Delta^h \Omega^h}$, then by Assumption 4.6 the squares of the integrands are bounded as follows:

$$\begin{aligned} \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \boldsymbol{\xi}_i^h(\mathbf{x}) \right\| &= \left\| \sum_{j=1}^{n_e} \sigma_j^e(\mathbf{v}|_{\overline{K^e}}) \mathbf{S}^e \cdot \widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}}) \right\| \leq n_e C_{12} \max_{j=1, \dots, n_e} \max_{\widehat{\mathbf{x}} \in \overline{K^r}} \left\| \widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}}) \right\|, \\ \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \mathbf{B}_x(\boldsymbol{\xi}_i^h(\mathbf{x})) \right\| &= \left\| \sum_{j=1}^{n_e} \sigma_j^e(\mathbf{v}|_{\overline{K^e}}) \mathbf{S}_B^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}})) \right\| \leq \\ &\leq n_e C_{13} \max_{j=1, \dots, n_e} \max_{\widehat{\mathbf{x}} \in \overline{K^r}} \left\| \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}})) \right\|, \end{aligned}$$

therefore, the integrands themselves are also bounded. Having the boundness and by (4.41) having also the convergence of the integrands to zero in Ω , as $h \rightarrow 0_+$, we now apply Theorem 3.6 to both the integrals in (4.40), which yields

$$\int_{\Omega} \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \boldsymbol{\xi}_i^h(\mathbf{x}) \right\|^2 dx \rightarrow 0 \text{ and } \int_{\Omega} \left\| \sum_{i \in \mathcal{I}_0^h} \sigma_i^h(\mathbf{v}|_{\overline{\Omega^h}}) \mathbf{B}_x(\boldsymbol{\xi}_i^h(\mathbf{x})) \right\|^2 dx \rightarrow 0, \text{ as } h \rightarrow 0_+.$$

□

4.3.4 Convergence of the finite element method

The following theorem states the convergence property of the finite element method.

Theorem 4.2. *Let $\Omega \subset \mathbb{R}^m$, $\Omega \in \mathcal{L}$ be a domain, and let $\{\Omega^h\}_{h>0}$ be a class of its nonempty polyhedral subdomains such that $\Omega^h \nearrow \Omega$. Further, we assume that*

$$\max_{i,j} \left| d_{i,j}^h(\mathbf{x}) - d_{i,j}(\mathbf{x}) \right| \rightarrow 0 \text{ a.e. in } \Omega, \text{ as } h \rightarrow 0_+, \quad (4.42)$$

where $d_{i,j}^h(\mathbf{x}) := d_{i,j}(\mathbf{x})$ in $\Omega \setminus \overline{\Omega^h}$, and

$$\left\| \mathbf{f}^h - \mathbf{f} \right\|_{\nu_1, 0, \Omega} \rightarrow 0, \text{ as } h \rightarrow 0_+, \quad (4.43)$$

where $\mathbf{f}^h(\mathbf{x}) := \mathbf{f}(\mathbf{x})$ in $\Omega \setminus \overline{\Omega^h}$. Then for each $\varepsilon > 0$ the following convergence holds

$$\mathbf{X}_{\nu_1}^h \left(\mathbf{u}_{\varepsilon}^h(\Omega^h) \right) \rightarrow \mathbf{u}_{\varepsilon}(\Omega) \text{ in } \mathbf{H}_0(\mathbf{B}; \Omega), \text{ as } h \rightarrow 0_+,$$

where $\mathbf{X}_{\nu_1}^h : [L^2(\Omega^h)]^{\nu_1} \mapsto [L^2(\Omega)]^{\nu_1}$ is the linear extension operator defined by (4.33), $\mathbf{u}_{\varepsilon}^h(\Omega^h) \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ is the solution to $(W_{\varepsilon}^h(\Omega^h))$, and $\mathbf{u}_{\varepsilon}(\Omega) \in \mathbf{H}_0(\mathbf{B}; \Omega)$ is the solution to $(W_{\varepsilon}(\Omega))$.

Proof. Let $\varepsilon > 0$ be arbitrary. Let us denote

$$\mathbf{u}_\varepsilon := \mathbf{u}_\varepsilon(\Omega) \text{ and } \mathbf{u}_\varepsilon^h := \mathbf{X}_{\nu_1}^h \left(\mathbf{u}_\varepsilon^h(\Omega^h) \right). \quad (4.44)$$

From Lemma 4.3 we know that the set $\mathbf{X}_0(\mathbf{B}; \Omega; \Omega^h)^h$ is a closed subspace of $\mathbf{H}_0(\mathbf{B}; \Omega)$. Therefore, the function \mathbf{u}_ε^h is the Galerkin approximation to the solution \mathbf{u}_ε of $(W_\varepsilon(\Omega))$ in the space $\mathbf{X}_0(\mathbf{B}; \Omega; \Omega^h)^h$ and we can employ Lemma 4.1, which for any $\mathbf{v}^h \in \mathbf{X}_0(\mathbf{B}; \Omega; \Omega^h)^h$ yields

$$\begin{aligned} \left\| \mathbf{u}_\varepsilon - \mathbf{u}_\varepsilon^h \right\|_{\mathbf{B}, \Omega} \leq C_9(\varepsilon) \left\{ \left\| \mathbf{u}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} + \frac{|a_\varepsilon(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h) - a_\varepsilon^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h)|}{\left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega}} + \right. \\ \left. + \frac{|f(\mathbf{u}_\varepsilon^h - \mathbf{v}^h) - f^h(\mathbf{u}_\varepsilon^h - \mathbf{v}^h)|}{\left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega}} \right\}. \end{aligned} \quad (4.45)$$

Now the idea of the proof is like in KŘÍŽEK AND NEITTAANMÄKI [115, Th. 4.16], originally from DOKTOR [56]. Let $\tau > 0$ be arbitrary. By Assumption 3.3, there exists $\widetilde{\mathbf{u}}_\varepsilon \in [C_0^\infty(\Omega)]^{\nu_1}$ such that

$$\left\| \mathbf{u}_\varepsilon(\Omega) - \widetilde{\mathbf{u}}_\varepsilon \right\|_{\mathbf{B}, \Omega} \leq \frac{\tau}{6C_9(\varepsilon)}. \quad (4.46)$$

In the estimate (4.45) we choose

$$\mathbf{v}^h := \mathbf{X}_{\nu_1}^h \left(\boldsymbol{\pi}_0^h(\widetilde{\mathbf{u}}_\varepsilon|_{\Omega^h}) \right).$$

We estimate the first term on the right-hand side of (4.45). By the triangle inequality (3.1), by Lemma 4.3, and by (4.46), we get

$$\begin{aligned} \left\| \mathbf{u}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} &= \left\| \mathbf{u}_\varepsilon - \widetilde{\mathbf{u}}_\varepsilon + \widetilde{\mathbf{u}}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} \leq \left\| \mathbf{u}_\varepsilon - \widetilde{\mathbf{u}}_\varepsilon \right\|_{\mathbf{B}, \Omega} + \left\| \widetilde{\mathbf{u}}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} \leq \\ &\leq \frac{\tau}{6C_9(\varepsilon)} + \left\| \widetilde{\mathbf{u}}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega}. \end{aligned} \quad (4.47)$$

The second term on the right-hand side of (4.47) reads

$$\begin{aligned} \left\| \widetilde{\mathbf{u}}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega}^2 &= \int_{\Omega} \chi_{\Omega \setminus \Omega^h}(\mathbf{x}) \left(\left\| \widetilde{\mathbf{u}}_\varepsilon(\mathbf{x}) \right\|^2 + \left\| \mathbf{B}(\widetilde{\mathbf{u}}_\varepsilon(\mathbf{x})) \right\|^2 \right) d\mathbf{x} + \left\| \widetilde{\mathbf{u}}_\varepsilon - \boldsymbol{\pi}_0^h(\widetilde{\mathbf{u}}_\varepsilon|_{\Omega^h}) \right\|_{\mathbf{B}, \Omega^h}^2 = \\ &= \int_{\Omega} (1 - \chi_{\Omega^h}(\mathbf{x})) \left(\left\| \widetilde{\mathbf{u}}_\varepsilon(\mathbf{x}) \right\|^2 + \left\| \mathbf{B}(\widetilde{\mathbf{u}}_\varepsilon(\mathbf{x})) \right\|^2 \right) d\mathbf{x} + \left\| \widetilde{\mathbf{u}}_\varepsilon - \boldsymbol{\pi}_0^h(\widetilde{\mathbf{u}}_\varepsilon|_{\Omega^h}) \right\|_{\mathbf{B}, \Omega^h}^2, \end{aligned} \quad (4.48)$$

where $\|\cdot\|$ denotes the Euclidean norm. Since $\Omega^h \nearrow \Omega$, then (4.35) holds and we get

$$(1 - \chi_{\Omega^h}(\mathbf{x})) \left(\left\| \widetilde{\mathbf{u}}_\varepsilon(\mathbf{x}) \right\|^2 + \left\| \mathbf{B}(\widetilde{\mathbf{u}}_\varepsilon(\mathbf{x})) \right\|^2 \right) \rightarrow 0 \text{ a.e. in } \Omega, \text{ as } h \rightarrow 0_+.$$

Moreover, due to (4.36)

$$\forall h > 0 : (1 - \chi_{\Omega^h}(\mathbf{x})) \left(\left\| \widetilde{\mathbf{u}}_\varepsilon(\mathbf{x}) \right\|^2 + \left\| \mathbf{B}(\widetilde{\mathbf{u}}_\varepsilon(\mathbf{x})) \right\|^2 \right) \leq \left\| \widetilde{\mathbf{u}}_\varepsilon(\mathbf{x}) \right\|^2 + \left\| \mathbf{B}(\widetilde{\mathbf{u}}_\varepsilon(\mathbf{x})) \right\|^2 \text{ a.e. in } \Omega.$$

Therefore, Theorem 3.6 yields

$$\int_{\Omega} (1 - \chi_{\Omega^h}(\mathbf{x})) \left(\left\| \widetilde{\mathbf{u}}_\varepsilon(\mathbf{x}) \right\|^2 + \left\| \mathbf{B}(\widetilde{\mathbf{u}}_\varepsilon(\mathbf{x})) \right\|^2 \right) d\mathbf{x} \rightarrow 0, \text{ as } h \rightarrow 0_+.$$

Using the triangle inequality (3.1), the second term on the right-hand side of (4.48) is estimated as follows:

$$\begin{aligned} \left\| \widetilde{\mathbf{u}}_\varepsilon - \pi_0^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) \right\|_{\mathbf{B}, \Omega^h} &= \left\| \widetilde{\mathbf{u}}_\varepsilon - \pi^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) + \pi^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) - \pi_0^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) \right\|_{\mathbf{B}, \Omega^h} \leq \\ &\leq \left\| \widetilde{\mathbf{u}}_\varepsilon - \pi^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) \right\|_{\mathbf{B}, \Omega^h} + \left\| \pi^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) - \pi_0^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) \right\|_{\mathbf{B}, \Omega^h}. \end{aligned} \quad (4.49)$$

The definition (4.38), Assumption 4.4, and Assumption 4.5, respectively, yield

$$\begin{aligned} \left\| \widetilde{\mathbf{u}}_\varepsilon - \pi^h(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{\Omega^h}}) \right\|_{\mathbf{B}, \Omega^h}^2 &= \sum_{K^e \in \mathcal{T}^h} \left\| \widetilde{\mathbf{u}}_\varepsilon - \pi^e(\widetilde{\mathbf{u}}_\varepsilon|_{K^e}) \right\|_{\mathbf{B}, K^e}^2 \leq \\ &\leq \sum_{K^e \in \mathcal{T}^h} \left(C_{10}(K^e) h^e \|\widetilde{\mathbf{u}}_\varepsilon\|_{\nu_1, 2, K^e} \right)^2 \leq C_{11}^2 h^2 \|\widetilde{\mathbf{u}}_\varepsilon\|_{\nu_1, 2, \Omega}^2. \end{aligned}$$

By Lemma 4.4, the second term on the right-hand side of (4.49) tends toward zero. Therefore, the right-hand side of (4.48) tends to zero, as $h \rightarrow 0_+$, i.e., there exists $h_1 > 0$ such that

$$\forall h \leq h_1 : \left\| \mathbf{u}_\varepsilon - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} \leq \frac{\tau}{3C_9(\varepsilon)}. \quad (4.50)$$

Further, we estimate the second term on the right-hand side of (4.45). The nominator reads as follows:

$$\begin{aligned} \left| a_\varepsilon(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h) - a_\varepsilon^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h) \right| &= \\ &= \left| \int_{\Omega} \mathbf{B}(\mathbf{u}_\varepsilon^h - \mathbf{v}^h) \cdot \left((\mathbf{D} - \mathbf{D}^h)^T \cdot \mathbf{B}(\mathbf{v}^h) \right) dx \right| \leq \\ &\leq \left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega} \sqrt{\int_{\Omega} \max_{i,j} \left| d_{i,j}^h(\mathbf{x}) - d_{i,j}(\mathbf{x}) \right|^2 \|\mathbf{B}(\mathbf{v}^h(\mathbf{x}))\|^2 dx}, \end{aligned}$$

where we used the Cauchy–Schwarz inequality (3.3) in $[L^2(\Omega)]^{\nu_2}$. After dividing the latter by $\left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega}$, we get

$$\frac{\left| a_\varepsilon(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h) - a_\varepsilon^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h) \right|}{\left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\mathbf{B}, \Omega}} \leq \sqrt{\int_{\Omega} \max_{i,j} \left| d_{i,j}^h(\mathbf{x}) - d_{i,j}(\mathbf{x}) \right|^2 \|\mathbf{B}(\mathbf{v}^h(\mathbf{x}))\|^2 dx}. \quad (4.51)$$

Now, we use Theorem 3.6 to show that the integral on the right-hand side of (4.51) vanishes, as $h \rightarrow 0_+$. First, we prove the boundeness of the integrand. We take an arbitrary $\mathbf{x} \in \Omega$. If $\mathbf{x} \notin \Omega^h$, then $\mathbf{v}^h(\mathbf{x}) = 0$ and the integrand vanishes. If $\mathbf{x} \in \overline{K^e}$ for some $K^e \in \mathcal{T}^h$, $\overline{K^e} \subset \overline{\Delta^h \Omega^h}$, then by (4.3) and Assumption 4.6 the square of the integrand reads

$$\begin{aligned} \max_{i,j} \left| d_{i,j}^h(\mathbf{x}) - d_{i,j}(\mathbf{x}) \right| \left\| \mathbf{B}(\mathbf{v}^h(\mathbf{x})) \right\| &= \\ &= \max_{i,j} \left| d_{i,j}^h(\mathbf{x}) - d_{i,j}(\mathbf{x}) \right| \left\| \sum_{j: \mathcal{G}^e(j) \notin \mathcal{I}_0^h} \sigma_j^e(\widetilde{\mathbf{u}}_\varepsilon|_{\overline{K^e}}) \mathbf{S}_\mathbf{B}^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}})) \right\| \leq \\ &\leq 2 \max_{i,j} \|d_{i,j}\|_{L^\infty(\Omega)} n_e C_{13}(\widetilde{\mathbf{u}}_\varepsilon) \max_{j=1, \dots, n_e} \max_{\widehat{\mathbf{x}} \in \overline{K^r}} \left\| \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}})) \right\|. \end{aligned}$$

Otherwise, $\mathbf{x} \in \overline{K^e}$ for some $K^e \in \mathcal{T}^h$, $K^e \subset \Omega^h \setminus \overline{\Delta^h \Omega^h}$, and the square of the integrand reads

$$\begin{aligned} \max_{i,j} \left| d_{i,j}^h(\mathbf{x}) - d_{i,j}(\mathbf{x}) \right| \left\| \mathbf{B}(\mathbf{v}^h(\mathbf{x})) \right\| &= \\ &= \max_{i,j} \left| d_{i,j}^h(\mathbf{x}) - d_{i,j}(\mathbf{x}) \right| \left\| \sum_{j=1}^{n_e} \sigma_j^e(\widetilde{\mathbf{u}}_\varepsilon|_{K^e}) \mathbf{S}_B^e \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}})) \right\| \leq \\ &\leq 2 \max_{i,j} \|d_{i,j}\|_{L^\infty(\Omega)} C_{14}(\widetilde{\mathbf{u}}_\varepsilon), \end{aligned}$$

where we used (4.3) and Assumption 4.7. Therefore, the integrand on the right-hand side of (4.51) is bounded by a constant independent of h , and due to the assumption (4.42) it converges to zero almost everywhere in Ω . Then, by Theorem 3.6

$$\int_{\Omega} \max_{i,j} \left| d_{i,j}^h - d_{i,j} \right|^2 \left\| \mathbf{B}(\mathbf{v}^h) \right\|^2 d\mathbf{x} \rightarrow 0, \text{ as } h \rightarrow 0_+.$$

Hence, there exists $h_2 > 0$ such that

$$\forall h \leq h_2 : \frac{|a_\varepsilon(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h) - a_\varepsilon^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^h - \mathbf{v}^h)|}{\|\mathbf{u}_\varepsilon^h - \mathbf{v}^h\|_{\mathbf{B},\Omega}} \leq \frac{\tau}{3C_9(\varepsilon)}. \quad (4.52)$$

Finally, we estimate the third term on the right-hand side of (4.45). The nominator reads as follows:

$$\begin{aligned} \left| f(\mathbf{u}_\varepsilon^h - \mathbf{v}^h) - f^h(\mathbf{u}_\varepsilon^h - \mathbf{v}^h) \right| &= \left| \int_{\Omega} (\mathbf{f} - \mathbf{f}^h) \cdot (\mathbf{u}_\varepsilon^h - \mathbf{v}^h) d\mathbf{x} \right| \leq \\ &\leq \left\| \mathbf{f} - \mathbf{f}^h \right\|_{\nu_1,0,\Omega} \left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\nu_1,0,\Omega} \leq \\ &\leq \left\| \mathbf{f} - \mathbf{f}^h \right\|_{\nu_1,0,\Omega} \left\| \mathbf{u}_\varepsilon^h - \mathbf{v}^h \right\|_{\mathbf{B},\Omega}, \end{aligned} \quad (4.53)$$

where we used the Cauchy–Schwarz inequality in $[L^2(\Omega)]^{\nu_1}$. Dividing (4.53) by $\|\mathbf{u}_\varepsilon^h - \mathbf{v}^h\|_{\mathbf{B},\Omega}$ and using the assumption (4.43), it follows that there exists $h_3 > 0$ such that

$$\forall h \leq h_3 : \frac{|f(\mathbf{u}_\varepsilon^h - \mathbf{v}^h) - f^h(\mathbf{u}_\varepsilon^h - \mathbf{v}^h)|}{\|\mathbf{u}_\varepsilon^h - \mathbf{v}^h\|_{\mathbf{B},\Omega}} \leq \frac{\tau}{3C_9(\varepsilon)}. \quad (4.54)$$

At the end, combining (4.45), (4.50), (4.52), and (4.54), and recalling the notation (4.44), we have proven the statement, i.e., for any $\tau > 0$ there exists $h_0 := \min\{h_1, h_2, h_3\}$ such that

$$\forall h \leq h_0 : \left\| \mathbf{u}_\varepsilon(\Omega) - \mathbf{X}_{\nu_1}^h(\mathbf{u}_\varepsilon^h(\Omega^h)) \right\|_{\mathbf{B},\Omega} \leq \tau.$$

□

4.4 Finite elements for magnetostatics

In this section, we derive two basic types of finite elements which are used for solving the 2–dimensional or 3–dimensional magnetostatic problem, respectively. We will validate Assumptions 4.1–4.2 and Assumptions 4.4–4.7.

4.4.1 Linear Lagrange elements on triangles

These finite elements approximate the space $H^1(\Omega)$, where $\Omega \subset \mathbb{R}^2$, $\Omega \in \mathcal{L}$. They are used, e.g., for solving the 2-dimensional linear magnetostatic problem introduced in Section 3.4.2. The elements are characterized by triangular domains and by the degrees of freedom that are nodal values in the corners.

The *linear Lagrange element* is a triple $E := (\overline{K^e}, P^e, \Sigma^e)$, where $K^e \subset \mathbb{R}^2$ is a triangular domain,

$$P^e := \{ p(\mathbf{x}) := a_0^e + a_1^e x_1 + a_2^e x_2 \in C(\overline{K^e}) \mid a_0^e, a_1^e, a_2^e \in \mathbb{R}, \mathbf{x} := (x_1, x_2) \in \overline{K^e} \},$$

and the degrees of freedom are

$$\Sigma^e := \{ \sigma_1^e, \sigma_2^e, \sigma_3^e \},$$

where $\sigma_i^e : C(\overline{K^e}) \mapsto \mathbb{R}$ is such that for $v \in C(\overline{K^e})$

$$\sigma_i^e(v) := v(\mathbf{x}_i^e), \quad i = 1, 2, 3, \quad (4.55)$$

where $\mathbf{x}_1^e := (x_{1,1}^e, x_{1,2}^e)$, $\mathbf{x}_2^e := (x_{2,1}^e, x_{2,2}^e)$, $\mathbf{x}_3^e := (x_{3,1}^e, x_{3,2}^e)$ are the corners of $\overline{K^e}$.

We concern the space $H^1(K^e)$ with the trace operator $\gamma_{K^e}(v) := v|_{\partial K^e}$, $v \in P^e$. From (4.55) it is easy to see that the following couples of degrees of freedom (σ_1^e, σ_2^e) , (σ_2^e, σ_3^e) , and (σ_3^e, σ_1^e) for any $v \in P^e$ determine the traces $v|_{h_3^e}$, $v|_{h_2^e}$, and $v|_{h_1^e}$ along the edges $h_3^e := (\mathbf{x}_1^e, \mathbf{x}_2^e)$, $h_1^e := (\mathbf{x}_2^e, \mathbf{x}_3^e)$, and $h_2^e := (\mathbf{x}_3^e, \mathbf{x}_1^e)$, respectively, see also Fig. 4.4. Therefore, Assumption 4.1 is fulfilled and we say that the linear Lagrange elements are $H^1(K^e)$ -conforming.

According to (4.27), (4.28), and Fig. 4.2, we specify the transformation from the reference element to the element e by

$$\mathbf{R}^e := \begin{pmatrix} x_{2,1}^e - x_{1,1}^e & x_{3,1}^e - x_{1,1}^e \\ x_{2,2}^e - x_{1,2}^e & x_{3,2}^e - x_{1,2}^e \end{pmatrix}, \quad \mathbf{r}^e := \begin{pmatrix} x_{1,1}^e \\ x_{1,2}^e \end{pmatrix}, \quad (4.56)$$

where the corners of the reference triangle $\overline{K^r}$ are

$$\widehat{\mathbf{x}}_1^r := (0, 0), \quad \widehat{\mathbf{x}}_2^r := (1, 0), \quad \widehat{\mathbf{x}}_3^r := (0, 1). \quad (4.57)$$

Concerning Assumption 4.2, we specify \mathcal{S}^e by

$$\mathcal{S}^e := 1. \quad (4.58)$$

It is easy to see that

$$\mathbf{grad}_{\mathbf{x}}(v(\mathbf{x})) = (\mathbf{R}^e)^{-T} \cdot \mathbf{grad}_{\widehat{\mathbf{x}}}(\widehat{v}(\widehat{\mathbf{x}})), \quad (4.59)$$

where $v(\mathbf{x}) := \mathcal{S}^e \cdot \widehat{v}(\widehat{\mathbf{x}})$ and $\mathbf{x} := \mathbf{R}^e \cdot \widehat{\mathbf{x}} + \mathbf{r}^e$. The reference shape functions read as follows:

$$\widehat{\xi}_1^r(\widehat{\mathbf{x}}) := 1 - \widehat{x}_1 - \widehat{x}_2, \quad \widehat{\xi}_2^r(\widehat{\mathbf{x}}) := \widehat{x}_1, \quad \widehat{\xi}_3^r(\widehat{\mathbf{x}}) := \widehat{x}_2, \quad \text{where } \widehat{\mathbf{x}} := (\widehat{x}_1, \widehat{x}_2) \in \overline{K^r} \quad (4.60)$$

and where $\overline{K^r}$ is the triangle in Fig. 4.2 the corners of which are given by (4.57).

Now we will state the element approximation property such that both Assumption 4.4 and Assumption 4.5 will be fulfilled. Suppose that we have a discretization $\mathcal{T}^h := \{K^{e_1}, \dots, K^{e_n}\}$ – a *triangulation*. The following definition is due to ZLÁMAL [219, p. 397].

Definition 4.1. A family $\mathcal{F} := \{\mathcal{T}^h \mid h > 0\}$ of triangulations is said to satisfy the minimum angle condition if there exists a constant γ_0 such that for any $\mathcal{T}^h \in \mathcal{F}$ and any $K^e \in \mathcal{T}^h$ we have

$$0 < \gamma_0 \leq \gamma_{K^e} < \frac{\pi}{2}, \quad (4.61)$$

where γ_{K^e} is the minimum angle of the triangle K^e .

The next lemma is due to ZLÁMAL [219] and it replaces both Assumption 4.4 and Assumption 4.5.

Lemma 4.5. Let \mathcal{F} be a family of triangulations satisfying the minimum angle condition (4.61). Then there exists a constant $C_{11} > 0$ such that for any $\mathcal{T}^h \in \mathcal{F}$ with $h \leq \bar{h}$ we have

$$\forall v \in H^2(\Omega^h) : \left\| v - \pi^h(v) \right\|_{1,\Omega^h} \leq C_{11} h |v|_{2,\Omega^h},$$

where $\pi^h : C(\overline{\Omega^h}) \mapsto H^1(\Omega^h)^h$ is defined by (4.38), using the degrees of freedom (4.55).

Proof. See ZLÁMAL [219]. □

The next two lemmas fulfill Assumptions 4.6 and 4.7, respectively.

Lemma 4.6. Let $v \in C_0^\infty(\Omega)$. Then there exist positive constants $C_{12} \equiv C_{12}(v)$ and $C_{13} \equiv C_{13}(v)$ such that for any discretization parameter $h > 0$ satisfying (4.32), for any subdomain $\Omega^h \subset \Omega$ satisfying Assumption 4.3, and for any discretization \mathcal{T}^h which satisfies the minimum angle condition (4.61) the following holds

$$\forall K^e \in \mathcal{T}^h : K^e \subset \Delta^h \Omega^h \Rightarrow |\sigma_j^e(v|_{\overline{K^e}})| \|\mathbf{S}^e\| \leq C_{12} \text{ and } |\sigma_j^e(v|_{\overline{K^e}})| \|\mathbf{S}_{\text{grad}}^e\| \leq C_{13},$$

where $j = 1, 2, 3$ and where $\Delta^h \Omega^h$ is defined by (4.39).

Proof. Let $v \in C_0^\infty(\Omega)$ be an arbitrary function, $h > 0$ be a discretization parameter satisfying (4.32), $\Omega^h \subset \Omega$ be a polygonal subdomain satisfying Assumption 4.3, and \mathcal{T}^h be a discretization which satisfies the minimum angle condition (4.61).

Let $K^e \subset \overline{\Omega^h}$ be arbitrary and let $\mathbf{x} \in \overline{K^e}$. Since $\mathbf{S}^e = 1$, the first estimate is as follows:

$$|\sigma_j^e(v|_{\overline{K^e}})| \|\mathbf{S}^e\| = |v(\mathbf{x}_j^e)| \leq \max_{\mathbf{z} \in \Omega} |v(\mathbf{z})|, \quad j = 1, 2, 3,$$

where $\mathbf{x}_j^e \in \partial K^e$ is a corner of K^e . Hence,

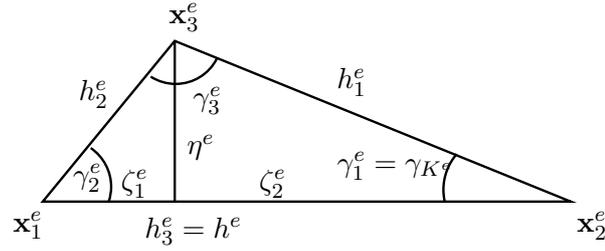
$$C_{12} := \max_{\mathbf{z} \in \Omega} |v(\mathbf{z})|.$$

Let now $K^e \in \mathcal{T}^h$ be such that $K^e \subset \Delta^h \Omega^h$ and let $\mathbf{x} \in \overline{K^e}$. Then, by Assumption 4.3, there exist $\mathbf{y} \in \partial\Omega$ and $\mathbf{x}^h \in \partial\Omega^h$ such that

$$\|\mathbf{x} - \mathbf{y}\| = \left\| \mathbf{x} - \mathbf{x}^h + \mathbf{x}^h - \mathbf{y} \right\| \leq \left\| \mathbf{x} - \mathbf{x}^h \right\| + \left\| \mathbf{x}^h - \mathbf{y} \right\| \leq 2h^e,$$

where h^e is by definition the maximum side of K^e . Since $\mathbf{y} \in \partial\Omega$, then $v(\mathbf{y}) = 0$. Now, we use Theorem 3.3

$$v(\mathbf{x}) = v(\mathbf{y}) + \mathbf{grad}(v(\mathbf{z})) \cdot (\mathbf{x} - \mathbf{z}) = \mathbf{grad}(v(\mathbf{z})) \cdot (\mathbf{x} - \mathbf{z}), \quad (4.62)$$

Figure 4.4: A Lagrange triangle K^e

where $\mathbf{z} \in \overline{K^e}$ lies on the line between \mathbf{x} and \mathbf{y} . Therefore,

$$\forall \mathbf{x} \in \overline{K^e} : |v(\mathbf{x})| \leq \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \cdot 2h^e. \quad (4.63)$$

To prove the second estimate, we exploit the structure of the matrix $(\mathbf{R}^e)^{-T}$. Using (3.9) and (3.14), we get

$$\begin{aligned} \left\| (\mathbf{R}^e)^{-T} \right\| &= \max_{i,j} \left| \left((\mathbf{R}^e)^{-T} \right)_{i,j} \right| = \max_{i,j} \left| \left((\mathbf{R}^e)^{-1} \right)_{i,j} \right| = \max_{i,j} \left| \frac{\mathbf{x}_{j+1,i}^e - \mathbf{x}_{1,i}^e}{\det(\mathbf{R}^e)} \right| \leq \\ &\leq \frac{h^e}{2 \text{meas}(K^e)}. \end{aligned} \quad (4.64)$$

From Fig. 4.4, it is clear that

$$\text{meas}(K^e) = \frac{h^e \eta^e}{2}.$$

Using (4.63), (4.64), Fig. 4.4, and the minimum angle condition (4.61), the second estimate reads as follows:

$$\begin{aligned} |\sigma_j^e(v|_{\overline{K^e}})| \|\mathbf{S}_B^e\| &= |v(\mathbf{x}_j^e)| \left\| (\mathbf{R}^e)^{-T} \right\| \leq \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{2h^e}{\eta^e} \leq \\ &\leq \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{2(\zeta_1^e + \zeta_2^e)}{\eta^e} \leq \\ &\leq \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{4\zeta_2^e}{\eta^e} = 4 \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{1}{\tan(\gamma_{K^e})} \leq \\ &\leq 4 \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{1}{\tan(\gamma_0)}, \end{aligned}$$

where γ_{K^e} denotes the minimum angle of the triangle K^e . Hence,

$$C_{13} := \frac{4 \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\|}{\tan(\gamma_0)}.$$

□

Lemma 4.7. *Let $v \in C_0^\infty(\Omega)$. Then there exists a positive constant $C_{14} \equiv C_{14}(v)$ such that for any discretization parameter $h > 0$, for any subdomain $\Omega^h \subset \Omega$, and for any discretization \mathcal{T}^h which satisfies the minimum angle condition (4.61), the following holds*

$$\begin{aligned} \forall K^e \in \mathcal{T}^h \quad \forall \mathbf{x} \equiv \mathcal{R}^e(\hat{\mathbf{x}}) \in \overline{K^e} : \\ \left\| \mathbf{grad}_{\mathbf{x}}(\pi^e(v|_{\overline{K^e}}(\mathbf{x}))) \right\| = \left\| \sum_{i=1}^3 \sigma_i^e(v|_{\overline{K^e}}) \mathbf{S}_{\mathbf{grad}}^e \cdot \mathbf{grad}_{\hat{\mathbf{x}}}(\hat{\xi}_i^e(\hat{\mathbf{x}})) \right\| \leq C_{14}, \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm.

Proof. Let $v \in C_0^\infty(\Omega)$ be an arbitrary function, $h > 0$ be a discretization parameter, $\Omega^h \subset \Omega$ be a polygonal subdomain, \mathcal{T}^h be a discretization of Ω^h which satisfies the minimum angle condition (4.61), and let $K^e \in \mathcal{T}^h$ be an element domain. The gradients of the reference shape functions, see (4.60), are constant over $\overline{K^r}$

$$\mathbf{grad}_{\widehat{\mathbf{x}}}\left(\widehat{\xi}_1^r(\widehat{\mathbf{x}})\right) = (-1, -1), \quad \mathbf{grad}_{\widehat{\mathbf{x}}}\left(\widehat{\xi}_2^r(\widehat{\mathbf{x}})\right) = (1, 0), \quad \mathbf{grad}_{\widehat{\mathbf{x}}}\left(\widehat{\xi}_3^r(\widehat{\mathbf{x}})\right) = (0, 1),$$

where $\widehat{\mathbf{x}} := (\widehat{x}_1, \widehat{x}_2) \in \overline{K^r}$. Now, using the latter and the definition of \mathbf{R}^e , we exploit the structure of the matrix $\mathbf{S}_{\mathbf{grad}}^e \equiv (\mathbf{R}^e)^{-T}$. It holds that

$$\begin{aligned} & \left\| \sum_{i=1}^3 \sigma_i^e(v|_{\overline{K^e}}) \mathbf{S}_{\mathbf{grad}}^e \cdot \mathbf{grad}_{\widehat{\mathbf{x}}}\left(\widehat{\xi}_i^r(\widehat{\mathbf{x}})\right) \right\| \leq \\ & \leq \frac{\sqrt{2}}{|\det(\mathbf{R}^e)|} \left\| \begin{pmatrix} (x_{3,2}^e - x_{1,2}^e)(v(\mathbf{x}_2^e) - v(\mathbf{x}_1^e)) + (x_{2,2}^e - x_{1,2}^e)(v(\mathbf{x}_1^e) - v(\mathbf{x}_3^e)) \\ (x_{3,1}^e - x_{1,1}^e)(v(\mathbf{x}_1^e) - v(\mathbf{x}_2^e)) + (x_{2,1}^e - x_{1,1}^e)(v(\mathbf{x}_3^e) - v(\mathbf{x}_1^e)) \end{pmatrix} \right\|, \end{aligned} \quad (4.65)$$

where we also used (3.14). Since h^e is the maximum side, it follows that

$$|x_{i,j}^e - x_{1,j}^e| \leq h^e, \quad i = 1, 2, 3, \quad j = 1, 2.$$

Similarly as in (4.62), Theorem 3.3 yields

$$|v(\mathbf{x}_i^e) - v(\mathbf{x}_1^e)| \leq \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| h^e, \quad i = 2, 3.$$

Finally, like at the end of the previous proof, from Fig. 4.4 it is clear that

$$|\det(\mathbf{R}^e)| = 2 \operatorname{meas}(K^e) = h^e \eta^e$$

and, due to the minimum angle condition (4.61) and Fig. 4.4, the estimate (4.65) is as follows:

$$\begin{aligned} & \left\| \sum_{i=1}^3 \sigma_i^e(v|_{\overline{K^e}}) \mathbf{S}_{\mathbf{grad}}^e \cdot \mathbf{grad}_{\widehat{\mathbf{x}}}\left(\widehat{\xi}_i^r(\widehat{\mathbf{x}})\right) \right\| \leq \frac{\sqrt{2}}{h^e \eta^e} 2h^e \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| h^e \leq \\ & \leq 2\sqrt{2} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{\zeta_1^e + \zeta_2^e}{\eta^e} \leq 2\sqrt{2} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{2\zeta_2^e}{\eta^e} = \\ & = 4\sqrt{2} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{1}{\tan(\gamma_{K^e})} \leq 4\sqrt{2} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{1}{\tan(\gamma_0)}, \end{aligned}$$

where γ_{K^e} is the minimum angle of the triangle K^e . Hence,

$$C_{14} := 4\sqrt{2} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v(\mathbf{z}))\| \frac{1}{\tan(\gamma_0)}.$$

□

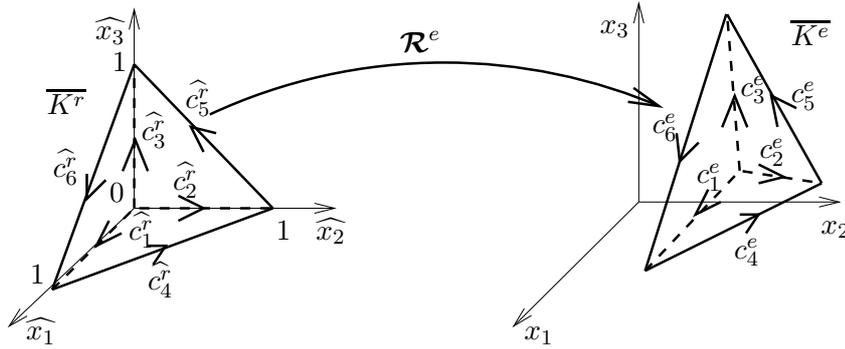


Figure 4.5: A transformation from the reference Nédélec tetrahedron

4.4.2 Linear Nédélec elements on tetrahedra

Here, we state a type of finite elements, which is frequently used for the approximation of the space $\mathbf{H}(\mathbf{curl}; \Omega)$, where $\Omega \subset \mathbb{R}^3$, $\Omega \in \mathcal{L}$. These elements can be used for solving the 3-dimensional linear magnetostatic problem, which was introduced in Section 3.4.3. The elements are defined over tetrahedra and the degrees of freedom are calculated as integrals along the edges. The elements were first introduced by NÉDÉLEC [142] and, since then, they have become a standard.

The *linear Nédélec element* is a triple $E := (\overline{K}^e, \mathbf{P}^e, \Sigma^e)$, where $K^e \subset \mathbb{R}^3$ is a tetrahedral domain,

$$\mathbf{P}^e := \{ \mathbf{p}(\mathbf{x}) := \mathbf{a}^e \times \mathbf{x} + \mathbf{b}^e \mid \mathbf{a}^e, \mathbf{b}^e \in \mathbb{R}^3, \mathbf{x} := (x_1, x_2, x_3) \in \overline{K}^e \},$$

and the degrees of freedom are

$$\Sigma^e := \{ \sigma_1^e, \sigma_2^e, \sigma_3^e, \sigma_4^e, \sigma_5^e, \sigma_6^e \},$$

where $\sigma_i^e : [C(\overline{K}^e)]^3 \mapsto \mathbb{R}$ is such that for $\mathbf{v} \in [C(\overline{K}^e)]^3$

$$\sigma_i^e(\mathbf{v}) := \int_{c_i^e} \mathbf{v} \cdot \mathbf{t}_i^e ds, \quad i = 1, \dots, 6,$$

where c_i^e stand for the oriented edges, see Fig. 4.5, and \mathbf{t}_i^e are the related unit tangential vectors.

Now, we concern the space $\mathbf{H}(\mathbf{curl}; K^e)$ and the corresponding trace operator $\gamma_{K^e}(\mathbf{v}) := \mathbf{n}^e \times \mathbf{v}$ on ∂K^e , where $\mathbf{v} \in \mathbf{P}^e$ and \mathbf{n}^e denotes the unit outer normal vector to ∂K^e . By NÉDÉLEC [142, Theorem 1], Assumption 4.1 is fulfilled, thus, the linear Nédélec finite elements are $\mathbf{H}(\mathbf{curl}; K^e)$ -conforming.

The transformation \mathcal{R}^e in Fig. 4.5 is determined by

$$\mathbf{R}^e := \begin{pmatrix} x_{2,1}^e - x_{1,1}^e & x_{3,1}^e - x_{1,1}^e & x_{4,1}^e - x_{1,1}^e \\ x_{2,2}^e - x_{1,2}^e & x_{3,2}^e - x_{1,2}^e & x_{4,2}^e - x_{1,2}^e \\ x_{2,3}^e - x_{1,3}^e & x_{3,3}^e - x_{1,3}^e & x_{4,3}^e - x_{1,3}^e \end{pmatrix}, \quad \mathbf{r}^e := \begin{pmatrix} x_{1,1}^e \\ x_{1,2}^e \\ x_{1,3}^e \end{pmatrix}, \quad (4.66)$$

where $\mathbf{x}_i^e := (x_{i,1}^e, x_{i,2}^e, x_{i,3}^e)$, $i=1, \dots, 4$, are the corners of the tetrahedron \overline{K}^e , which correspond to the following corners of \overline{K}^r

$$\widehat{\mathbf{x}}_1^r := (0, 0, 0), \quad \widehat{\mathbf{x}}_2^r := (1, 0, 0), \quad \widehat{\mathbf{x}}_3^r := (0, 1, 0), \quad \widehat{\mathbf{x}}_4^r := (0, 0, 1). \quad (4.67)$$

As far as Assumption 4.2 is considered, we determine \mathbf{S}^e by

$$\mathbf{S}^e := (\mathbf{R}^e)^{-T}. \quad (4.68)$$

It can be shown that following Piola's transformation holds, see RAVIART AND THOMAS [167, Formula 3.17],

$$\mathbf{curl}_{\mathbf{x}}(\mathbf{v}(\mathbf{x})) = \frac{1}{\det(\mathbf{R}^e)} \mathbf{R}^e \cdot \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\mathbf{v}}(\widehat{\mathbf{x}})), \quad (4.69)$$

where $\mathbf{v}(\mathbf{x}) := \mathbf{S}^e \cdot \widehat{\mathbf{v}}(\widehat{\mathbf{x}})$ and $\mathbf{x} := \mathbf{R}^e \cdot \widehat{\mathbf{x}} + \mathbf{r}^e$. The reference shape functions read as follows:

$$\begin{aligned} \widehat{\xi}_1^r(\widehat{\mathbf{x}}) &:= \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \times \widehat{\mathbf{x}} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, & \widehat{\xi}_2^r(\widehat{\mathbf{x}}) &:= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \times \widehat{\mathbf{x}} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \\ \widehat{\xi}_3^r(\widehat{\mathbf{x}}) &:= \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \times \widehat{\mathbf{x}} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, & \widehat{\xi}_4^r(\widehat{\mathbf{x}}) &:= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \times \widehat{\mathbf{x}} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \\ \widehat{\xi}_5^r(\widehat{\mathbf{x}}) &:= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \times \widehat{\mathbf{x}} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, & \widehat{\xi}_6^r(\widehat{\mathbf{x}}) &:= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \times \widehat{\mathbf{x}} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \end{aligned} \quad (4.70)$$

where $\widehat{\mathbf{x}} := (\widehat{x}_1, \widehat{x}_2, \widehat{x}_3) \in \overline{K^r}$ and $\overline{K^r}$ is the reference tetrahedron, see Fig. 4.5, the corners of which are given by (4.67).

Now, we will state the element approximation property such that both Assumption 4.4 and Assumption 4.5 will be fulfilled. Suppose that we have a decomposition $\mathcal{T}^h := \{K^{e_1}, \dots, K^{e_{n_{\Omega^h}}}\}$. The following definition and lemma are due to NÉDÉLEC [142, p. 327].

Definition 4.2. A family $\mathcal{F} := \{\mathcal{T}^h \mid h > 0\}$ of decompositions into tetrahedra is said to be regular if there exists a constant $C_{15} > 0$ such that for any $\mathcal{T}^h \in \mathcal{F}$ and any $K^e \in \mathcal{T}^h$ we have

$$\frac{h^e}{\rho^e} \leq C_{15}, \quad (4.71)$$

where ρ^e denotes the radius of the largest sphere inscribed in K^e .

Lemma 4.8. Let \mathcal{F} be a regular family of decompositions into tetrahedra in the sense of Definition 4.2. Then there exists a constant $C_{11} > 0$ such that for any $\mathcal{T}^h \in \mathcal{F}$ with $h \leq \bar{h}$ we have

$$\forall \mathbf{v} \in \left[H^2(\Omega^h) \right]^{\nu_1} : \|\mathbf{v} - \boldsymbol{\pi}^e(\mathbf{v})\|_{\mathbf{curl}, \Omega^h} \leq C_{11} h |\mathbf{v}|_{\nu_1, 2, \Omega^h}.$$

Proof. The assertion is a direct consequence of NÉDÉLEC [142, Th. 2]. \square

The next two lemmas fulfill Assumptions 4.6 and 4.7, respectively.

Lemma 4.9. Let $\mathbf{v} \in [C_0^\infty(\Omega)]^3$. Then there exist positive constants $C_{12} \equiv C_{12}(\mathbf{v})$ and $C_{13} \equiv C_{13}(\mathbf{v})$ such that for any discretization parameter $h > 0$, for any subdomain $\Omega^h \subset \Omega$ satisfying Assumption 4.3, and for any discretization \mathcal{T}^h which satisfies the regularity condition (4.71) the following holds

$$\forall K^e \in \mathcal{T}^h : \overline{K^e} \subset \overline{\Delta^h \Omega^h} \Rightarrow |\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| \|\mathbf{S}^e\| \leq C_{12} \text{ and } |\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| \|\mathbf{S}_{\mathbf{curl}}^e\| \leq C_{13},$$

where $j = 1, \dots, 6$.

Proof. Let $\mathbf{v} := (v_1, \dots, v_{\nu_1}) \in [C_0^\infty(\Omega)]^{\nu_1}$ be an arbitrary function, $h > 0$ be a discretization parameter satisfying (4.32), $\Omega^h \subset \Omega$ be a polygonal subdomain satisfying Assumption 4.3, \mathcal{T}^h be a discretization which satisfies the regularity condition (4.71), and let $K^e \in \mathcal{T}^h$ be an element domain. For $j = 1, \dots, 6$ we have the estimate

$$|\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| = \left| \int_{c_j^e} \mathbf{v}(\mathbf{x}) \cdot \mathbf{t}_j^e ds \right| \leq \max_{\mathbf{z} \in \Omega} \|\mathbf{v}(\mathbf{z})\| h^e, \quad (4.72)$$

where h^e stands for the maximum edge size. Since $\mathbf{S}^e := (\mathbf{R}^e)^{-T}$, then, using (3.9) and (3.14), it follows that

$$\|\mathbf{S}^e\| = \frac{1}{|\det(\mathbf{R}^e)|} \|\widetilde{\mathbf{R}}^e\| = \frac{1}{6 \text{meas}(K^e)} \|\widetilde{\mathbf{R}}^e\| \leq \frac{(h^e)^2}{3 \text{meas}(K^e)}. \quad (4.73)$$

Since ρ^e denotes the radius of the largest sphere inscribed in K^e , from the regularity condition (4.71) it is obvious that

$$\text{meas}(K^e) \geq \frac{4}{3}\pi(\rho^e)^3 \geq \frac{4}{3}\pi \left(\frac{h^e}{C_{15}} \right)^3. \quad (4.74)$$

Putting the latter into (4.73) and combining that with (4.72), the first estimate reads as follows:

$$|\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| \|\mathbf{S}^e\| \leq \max_{\mathbf{z} \in \Omega} \|\mathbf{v}(\mathbf{z})\| \frac{(C_{15})^3}{4\pi},$$

hence,

$$C_{12} := \max_{\mathbf{z} \in \Omega} \|\mathbf{v}(\mathbf{z})\| \frac{(C_{15})^3}{4\pi}.$$

Similarly as in the proof of Lemma 4.6, let $K^e \in \mathcal{T}^h$ be such that $K^e \subset \Delta^h \Omega^h$. Then there exists $\mathbf{x}^h \in \partial\Omega^h$ and, by Assumption 4.3, there exists $\mathbf{y} \in \partial\Omega$ such that

$$\|\mathbf{x} - \mathbf{y}\| = \left\| \mathbf{x} - \mathbf{x}^h + \mathbf{x}^h - \mathbf{y} \right\| \leq \left\| \mathbf{x} - \mathbf{x}^h \right\| + \left\| \mathbf{x}^h - \mathbf{y} \right\| \leq 2h^e,$$

where h^e is by definition the maximum side of K^e . Since $\mathbf{y} \in \partial\Omega$, then $\mathbf{v}(\mathbf{y}) = \mathbf{0}$. Now we use Theorem 3.3

$$v_i(\mathbf{x}) = v_i(\mathbf{y}) + \mathbf{grad}(v_i(\mathbf{z})) \cdot (\mathbf{x} - \mathbf{z}) = \mathbf{grad}(v_i(\mathbf{z})) \cdot (\mathbf{x} - \mathbf{z}) \quad \text{for } i = 1, 2, 3$$

and

$$\|\mathbf{v}(\mathbf{x})\| \leq \max_{i \in \{1,2,3\}} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v_i(\mathbf{z}))\| 2h^e \quad \text{for } i = 1, 2, 3,$$

where $\mathbf{z} \in \overline{K^e}$ lies on the line between \mathbf{x} and \mathbf{y} . Therefore,

$$|\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| = \left| \int_{c_j^e} \mathbf{v}(\mathbf{x}) \cdot \mathbf{t}_j^e ds \right| \leq \max_{i \in \{1,2,3\}} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v_i(\mathbf{z}))\| 2(h^e)^2. \quad (4.75)$$

Concerning the second estimate, we have

$$\|\mathbf{S}^e_{\text{curl}}\| = \frac{1}{|\det(\mathbf{R}^e)|} \|\mathbf{R}^e\| \leq \frac{\max_{i,j} |x_{i+1,j}^e - x_{1,j}^e|}{6 \text{meas}(K^e)} \leq \frac{h^e}{6 \text{meas}(K^e)} \leq \frac{(C_{15})^3}{8\pi (h^e)^2},$$

where we used (4.74). Combining the latter with (4.75), the second estimate is as follows:

$$\begin{aligned} |\sigma_j^e(\mathbf{v}|_{\overline{K^e}})| \|\mathbf{S}_{\text{curl}}^e\| &\leq \max_{i \in \{1,2,3\}} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v_i(\mathbf{z}))\| 2(h^e)^2 \frac{(C_{15})^3}{8\pi(h^e)^2} \leq \\ &\leq \frac{1}{4} \max_{i \in \{1,2,3\}} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v_i(\mathbf{z}))\| (C_{15})^3, \end{aligned} \quad (4.76)$$

hence,

$$C_{13} := \frac{1}{4} \max_{i \in \{1,2,3\}} \max_{\mathbf{z} \in \Omega} \|\mathbf{grad}(v_i(\mathbf{z}))\| (C_{15})^3.$$

□

Lemma 4.10. *Let $\mathbf{v} \in [C_0^\infty(\Omega)]^3$. Then there exists a positive constant $C_{14} \equiv C_{14}(\mathbf{v})$ such that for any discretization parameter $h > 0$, for any subdomain $\Omega^h \subset \Omega$, and for any discretization \mathcal{T}^h which satisfy the regularity condition (4.71), the following holds*

$$\begin{aligned} \forall K^e \in \mathcal{T}^h \quad \forall \mathbf{x} \equiv \mathcal{R}^e(\widehat{\mathbf{x}}) \in \overline{K^e} : \\ \|\mathbf{curl}_{\mathbf{x}}(\boldsymbol{\pi}^e(\mathbf{v}|_{\overline{K^e}}))\| = \left\| \sum_{i=1}^6 \sigma_i^e(\mathbf{v}|_{\overline{K^e}}) \mathbf{S}_{\text{curl}}^e \cdot \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_i^r(\widehat{\mathbf{x}})) \right\| \leq C_{14}. \end{aligned} \quad (4.77)$$

Proof. The proof is similar to that of Lemma 4.7. Let $\mathbf{v} \in [C_0^\infty(\Omega)]^3$ be an arbitrary function, $h > 0$ be a discretization parameter, $\Omega^h \subset \Omega$ be a polygonal subdomain, \mathcal{T}^h be a discretization of Ω^h which satisfies the regularity condition (4.71), and let $K^e \in \mathcal{T}^h$ be an element domain. The rotations of the reference shape functions, see (4.70), are constant over $\overline{K^r}$

$$\begin{aligned} \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_1^r(\widehat{\mathbf{x}})) &= (0, -2, 2), & \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_2^r(\widehat{\mathbf{x}})) &= (2, 0, -2), & \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_3^r(\widehat{\mathbf{x}})) &= (-2, 2, 0), \\ \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_4^r(\widehat{\mathbf{x}})) &= (0, 0, 2), & \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_5^r(\widehat{\mathbf{x}})) &= (2, 0, 0), & \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_6^r(\widehat{\mathbf{x}})) &= (0, 2, 0), \end{aligned}$$

where $\widehat{\mathbf{x}} := (\widehat{x}_1, \widehat{x}_2, \widehat{x}_3) \in \overline{K^r}$. Let us simplify the rest of the proof by the following notation

$$\sigma_i^e := \sigma_i^e(\mathbf{v}|_{\overline{K^e}}) \quad \text{for } i = 1, 2, \dots, 6.$$

Now, we exploit the structure of the matrix $\mathbf{S}_{\text{curl}}^e$. It holds that

$$\begin{aligned} \mathbf{curl}_{\mathbf{x}}(\boldsymbol{\pi}^e(\mathbf{v}|_{\overline{K^e}(\mathbf{x})})) &= \frac{1}{\det(\mathbf{R}^e)} \sum_{i=1}^6 \sigma_i^e(\mathbf{v}|_{\overline{K^e}}) \mathbf{R}^e \cdot \mathbf{curl}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_i^r(\widehat{\mathbf{x}})) = \\ &= \frac{2}{6 \text{meas}(K^e)} \left(\begin{aligned} &(x_{2,1}^e - x_{1,1}^e)(\sigma_2^e - \sigma_3^e + \sigma_5^e) + \\ &(x_{2,2}^e - x_{1,2}^e)(\sigma_2^e - \sigma_3^e + \sigma_5^e) + \\ &(x_{2,3}^e - x_{1,3}^e)(\sigma_2^e - \sigma_3^e + \sigma_5^e) + \\ &+ (x_{3,1}^e - x_{1,1}^e)(\sigma_3^e - \sigma_1^e + \sigma_6^e) + (x_{4,1}^e - x_{1,1}^e)(\sigma_1^e - \sigma_2^e + \sigma_4^e) \\ &+ (x_{3,2}^e - x_{1,2}^e)(\sigma_3^e - \sigma_1^e + \sigma_6^e) + (x_{4,2}^e - x_{1,2}^e)(\sigma_1^e - \sigma_2^e + \sigma_4^e) \\ &+ (x_{3,3}^e - x_{1,3}^e)(\sigma_3^e - \sigma_1^e + \sigma_6^e) + (x_{4,3}^e - x_{1,3}^e)(\sigma_1^e - \sigma_2^e + \sigma_4^e) \end{aligned} \right). \end{aligned}$$

Let f_2^e , f_3^e , and f_4^e stand for the faces that are respectively opposite to the nodes \mathbf{x}_2^e , \mathbf{x}_3^e , and \mathbf{x}_4^e . The following oriented closed curves

$$(\mathbf{x}_1^e, \mathbf{x}_4^e, \mathbf{x}_3^e, \mathbf{x}_1^e), \quad (\mathbf{x}_1^e, \mathbf{x}_2^e, \mathbf{x}_4^e, \mathbf{x}_1^e), \quad \text{and} \quad (\mathbf{x}_1^e, \mathbf{x}_3^e, \mathbf{x}_4^e, \mathbf{x}_1^e),$$

see also Fig. 4.5, are respectively the positively oriented boundaries of the faces f_2^e , f_3^e , and f_4^e with the outer unit normal vectors \mathbf{n}_2^e , \mathbf{n}_3^e , and \mathbf{n}_4^e . Now, using Theorem 3.5, we arrive at

$$\mathbf{curl}_{\mathbf{x}}(\boldsymbol{\pi}^e(\mathbf{v}|_{\overline{K^e}}(\mathbf{x}))) = \frac{-1}{3 \text{meas}(K^e)} \mathbf{R}^e \cdot \begin{pmatrix} \int_{f_2^e} \mathbf{curl}_{\mathbf{x}}(\mathbf{v}|_{\overline{K^e}}(\mathbf{x})) \cdot \mathbf{n}_2^e(\mathbf{x}) \, d\mathbf{S} \\ \int_{f_3^e} \mathbf{curl}_{\mathbf{x}}(\mathbf{v}|_{\overline{K^e}}(\mathbf{x})) \cdot \mathbf{n}_3^e(\mathbf{x}) \, d\mathbf{S} \\ \int_{f_4^e} \mathbf{curl}_{\mathbf{x}}(\mathbf{v}|_{\overline{K^e}}(\mathbf{x})) \cdot \mathbf{n}_4^e(\mathbf{x}) \, d\mathbf{S} \end{pmatrix}.$$

Since for $i = 2, 3, 4$ and $j = 1, 2, 3$

$$|x_{i,j}^e - x_{1,j}^e| \leq h^e \quad \text{and} \quad \left| \int_{f_i^e} \mathbf{curl}_{\mathbf{x}}(\mathbf{v}|_{\overline{K^e}}(\mathbf{x})) \cdot \mathbf{n}_j^e(\mathbf{x}) \, d\mathbf{S} \right| \leq \frac{1}{2} \max_{\mathbf{x} \in \Omega} \|\mathbf{curl}_{\mathbf{x}}(\mathbf{v}(\mathbf{x}))\| (h^e)^2$$

and due to (4.71), the relation (4.74) holds and we get

$$\|\mathbf{curl}_{\mathbf{x}}(\boldsymbol{\pi}^e(\mathbf{v}|_{\overline{K^e}}(\mathbf{x})))\| \leq \frac{3 \max_{\mathbf{x} \in \Omega} \|\mathbf{curl}_{\mathbf{x}}(\mathbf{v}(\mathbf{x}))\| (C_{15})^3}{8\pi},$$

hence

$$C_{14} := \frac{3 \max_{\mathbf{x} \in \Omega} \|\mathbf{curl}_{\mathbf{x}}(\mathbf{v}(\mathbf{x}))\| (C_{15})^3}{8\pi}.$$

□

Chapter 5

Abstract optimal shape design problem

In this chapter, we will introduce a shape optimization problem governed by the abstract linear elliptic boundary vector–value problem, the weak formulation of which was introduced in Section 3.4. We will state a continuous setting of the shape optimization problem and prove the existence of a solution. Further, we will deal with a regularized formulation and with a convergence of the regularized solutions to the true one. Finally, we will introduce a discretized shape optimization problem and assumptions under which the discretized and regularized solutions converge to the true one. The theory here is very similar to the one presented by HASLINGER AND NEITTAANMÄKI [85]. The main difference is that we fix the computational domain Ω and the shapes control the material distribution rather than the boundary $\partial\Omega$, which is usual in mechanics. Moreover, we will be concerned with a *multi–state optimization*, where several state problems with the same bilinear form, but different linear functionals, in our case, different current excitations, are involved.

Let us recall some basic literature on shape optimization: BEGIS AND GLOWINSKI [19], MURAT AND SIMON [140], PIRONNEAU [159], HASLINGER AND NEITTAANMÄKI [85], HASLINGER AND MÄKINEN [83], BENDØE [21], SOKOLOWSKI AND ZOLESIO [196], BÖRNER [24], DELFOUR AND ZOLESIO [54], KAWOHL ET AL. [107], MOHAMMADI AND PIRONNEAU [135]. Besides the basic textbooks, one can find a lot of theoretical analysis in BUCUR AND ZOLESIO [35], CHLEBOUN AND MÄKINEN [44], PEICHL AND RING [155, 156], PETERSSON [157], PETERSSON AND HASLINGER [158]. Papers focused on applications in electromagnetism are, for example, DI BARBA ET AL. [18], BRANDSTÄTTER ET AL. [30], LUKÁŠ [123], MARROCCO AND PIRONNEAU [132], TAKAHASHI [206]. An optimization of mechanical components is presented in HAASE AND LINDNER [76].

5.1 A fundamental theorem

Let us suppose that we have a set \mathcal{U} being a subset of a normed linear space V and we have the *cost functional* $\mathcal{J} : \mathcal{U} \mapsto \mathbb{R}$. The optimization problem reads as follows:

$$\left. \begin{array}{l} \text{Find } \alpha^* \in \mathcal{U}: \\ \mathcal{J}(\alpha^*) \leq \mathcal{J}(\alpha) \quad \forall \alpha \in \mathcal{U} \end{array} \right\} . \quad (P)$$

We say that the set \mathcal{U} is *compact* if for any sequence $\{\alpha_n\} \subset \mathcal{U}$ there exist a subsequence $\{\alpha_{n_k}\}_{k=1}^\infty \subset \{\alpha_n\}_{n=1}^\infty$ and $\alpha \in \mathcal{U}$ such that $\alpha_{n_k} \rightarrow \alpha$ in V , as $k \rightarrow \infty$.

The next fundamental theorem of functional analysis examine the existence of a solution to the problem (P).

Theorem 5.1. *Let \mathcal{U} be a compact subset of the normed linear space V and let $\mathcal{J} : \mathcal{U} \mapsto \mathbb{R}$ be a continuous functional. Then there exists a solution α^* to the problem (P).*

Proof. See HASLINGER AND NEITTAANMÄKI [85, p. 6–7]. □

5.2 Continuous setting

Results of this section are presented in LUKÁŠ [120]. Let us recall that in Assumption 3.1 we suppose that $\Omega \subset \mathbb{R}^m$, $m \in \{2, 3\}$, $\Omega \in \mathcal{L}$ is a computational domain that is forever fixed independently from any parameter or variable.

5.2.1 Admissible shapes

The symbol α stands for a shape, which is a continuous function, i.e., $\alpha \in C(\bar{\omega})$, where $\omega \subset \mathbb{R}^{m-1}$ is a nonempty polyhedral domain, see also Fig. 5.1. We assume that for all the admissible shapes α there exists a common Lipschitz constant $C_{16} > 0$, i.e.,

$$\forall \mathbf{x}, \mathbf{y} \in \bar{\omega} : |\alpha(\mathbf{x}) - \alpha(\mathbf{y})| \leq C_{16} \|\mathbf{x} - \mathbf{y}\|. \quad (5.1)$$

We further employ the box constraints, i.e., there exist $\alpha_l, \alpha_u \in \mathbb{R}$ such that

$$\forall \mathbf{x} \in \bar{\omega} : \alpha_l \leq \alpha(\mathbf{x}) \leq \alpha_u. \quad (5.2)$$

Then the set of admissible shapes is as follows:

$$\mathcal{U} := \{\alpha \in C(\bar{\omega}) \mid (5.1) \text{ and } (5.2) \text{ hold}\}, \quad (5.3)$$

equipped with the uniform convergence, see (3.16),

$$\alpha_n \rightarrow \alpha \text{ in } \mathcal{U} \text{ if } \alpha_n \rightrightarrows \alpha, \text{ as } n \rightarrow \infty. \quad (5.4)$$

Lemma 5.1. *\mathcal{U} is compact.*

Proof. Let $\{\alpha_n\}_{n=1}^{\infty} \subset \mathcal{U}$ be an arbitrary sequence of shapes. By (5.2) the sequence is uniformly bounded and by (5.1) it is equicontinuous. Then by Theorem 3.2 there exist a subsequence $\{\alpha_{n_k}\}_{k=1}^{\infty} \subset \{\alpha_n\}_{n=1}^{\infty}$ and $\alpha \in C(\bar{\omega})$ such that

$$\alpha_{n_k} \rightrightarrows \alpha \text{ in } \bar{\omega}, \text{ as } k \rightarrow \infty.$$

It is easy to see that α satisfies both (5.1) and (5.2), which completes the proof. □

In Chapter 7, we will deal with an application where we will be at the end looking for smooth shapes, e.g., Bézier curves or patches, cf. FARIN [59], rather than for continuous ones. To this end, being inspired by CHLEBOUN AND MÄKINEN [44], we introduce a parameterization, i.e., a nonempty compact set of design parameters $\Upsilon \subset \mathbb{R}^{n_\Upsilon}$, $n_\Upsilon \in \mathbb{N}$, and a continuous nonsurjective mapping

$$F : \Upsilon \mapsto \mathcal{U}. \quad (5.5)$$

Finally, without losing generality we assume that the shape α controls the following decomposition of Ω into the subdomains $\Omega_0(\alpha)$ and $\Omega_1(\alpha)$

$$\bar{\Omega} = \overline{\Omega_0(\alpha)} \cup \overline{\Omega_1(\alpha)}, \quad \Omega_0(\alpha) \cap \Omega_1(\alpha) = \emptyset$$

such that $\text{graph}(\alpha) \subset \partial\Omega_0(\alpha) \cap \partial\Omega_1(\alpha)$, $\text{meas}(\Omega_0(\alpha)) > 0$, and $\text{meas}(\Omega_1(\alpha)) > 0$, (5.6)

as depicted in Fig. 5.1, where the graph is defined by

$$\text{graph}(\alpha) := \{(x_1, \dots, x_{m-1}, y) \in \mathbb{R}^m \mid \mathbf{x} := (x_1, \dots, x_{m-1}) \in \bar{\omega} \text{ and } y = \alpha(\mathbf{x})\}.$$

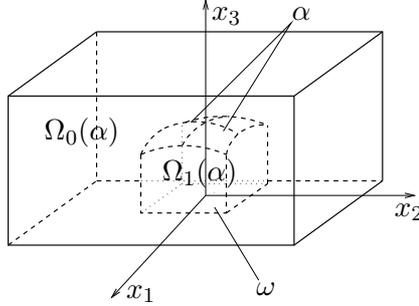


Figure 5.1: Decomposition of Ω

5.2.2 Multistate problem

The shape optimization problem is governed by a *state problem* that describes the related physical field. This is in our case the weak formulation (W) which is controlled by the shape α via the material distribution \mathbf{D} . We restrict ourselves to the case of two materials with spatial constant physical properties.

Assumption 5.1. *We assume that the material function $\mathbf{D} \equiv \mathbf{D}_\alpha$ is controlled by the shape $\alpha \in \mathcal{U}$ as follows:*

$$\mathbf{D}_\alpha(\mathbf{x}) := \begin{cases} \mathbf{D}_0 & , \mathbf{x} \in \Omega_0(\alpha) \\ \mathbf{D}_1 & , \mathbf{x} \in \Omega_1(\alpha) \end{cases}, \quad (5.7)$$

where $\mathbf{D}_0, \mathbf{D}_1 \in \mathbb{R}^{\nu_2 \times \nu_2}$, $\nu_2 \in \mathbb{N}$, are constant and positive definite matrices which correspond to the particular materials.

From the positive definiteness of \mathbf{D}_0 and \mathbf{D}_1 , the relation (3.39) follows. The bilinear form (3.40) now reads

$$a_\alpha(\mathbf{v}, \mathbf{u}) := \int_{\Omega_0(\alpha)} \mathbf{B}(\mathbf{v}) \cdot (\mathbf{D}_0 \cdot \mathbf{B}(\mathbf{u})) \, d\mathbf{x} + \int_{\Omega_1(\alpha)} \mathbf{B}(\mathbf{v}) \cdot (\mathbf{D}_1 \cdot \mathbf{B}(\mathbf{u})) \, d\mathbf{x}, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega), \quad (5.8)$$

where both the operator \mathbf{B} and the space $\mathbf{H}_0(\mathbf{B}; \Omega)$ were described in Section 3.3.6.

Concerning the linear functional (3.41), we distinguish several right-hand sides \mathbf{f} , e.g., several current excitations in case of magnetostatics. The linear functional (3.41) reads as follows:

$$f^v(\mathbf{v}) := \int_{\Omega} \mathbf{f}^v \cdot \mathbf{v} \, d\mathbf{x}, \quad \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega), \quad \text{for } v = 1, 2, \dots, n_v, \quad (5.9)$$

where $n_v \in \mathbb{N}$ is a number of the considered right-hand sides $\mathbf{f}^v \in [L^2(\Omega)]^{\nu_1}$, $\nu_1 \in \mathbb{N}$, such that they fulfill

$$\forall \mathbf{p} \in \mathbf{Ker}(\mathbf{B}; \Omega) : \int_{\Omega} \mathbf{f}^v \cdot \mathbf{p} \, d\mathbf{x} = 0 \quad \text{for each } v = 1, 2, \dots, n_v,$$

where $\mathbf{Ker}(\mathbf{B}; \Omega)$ is defined by (3.35).

Assumption 5.2. We assume that for each $v = 1, \dots, n_v$ the right-hand sides \mathbf{f}^v are independent from $\alpha \in \mathcal{U}$.

Now for each $v = 1, 2, \dots, n_v$ the state problem (W) is rewritten as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{u}^v(\alpha) \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega): \\ a_\alpha(\mathbf{v}, \mathbf{u}^v(\alpha)) = f^v(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega) \end{array} \right\}, \quad (W^v(\alpha))$$

where the space $\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ is defined by (3.36).

Lemma 5.2. For each $\alpha \in \mathcal{U}$ and $v = 1, 2, \dots, n_v$ there exists exactly one solution $\mathbf{u}^v(\alpha) \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ to the problem ($W^v(\alpha)$). Moreover, there exists a positive constant C_7 such that

$$\forall \alpha \in \mathcal{U} : \|\mathbf{u}^v(\alpha)\|_{\mathbf{B},\Omega} \leq C_7 \|\mathbf{f}^v\|_{\nu_1,0,\Omega}, \quad v = 1, 2, \dots, n_v.$$

Proof. Taking an arbitrary shape $\alpha \in \mathcal{U}$ and any $v = 1, 2, \dots, n_v$, the proof is the same as the one of Theorem 3.16, where the symbols a , f , \mathbf{D} , and \mathbf{f} are replaced by a_α , f^v , \mathbf{D}_α , and \mathbf{f}^v , respectively. \square

Lemma 5.3. For each $v = 1, 2, \dots, n_v$ the mapping $\mathbf{u}^v : \mathcal{U} \mapsto \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ is continuous on \mathcal{U} .

Proof. Let $v = 1, 2, \dots, n_v$ be arbitrary and let $\{\alpha_n\}_{n=1}^\infty \subset \mathcal{U}$ be a sequence such that $\alpha_n \rightrightarrows \alpha \in \mathcal{U}$. To make it readable, we denote

$$\mathbf{u} := \mathbf{u}^v(\alpha) \quad \text{and} \quad \mathbf{u}_n := \mathbf{u}^v(\alpha_n) . \quad (5.10)$$

We observe that (3.46) holds independently of $\alpha \in \mathcal{U}$. Therefore, by the definitions of ($W^v(\alpha_n)$) and ($W^v(\alpha)$), we have a simple case of Lemma 4.1

$$\begin{aligned} \|\mathbf{u}_n - \mathbf{u}\|_{\mathbf{B},\Omega}^2 &\leq \frac{1}{C_7} a_{\alpha_n}(\mathbf{u}_n - \mathbf{u}, \mathbf{u}_n - \mathbf{u}) = \frac{1}{C_7} (f^v(\mathbf{u}_n - \mathbf{u}) - a_{\alpha_n}(\mathbf{u}, \mathbf{u}_n - \mathbf{u})) = \\ &= \frac{1}{C_7} (a_\alpha(\mathbf{u}, \mathbf{u}_n - \mathbf{u}) - a_{\alpha_n}(\mathbf{u}, \mathbf{u}_n - \mathbf{u})), \end{aligned} \quad (5.11)$$

where C_7 is the $\mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ -ellipticity constant.

Further, we denote the characteristic functions of the sets $\Omega_0(\alpha)$ and $\Omega_1(\alpha)$ by $\chi_{\Omega_0(\alpha)}(\mathbf{x})$ and $\chi_{\Omega_1(\alpha)}(\mathbf{x})$, respectively. Since $\alpha_n \rightrightarrows \alpha$, the following holds

$$\chi_{\Omega_0(\alpha_n)}(\mathbf{x}) \rightarrow \chi_{\Omega_0(\alpha)}(\mathbf{x}) \quad \text{and} \quad \chi_{\Omega_1(\alpha_n)}(\mathbf{x}) \rightarrow \chi_{\Omega_1(\alpha)}(\mathbf{x}) \quad \text{a.e. in } \Omega, \quad \text{as } n \rightarrow \infty . \quad (5.12)$$

Now, we write down the right-hand side of (5.11) and using (3.45) and the Cauchy–Schwarz

inequality in $[L^2(\Omega)]^{\nu_2}$ we get

$$\begin{aligned}
|a_\alpha(\mathbf{u}, \mathbf{u}_n - \mathbf{u}) - a_{\alpha_n}(\mathbf{u}, \mathbf{u}_n - \mathbf{u})| &= \left| \int_{\Omega_0(\alpha)} \mathbf{B}(\mathbf{u}) \cdot (\mathbf{D}_0 \cdot \mathbf{B}(\mathbf{u}_n - \mathbf{u})) \, d\mathbf{x} + \right. \\
&\quad + \int_{\Omega_1(\alpha)} \mathbf{B}(\mathbf{u}) \cdot (\mathbf{D}_1 \cdot \mathbf{B}(\mathbf{u}_n - \mathbf{u})) \, d\mathbf{x} - \int_{\Omega_0(\alpha_n)} \mathbf{B}(\mathbf{u}) \cdot (\mathbf{D}_0 \cdot \mathbf{B}(\mathbf{u}_n - \mathbf{u})) \, d\mathbf{x} \\
&\quad \left. - \int_{\Omega_1(\alpha_n)} \mathbf{B}(\mathbf{u}) \cdot (\mathbf{D}_1 \cdot \mathbf{B}(\mathbf{u}_n - \mathbf{u})) \, d\mathbf{x} \right| \leq \\
&\leq \left| \int_{\Omega} \{(\chi_{\Omega_0(\alpha)} - \chi_{\Omega_0(\alpha_n)}) \mathbf{B}(\mathbf{u})\} \cdot (\mathbf{D}_0 \cdot \mathbf{B}(\mathbf{u}_n - \mathbf{u})) \, d\mathbf{x} \right| + \\
&\quad + \left| \int_{\Omega} \{(\chi_{\Omega_1(\alpha)} - \chi_{\Omega_1(\alpha_n)}) \mathbf{B}(\mathbf{u})\} \cdot (\mathbf{D}_1 \cdot \mathbf{B}(\mathbf{u}_n - \mathbf{u})) \, d\mathbf{x} \right| \leq \\
&\leq d \left(\|(\chi_{\Omega_0(\alpha)} - \chi_{\Omega_0(\alpha_n)}) \mathbf{B}(\mathbf{u})\|_{0,\nu_2,\Omega} + \right. \\
&\quad \left. + \|(\chi_{\Omega_1(\alpha)} - \chi_{\Omega_1(\alpha_n)}) \mathbf{B}(\mathbf{u})\|_{0,\nu_2,\Omega} \right) \cdot \|\mathbf{B}(\mathbf{u}_n - \mathbf{u})\|_{0,\nu_2,\Omega}.
\end{aligned} \tag{5.13}$$

For a better clarity, we introduce the symbols

$$\mathcal{A}_0(n) := \|(\chi_{\Omega_0(\alpha)} - \chi_{\Omega_0(\alpha_n)}) \mathbf{B}(\mathbf{u})\|_{0,\nu_2,\Omega}, \quad \mathcal{A}_1(n) := \|(\chi_{\Omega_1(\alpha)} - \chi_{\Omega_1(\alpha_n)}) \mathbf{B}(\mathbf{u})\|_{0,\nu_2,\Omega}.$$

Then, the relation (5.13) reads as follows:

$$\begin{aligned}
|a_\alpha(\mathbf{u}, \mathbf{u}_n - \mathbf{u}) - a_{\alpha_n}(\mathbf{u}, \mathbf{u}_n - \mathbf{u})| &\leq d(\mathcal{A}_0(n) + \mathcal{A}_1(n)) \|\mathbf{B}(\mathbf{u}_n - \mathbf{u})\|_{0,\nu_2,\Omega} \leq \\
&\leq d(\mathcal{A}_0(n) + \mathcal{A}_1(n)) \|\mathbf{u}_n - \mathbf{u}\|_{\mathbf{B},\Omega}.
\end{aligned} \tag{5.14}$$

From the relation (5.12), it follows that

$$\left. \begin{aligned} &|\chi_{\Omega_0(\alpha)}(\mathbf{x}) - \chi_{\Omega_0(\alpha_n)}(\mathbf{x})|^2 \|\mathbf{B}(\mathbf{u}(\mathbf{x}))\|^2 \rightarrow 0 \\ &|\chi_{\Omega_1(\alpha)}(\mathbf{x}) - \chi_{\Omega_1(\alpha_n)}(\mathbf{x})|^2 \|\mathbf{B}(\mathbf{u}(\mathbf{x}))\|^2 \rightarrow 0 \end{aligned} \right\} \text{a.e. in } \Omega, \text{ as } n \rightarrow \infty \tag{5.15}$$

and, since $\mathbf{B}(\mathbf{u}) \in [L^2(\Omega)]^{\nu_2}$, the functions on the left-hand side of (5.15) are in $L^1(\Omega)$ and each bounded by $\|\mathbf{B}(\mathbf{u})\|^2 \in L^1(\Omega)$ from above. Now, using Theorem 3.6, we arrive at

$$\mathcal{A}_0(n) \rightarrow 0 \text{ and } \mathcal{A}_1(n) \rightarrow 0, \text{ as } n \rightarrow \infty. \tag{5.16}$$

Combining (5.10), (5.11), (5.14), and (5.16), we have proven the statement

$$\mathbf{u}^v(\alpha_n) \rightarrow \mathbf{u}^v(\alpha) \text{ in } \mathbf{H}(\mathbf{B}; \Omega), \text{ as } n \rightarrow \infty.$$

□

5.2.3 Shape optimization problem

Let $\mathcal{I} : \mathcal{U} \times [[L^2(\Omega)]^{\nu_2}]^{n_v} \mapsto \mathbb{R}$ be a continuous functional. Using $(W^v(\alpha))$, we define the cost functional $\mathcal{J} : \mathcal{U} \mapsto \mathbb{R}$ by

$$\mathcal{J}(\alpha) := \mathcal{I}(\alpha, \mathbf{B}(\mathbf{u}^1(\alpha)), \mathbf{B}(\mathbf{u}^2(\alpha)), \dots, \mathbf{B}(\mathbf{u}^{n_v}(\alpha))), \quad \alpha \in \mathcal{U}.$$

The continuous optimization problem then, in accordance with Section 5.1, reads as follows:

$$\left. \begin{array}{l} \text{Find } \alpha^* \in \mathcal{U}: \\ \mathcal{J}(\alpha^*) \leq \mathcal{J}(\alpha) \quad \forall \alpha \in \mathcal{U} \end{array} \right\}. \quad (P)$$

Theorem 5.2. *There exists $\alpha^* \in \mathcal{U}$ that is a solution to (P).*

Proof. By Lemma 5.1, \mathcal{U} is a compact subset of the normed linear space $C(\bar{\omega})$. Using the continuity of \mathcal{I} on $\mathcal{U} \times [[L^2(\Omega)]^{\nu_2}]^{n_\nu}$ and Lemma 5.3, the continuity of \mathcal{J} on \mathcal{U} follows. Now, Theorem 5.1 completes the proof. \square

Moreover, we use (5.5) to define the cost functional $\tilde{\mathcal{J}} : \Upsilon \mapsto \mathbb{R}$

$$\tilde{\mathcal{J}}(\mathbf{p}) := \mathcal{J}(F(\mathbf{p})), \quad \mathbf{p} \in \Upsilon.$$

Then, by the compactness of Υ , by the continuity of F on Υ , and by the same arguments as in the proof of Theorem 5.2, there exists a solution $\mathbf{p}^* \in \Upsilon$ to the optimization problem

$$\left. \begin{array}{l} \text{Find } \mathbf{p}^* \in \Upsilon: \\ \tilde{\mathcal{J}}(\mathbf{p}^*) \leq \tilde{\mathcal{J}}(\mathbf{p}) \quad \forall \mathbf{p} \in \Upsilon \end{array} \right\}. \quad (\tilde{P})$$

5.3 Regularized setting

In this section, we will show a convergence of solutions of optimization problems whose state problems are regularized, as described in Section 3.4.1, to a continuous solution α^* .

Let $\varepsilon > 0$ be a regularization parameter. Due to (3.47) and (5.8), we introduce the regularized bilinear form controlled by the shape $\alpha \in \mathcal{U}$

$$a_{\varepsilon, \alpha}(\mathbf{v}, \mathbf{u}) := a_\alpha(\mathbf{v}, \mathbf{u}) + \varepsilon \int_{\Omega} \mathbf{v} \cdot \mathbf{u} \, dx, \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}(\mathbf{B}; \Omega).$$

For each $v = 1, 2, \dots, n_\nu$ the regularized weak formulation reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{u}_\varepsilon^v(\alpha) \in \mathbf{H}_0(\mathbf{B}; \Omega): \\ a_{\varepsilon, \alpha}(\mathbf{v}, \mathbf{u}_\varepsilon^v(\alpha)) = f^v(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{H}_0(\mathbf{B}; \Omega) \end{array} \right\}. \quad (W_\varepsilon^v(\alpha))$$

Lemma 5.4. *Let $\varepsilon > 0$, $\alpha \in \mathcal{U}$. Then for each $v = 1, 2, \dots, n_\nu$ there exists a unique solution $\mathbf{u}_\varepsilon^v(\alpha) \in \mathbf{H}_0(\mathbf{B}; \Omega)$ to the problem $(W_\varepsilon^v(\alpha))$. Moreover, there exists a positive constant $C_8(\varepsilon)$ such that*

$$\|\mathbf{u}_\varepsilon^v(\alpha)\|_{\mathbf{B}, \Omega} \leq C_8(\varepsilon) \|\mathbf{f}^v\|_{\nu_1, 0, \Omega} \quad \text{for each } v = 1, 2, \dots, n_\nu.$$

Proof. Taking $\varepsilon > 0$, an arbitrary shape $\alpha \in \mathcal{U}$, and any $v = 1, 2, \dots, n_\nu$, the proof is the same as the one of Theorem 3.17, while the symbols a_ε , f , \mathbf{D} , and \mathbf{f} are replaced by $a_{\varepsilon, \alpha}$, f^v , \mathbf{D}_α , and \mathbf{f}^v , respectively. \square

Lemma 5.5. *Let $\varepsilon > 0$. Then for each $v = 1, 2, \dots, n_\nu$ the mapping $\mathbf{u}_\varepsilon^v : \mathcal{U} \mapsto \mathbf{H}_0(\mathbf{B}; \Omega)$ is continuous on \mathcal{U} .*

Proof. Taking any $\varepsilon > 0$, the proof is the same as the one of Lemma 5.3, where all the proper symbols are subscribed with ε . \square

Lemma 5.6. *Let $\alpha \in \mathcal{U}$. Then for each $v = 1, 2, \dots, n_v$ the following holds*

$$\mathbf{B}(\mathbf{u}_\varepsilon^v(\alpha)) \rightarrow \mathbf{B}(\mathbf{u}^v(\alpha)) \text{ in } [L^2(\Omega)]^{\nu^2}, \text{ as } \varepsilon \rightarrow 0_+, \quad (5.17)$$

where $\mathbf{u}_\varepsilon^v(\alpha) \in \mathbf{H}_0(\mathbf{B}; \Omega)$ are the solutions to $(W_\varepsilon^v(\alpha))$ and $\mathbf{u}^v(\alpha) \in \mathbf{H}_{0,\perp}(\mathbf{B}; \Omega)$ is the solution to $(W^v(\alpha))$.

Proof. Taking an arbitrary shape $\alpha \in \mathcal{U}$ and any $v = 1, 2, \dots, n_v$, the proof is the same as the one of Theorem 3.18, where we replace the symbols related to the problems (W) and (W_ε) by the corresponding symbols related to $(W^v(\alpha))$ and $(W_\varepsilon^v(\alpha))$, respectively. \square

Now, we return to the shape optimization problem. We introduce the regularized cost functional by

$$\mathcal{J}_\varepsilon(\alpha) := \mathcal{I}(\alpha, \mathbf{B}(\mathbf{u}_\varepsilon^1(\alpha)), \mathbf{B}(\mathbf{u}_\varepsilon^2(\alpha)), \dots, \mathbf{B}(\mathbf{u}_\varepsilon^{n_v}(\alpha))), \quad \alpha \in \mathcal{U}.$$

The regularized shape optimization problem then reads as follows:

$$\left. \begin{array}{l} \text{Find } \alpha_\varepsilon^* \in \mathcal{U}: \\ \mathcal{J}_\varepsilon(\alpha_\varepsilon^*) \leq \mathcal{J}_\varepsilon(\alpha) \quad \forall \alpha \in \mathcal{U} \end{array} \right\}. \quad (P_\varepsilon)$$

Theorem 5.3. *Let $\varepsilon > 0$. Then there exists $\alpha_\varepsilon^* \in \mathcal{U}$ that is a solution to (P_ε) .*

Proof. Taking any $\varepsilon > 0$, the proof is fairly the same as the one of Theorem 5.2, where we use the symbol \mathcal{J}_ε instead of \mathcal{J} , and Lemma 5.5 instead of Lemma 5.3. \square

Theorem 5.4. *Let $\{\varepsilon_n\}_{n=1}^\infty \subset \mathbb{R}$ be a sequence of positive regularization parameters such that $\varepsilon_n \rightarrow 0_+$, as $n \rightarrow \infty$ and let $\alpha_{\varepsilon_n}^* \in \mathcal{U}$ be the corresponding solutions to the problems (P_{ε_n}) . Then there exist a subsequence $\{\varepsilon_{n_k}\}_{k=1}^\infty \subset \{\varepsilon_n\}_{n=1}^\infty$ and a shape $\alpha^* \in \mathcal{U}$ such that*

$$\alpha_{\varepsilon_{n_k}}^* \rightarrow \alpha^* \text{ in } \mathcal{U}, \text{ as } k \rightarrow \infty$$

holds and, moreover, α^* is a solution to the problem (P) .

Proof. By Theorem 5.3, for each $\varepsilon_n > 0$ there exists $\alpha_{\varepsilon_n}^* \in \mathcal{U}$ which is a solution to (P_{ε_n}) . By Lemma 5.1 there exists a subsequence of shapes $\{\alpha_{\varepsilon_{n_k}}^*\}_{k=1}^\infty \subset \{\alpha_{\varepsilon_n}^*\}_{n=1}^\infty$ and a shape $\alpha^* \in \mathcal{U}$ such that

$$\alpha_{\varepsilon_{n_k}}^* \rightarrow \alpha^* \text{ in } \mathcal{U}, \text{ as } k \rightarrow \infty. \quad (5.18)$$

Let $\alpha \in \mathcal{U}$ be arbitrary. Then, due to the definition of $(P_{\varepsilon_{n_k}})$, for any $k \in \mathbb{N}$ we get

$$\mathcal{J}_{\varepsilon_{n_k}}(\alpha_{\varepsilon_{n_k}}^*) \leq \mathcal{J}_{\varepsilon_{n_k}}(\alpha). \quad (5.19)$$

Using Lemma 5.6 and the continuity of \mathcal{I} , the right-hand side of (5.19) converges as follows:

$$\mathcal{J}_{\varepsilon_{n_k}}(\alpha) \rightarrow \mathcal{J}(\alpha), \text{ as } k \rightarrow \infty.$$

Using (5.18), Lemma 5.5, Lemma 5.6, and the continuity of \mathcal{I} , the left-hand side of (5.19) also converges

$$\mathcal{J}_{\varepsilon_{n_k}}(\alpha_{\varepsilon_{n_k}}^*) \rightarrow \mathcal{J}(\alpha^*), \text{ as } k \rightarrow \infty.$$

Therefore, we have proven that for any $\alpha \in \mathcal{U}$

$$\mathcal{J}(\alpha^*) \leq \mathcal{J}(\alpha).$$

\square

Finally, we introduce the regularized cost functional $\tilde{\mathcal{J}}_\varepsilon : \Upsilon \mapsto \mathbb{R}$ by

$$\tilde{\mathcal{J}}_\varepsilon(\mathbf{p}) := \mathcal{J}_\varepsilon(F(\mathbf{p})), \quad \mathbf{p} \in \Upsilon.$$

Then, the regularized optimization problem reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{p}_\varepsilon^* \in \Upsilon: \\ \tilde{\mathcal{J}}_\varepsilon(\mathbf{p}_\varepsilon^*) \leq \tilde{\mathcal{J}}_\varepsilon(\mathbf{p}) \quad \forall \mathbf{p} \in \Upsilon \end{array} \right\}. \quad (\tilde{P}_\varepsilon)$$

In the same fashion as in the case of (P_ε) , we can derive the existence theory as well as the approximation property of the problem (\tilde{P}_ε) .

5.4 Discretized setting

In this section, we introduce a setting of the shape optimization problem (P_ε) discretized by the finite element method. We will prove a convergence of the approximate solutions to the true one.

Let $\varepsilon > 0$ be a regularization parameter and let $h > 0$ be a discretization parameter, see (4.31). With any $h > 0$ we associate a nonempty polyhedral computational subdomain $\Omega^h \subset \Omega$ such that (4.35) is satisfied.

5.4.1 Discretized set of admissible shapes

First, we introduce a finite-dimensional approximation of the set \mathcal{U} of admissible shapes. Let $\mathcal{T}_\omega^h := \{\omega_1^h, \dots, \omega_{n_\omega^h}^h\}$, $n_\omega^h \in \mathbb{N}$, be a discretization of the nonempty polyhedral domain $\omega \subset \mathbb{R}^{m-1}$. Let $P^1(\mathcal{T}_\omega^h) \subset C(\bar{\omega})$ denote a space of continuous functions that are linear over $\bar{\omega}_i^h$ for each $i = 1, \dots, n_\omega^h$. We denote the corners of $\bar{\omega}_i^h$ by $\mathbf{x}_{\omega_i^h,1}^h, \dots, \mathbf{x}_{\omega_i^h,m}^h \in \mathbb{R}^{m-1}$. By $\mathbf{x}_{\omega,1}^h, \dots, \mathbf{x}_{\omega,n_{\mathbf{x}_\omega^h}^h}^h \in \mathbb{R}^{m-1}$ we denote all the nodes of the discretization \mathcal{T}_ω^h .

Now, we discretize the condition (5.1) as follows:

$$\left| \alpha^h(\mathbf{x}_{\omega_i^h,j}^h) - \alpha^h(\mathbf{x}_{\omega_i^h,k}^h) \right| \leq C_{16} \left\| \mathbf{x}_{\omega_i^h,j}^h - \mathbf{x}_{\omega_i^h,k}^h \right\| \quad \text{for } i = 1, \dots, n_\omega^h, \quad j, k = 1, \dots, m, \quad j \neq k, \quad (5.20)$$

which in total involves n_ω^h or $3n_\omega^h$ conditions in case of $m = 2$ or $m = 3$, respectively. Discretized box constraints (5.2) include the following $n_{\mathbf{x}_\omega^h}$ conditions

$$\alpha_1 \leq \alpha^h(\mathbf{x}_{\omega,i}^h) \leq \alpha_u \quad \text{for } i = 1, \dots, n_{\mathbf{x}_\omega^h}. \quad (5.21)$$

Then the discretized set of admissible shapes is as follows:

$$\mathcal{U}^h := \left\{ \alpha^h \in P^1(\mathcal{T}_\omega^h) \mid (5.20) \text{ and } (5.21) \text{ hold} \right\},$$

equipped with the uniform convergence (5.4). Obviously for each $h > 0$: $\mathcal{U}^h \subset \mathcal{U}$ and, by the definition of $P^1(\mathcal{T}_\omega^h)$, \mathcal{U}^h is finite dimensional.

Lemma 5.7. *For any $h > 0$ the set \mathcal{U}^h is compact.*

Proof. See the proof of Lemma 5.1. □

We introduce an interpolation operator $\pi_\omega^h : \mathcal{U} \mapsto P^1(\mathcal{T}_\omega^h)$ such that for each $\mathbf{x}_{\omega_i^h, j}^h$, a corner of ω_i^h , it holds that

$$\forall \alpha \in \mathcal{U} : \left[\pi_\omega^h(\alpha) \right] \left(\mathbf{x}_{\omega_i^h, j}^h \right) = \alpha \left(\mathbf{x}_{\omega_i^h, j}^h \right), \quad i = 1, \dots, n_\omega^h, \quad j = 1, \dots, m. \quad (5.22)$$

Lemma 5.8. *Let $\alpha \in \mathcal{U}$ be an arbitrary shape and let $\{h_n\}_{n=1}^\infty \subset \mathbb{R}$ be a sequence of positive discretization parameters. Then the following holds*

$$\pi_\omega^{h_n}(\alpha) \rightarrow \alpha \text{ in } \mathcal{U}, \text{ as } h_n \rightarrow 0_+.$$

Proof. See BEGIS AND GLOWINSKI [19]. □

Further, let $h > 0$ be given. Since for any $\alpha \in \mathcal{U}$: (5.1) implies (5.20), and (5.2) implies (5.21), then also $\alpha \in \mathcal{U}$ implies $\pi_\omega^h(\alpha) \in \mathcal{U}^h$. Moreover, as $F : \Upsilon \mapsto \mathcal{U}$, then for any $\mathbf{p} \in \Upsilon$ it follows that $\pi_\omega^h(F(\mathbf{p})) \in \mathcal{U}^h$. Therefore, we use Υ also for the discretized setting.

Finally, like in the continuous case we assume that a discretized shape α^h controls the decomposition of Ω^h into the subdomains $\Omega_0^h(\alpha^h)$ and $\Omega_1^h(\alpha^h)$ as follows:

$$\overline{\Omega^h} = \overline{\Omega_0^h(\alpha^h)} \cup \overline{\Omega_1^h(\alpha^h)}, \quad \Omega_0^h(\alpha^h) \cap \Omega_1^h(\alpha^h) = \emptyset$$

such that $\text{graph}(\alpha^h) \subset \partial\Omega_0^h(\alpha^h) \cap \partial\Omega_1^h(\alpha^h)$, $\text{meas}(\Omega_0^h(\alpha^h)) > 0$, and $\text{meas}(\Omega_1^h(\alpha^h)) > 0$, (5.23)

where the boundaries $\partial\Omega_0^h(\alpha^h)$, $\partial\Omega_1^h(\alpha^h)$ are polyhedral, as depicted in Fig. 5.2.

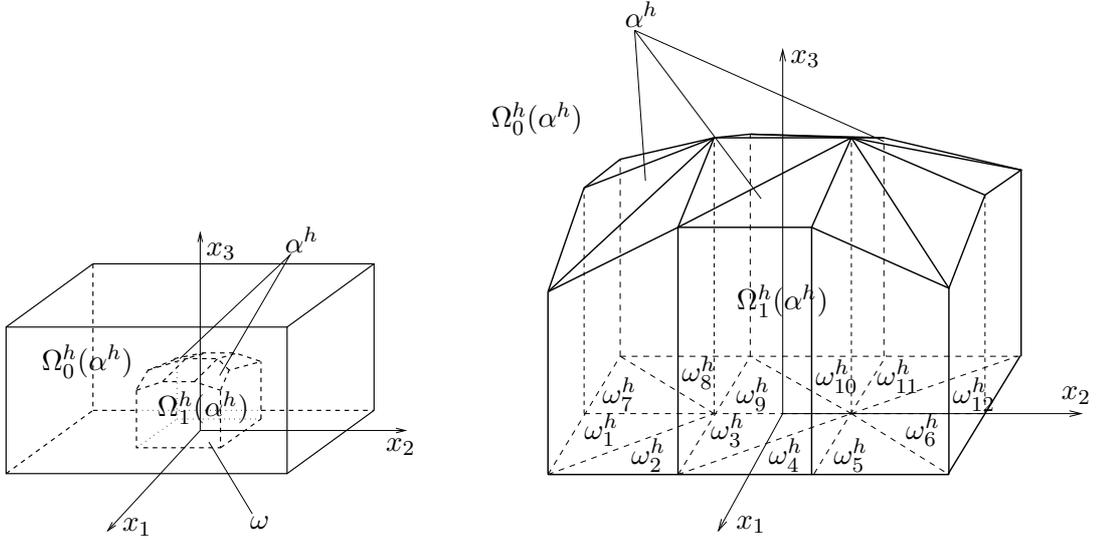


Figure 5.2: Decomposition of Ω^h

5.4.2 Discretized multistate problem

Let $\varepsilon > 0$ be a regularization parameter. For each $\alpha^h \in \mathcal{U}^h$ we provide a discretization $\mathcal{T}^h(\alpha^h) := \{K_1(\alpha^h), \dots, K_{n_{\Omega^h}}(\alpha^h)\}$ of the computational domain Ω^h such that

$$\forall K_i(\alpha^h) \in \mathcal{T}^h(\alpha^h) : K_i(\alpha^h) \subset \Omega_0^h(\alpha^h) \text{ or } K_i(\alpha^h) \subset \Omega_1^h(\alpha^h). \quad (5.24)$$

By $E^h := \{e_1, \dots, e_{n_{\Omega^h}}\}$ we denote the corresponding set of finite elements. For any $h > 0$ and $\alpha^h \in \mathcal{U}^h$ we suppose that Assumptions 4.1–4.7 hold. We introduce another assumption, which is due to HASLINGER AND NEITTAANMÄKI [85, p. 67].

Assumption 5.3. *We assume that for any $h > 0$ fixed the topology of the discretization grid $\mathcal{T}^h(\alpha^h)$ is independent from $\alpha^h \in \mathcal{U}^h$, we further assume that the coordinates of $\mathbf{x}_1^{e_i}(\alpha^h), \dots, \mathbf{x}_{m+1}^{e_i}(\alpha^h) \in \mathbb{R}^m$, see (4.25), which are the corners of $K_i(\alpha^h) \in \mathcal{T}^h(\alpha^h)$, still form a triangle ($m = 2$) or a tetrahedron ($m = 3$), and they depend continuously on $\alpha^h \in \mathcal{U}^h$.*

The regularized and discretized setting of the multistate problem reads as follows:

$$\left. \begin{aligned} &\text{Find } \mathbf{u}_\varepsilon^{v,h}(\alpha^h) \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h : \\ &a_{\varepsilon, \alpha^h}^h(\mathbf{v}^h, \mathbf{u}_\varepsilon^{v,h}(\alpha^h)) = f^{v,h}(\mathbf{v}^h) \quad \forall \mathbf{v}^h \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h \end{aligned} \right\}, \quad v = 1, \dots, n_v, \quad (W_\varepsilon^{v,h}(\alpha^h))$$

where the finite element space $\mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ is defined by (4.17), where further for each $\mathbf{v}^h, \mathbf{w}^h \in \mathbf{H}(\mathbf{B}; \Omega^h)$ we define

$$a_{\varepsilon, \alpha^h}^h(\mathbf{v}^h, \mathbf{w}^h) := \int_{\Omega^h} \mathbf{B}(\mathbf{v}^h) \cdot (\mathbf{D}_{\alpha^h}^h \cdot \mathbf{B}(\mathbf{w}^h)) \, d\mathbf{x} + \varepsilon \int_{\Omega^h} \mathbf{v}^h \cdot \mathbf{w}^h \, d\mathbf{x}, \quad (5.25)$$

in which, in virtue of Assumption 5.1,

$$\mathbf{D}_{\alpha^h}^h(\mathbf{x}) := \begin{cases} \mathbf{D}_0 & , \mathbf{x} \in \Omega_0^h(\alpha^h) \\ \mathbf{D}_1 & , \mathbf{x} \in \Omega_1^h(\alpha^h) \\ \mathbf{0} & , \mathbf{x} \in \Omega \setminus \Omega^h \end{cases}, \quad (5.26)$$

and where for each $\mathbf{v}^h \in \mathbf{H}(\mathbf{B}; \Omega^h)$ we set

$$f^{v,h}(\mathbf{v}^h) := \int_{\Omega^h} \mathbf{f}^{v,h} \cdot \mathbf{v}^h \, d\mathbf{x}, \quad v = 1, \dots, n_v,$$

in which, due to (4.18), $\mathbf{f}^{v,h} \in [L^2(\Omega^h)]^{\nu_1}$ are elementwise constant and such that

$$\left\| \mathbf{f}^{v,h} - \mathbf{f}^v \right\|_{\nu_1, 0, \Omega} \rightarrow 0, \quad \text{as } h \rightarrow 0_+, \quad v = 1, \dots, n_v, \quad (5.27)$$

where $\mathbf{f}^{v,h}(\mathbf{x}) := \mathbf{f}^v(\mathbf{x})$ in $\Omega \setminus \overline{\Omega^h}$. The following is in virtue of Assumption 5.2.

Assumption 5.4. *We assume that for each $v = 1, \dots, n_v$ the right-hand side $\mathbf{f}^{v,h}$ is independent of the shape $\alpha^h \in \mathcal{U}^h$.*

The following lemma assures that for any $\varepsilon > 0$, $h > 0$, and $v = 1, \dots, n_v$ fixed the mapping $\mathbf{u}_\varepsilon^{v,h} : \mathcal{U}^h \mapsto \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ is well defined.

Lemma 5.9. *For each $\varepsilon > 0$, $h > 0$, $\alpha^h \in \mathcal{U}^h$, and $v = 1, \dots, n_v$ there exists a unique solution $\mathbf{u}_\varepsilon^{v,h}(\alpha^h) \in \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ to the problem $(W_\varepsilon^{v,h}(\alpha^h))$.*

Proof. Since Ω^h is a polyhedron, then $\Omega^h \in \mathcal{L}$ and the statement follows by the same arguments as in the proof of Theorem 3.17. \square

Lemma 5.10. *Let $\varepsilon > 0$, $h > 0$. Then for each $v = 1, 2, \dots, n_v$ the mapping $\mathbf{u}_\varepsilon^{v,h} : \mathcal{U}^h \mapsto \mathbf{H}_0(\mathbf{B}; \Omega^h)^h$ is continuous on \mathcal{U}^h .*

Proof. We take an arbitrary $\varepsilon > 0$, $h > 0$, and $v = 1, \dots, n_v$. Note that we cannot use the same technique as in the proof of Lemma 5.3, since the settings $(W_\varepsilon^{v,h}(\alpha^h))$ differ from $\alpha^h \in \mathcal{U}^h$. Therefore, the estimate (5.11) cannot be established. We will rather exploit the algebraic structure of the mapping $\mathbf{u}_\varepsilon^{v,h}$.

In the similar manner as in (4.24), the solution to $(W_\varepsilon^{v,h}(\alpha^h))$ reads as follows:

$$\mathbf{u}_\varepsilon^{v,h}(\alpha^h) = \sum_{i=1}^n u_{\varepsilon,i}^{v,n}(\mathbf{x}^h(\alpha^h)) \boldsymbol{\xi}_i^h(\mathbf{x}^h(\alpha^h)), \quad (5.28)$$

where $\mathbf{x}^h(\alpha^h)$ denotes a vector of global coordinates of all element domains corners, which are by Assumption 5.3 continuously dependent on the shape $\alpha^h \in \mathcal{U}^h$, where further $\boldsymbol{\xi}_i^h(\mathbf{x}^h)$ denotes the global shape functions, and where we use the same notation for both the functions $\mathbf{u}_\varepsilon^{v,n}(\alpha^h)$ and $\mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h(\alpha^h))$

$$\mathbf{u}_\varepsilon^{v,n}(\alpha^h) \equiv \mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h(\alpha^h)) := \left(u_{\varepsilon,1}^{v,n}(\mathbf{x}^h(\alpha^h)), \dots, u_{\varepsilon,n}^{v,n}(\mathbf{x}^h(\alpha^h)) \right) \in \mathbb{R}^n,$$

which is the solution to the linear system (4.9). In this case, (4.9) reads as follows:

$$\mathbf{A}_\varepsilon^n(\mathbf{x}^h) \cdot \mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h) = \mathbf{f}^{v,n}(\mathbf{x}^h). \quad (5.29)$$

Now, let us take a look into the assembling of the matrix and the right-hand side vector in (5.29). Due to (4.21), the element contributions to them are

$$\begin{aligned} \left(\mathbf{A}_\varepsilon^n(\mathbf{x}^h) \right)_{i,j} &= \sum_{e \in E_i^h \cap E_j^h} \sum_{k,l=1}^{n^e} a_{\varepsilon,\mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)), \\ \left(\mathbf{f}^{v,n}(\mathbf{x}^h) \right)_i &= \sum_{e \in E_i^h} \sum_{k=1}^{n^e} f^{v,e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)), \end{aligned} \quad (5.30)$$

for $i, j = 1, \dots, n$, where E_i^h denotes the set of elements neighbouring with e_i , see (4.20), and \mathbf{x}^e is the vector of coordinates of the element domain corners, see (4.25). Using the map from the reference element r , the element contributions to the bilinear form and linear functional, see also (4.29) and (4.30), respectively are

$$\begin{aligned} a_{\varepsilon,\mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)) &= \\ &= \int_{K^r} \left(\mathbf{S}_B^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_k^r) \right) \cdot \left(\mathbf{D}_{\mathbf{x}^e}^e \cdot \left(\mathbf{S}_B^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_l^r) \right) \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}} + \\ &\quad + \varepsilon \int_{K^r} \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_k^r \right) \cdot \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_l^r \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}}, \\ f^{v,e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)) &= \int_{K^r} \mathbf{f}^e \cdot \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_k^r \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}}, \end{aligned} \quad (5.31)$$

where $k, l = 1, \dots, n^e$ and where we also used Assumptions 5.1 and 5.4.

The expressions (5.28)–(5.31) specify the function $\mathbf{u}_\varepsilon^{v,h}(\alpha^h)$. Now, we will prove its continuity. Let $\alpha^h \in \mathcal{U}^h$ be an arbitrary discretized admissible shape and let $\{\alpha_p^h\}_{p=1}^\infty \subset \mathcal{U}^h$ be such a sequence that

$$\alpha_p^h \rightarrow \alpha^h \text{ in } \mathcal{U}^h, \text{ as } p \rightarrow \infty.$$

Let us denote for each element $e \in E^h$, where E^h stands for the set of finite elements,

$$\mathbf{x}_p^e := \mathbf{x}^e(\alpha_p^h) \text{ and } \mathbf{x}^e := \mathbf{x}^e(\alpha^h).$$

By Assumption 5.3, for each element $e \in E^h$ we get

$$\mathbf{x}_p^e \rightarrow \mathbf{x}^e \text{ in } \mathbb{R}^{m(m+1)}, \text{ as } p \rightarrow \infty.$$

Again by Assumption 5.3, for each $e \in E^h$ and each $p \in \mathbb{N}$ the element domains $K^e(\mathbf{x}_p^e)$ as well as $K^e(\mathbf{x}^e)$ still form a triangle (in case of $m = 2$) or a tetrahedron (in case of $m = 3$), and it follows from the definition (4.28) that the matrices $\mathbf{R}^e(\mathbf{x}_p^e)$, $\mathbf{R}^e(\mathbf{x}^e)$ are nonsingular, and therefore

$$|\det(\mathbf{R}^e(\mathbf{x}_p^e))| > 0, \quad |\det(\mathbf{R}^e(\mathbf{x}^e))| > 0.$$

Moreover, as both \mathbf{R}^e and \mathbf{r}^e by (4.28) continuously depend on \mathbf{x}^e , we get:

$$\det(\mathbf{R}^e(\mathbf{x}_p^e)) \rightarrow \det(\mathbf{R}^e(\mathbf{x}^e)) \text{ in } \mathbb{R}, \text{ as } p \rightarrow \infty. \quad (5.32)$$

From Assumption 4.2 it follows that

$$\mathbf{S}^e(\mathbf{x}_p^e) \rightarrow \mathbf{S}^e(\mathbf{x}^e) \text{ in } \mathbb{R}^{\nu_1 \times \nu_1} \text{ and } \mathbf{S}_B^e(\mathbf{x}_p^e) \rightarrow \mathbf{S}_B^e(\mathbf{x}^e) \text{ in } \mathbb{R}^{\nu_2 \times \nu_2}, \text{ as } p \rightarrow \infty. \quad (5.33)$$

Now, the only symbols in the integrals (5.31) that depend on the integration variable $\widehat{\mathbf{x}}$ are $\widehat{\boldsymbol{\xi}}_k^r(\widehat{\mathbf{x}})$, $\widehat{\boldsymbol{\xi}}_l^r(\widehat{\mathbf{x}})$, $\mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_k^r(\widehat{\mathbf{x}}))$, and $\mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_l^r(\widehat{\mathbf{x}}))$. However, they are each independent from the vector \mathbf{x}^e . We can expand the matrix multiplications in the integrands, which leads to the following finite linear combinations of integrals

$$\begin{aligned} a_{\varepsilon, \mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)) &= \sum_{i=1}^N c_{\varepsilon, i}^e(\mathbf{x}^e) \int_{K^r} F_i(\widehat{\mathbf{x}}) d\widehat{\mathbf{x}}, \\ f^{v,e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)) &= \sum_{j=1}^M d_j^{v,e}(\mathbf{x}^e) \int_{K^r} G_j(\widehat{\mathbf{x}}) d\widehat{\mathbf{x}}. \end{aligned}$$

From (5.32) and (5.33) for $i = 1, \dots, N$, $j = 1, \dots, M$ it follows that

$$c_{\varepsilon, i}^e(\mathbf{x}_p^e) \rightarrow c_{\varepsilon, i}^e(\mathbf{x}^e) \text{ and } d_j^{v,e}(\mathbf{x}_p^e) \rightarrow d_j^{v,e}(\mathbf{x}^e) \text{ in } \mathbb{R}, \text{ as } p \rightarrow \infty,$$

which consequently yields

$$a_{\varepsilon, \mathbf{x}_p^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}_p^e), \boldsymbol{\xi}_l^e(\mathbf{x}_p^e)) \rightarrow a_{\varepsilon, \mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)) \text{ in } \mathbb{R}, \text{ as } p \rightarrow \infty$$

and

$$f^{v,e}(\boldsymbol{\xi}_k^e(\mathbf{x}_p^e)) \rightarrow f^{v,e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)) \text{ in } \mathbb{R}, \text{ as } p \rightarrow \infty$$

for $k, l = 1, \dots, n^e$. By Assumption 5.3 the topology of the mesh $\mathcal{T}^h(\alpha^h)$ does not change with any $\alpha^h \in \mathcal{U}^h$, hence, the sets E_i^h and E_j^h in (5.30) remain unchanged. It follows that

$$\left(\mathbf{A}_\varepsilon^n(\mathbf{x}^h(\alpha_p^h)) \right)_{i,j} \rightarrow \left(\mathbf{A}_\varepsilon^n(\mathbf{x}^h(\alpha^h)) \right)_{i,j}, \text{ as } p \rightarrow \infty,$$

and

$$\left(\mathbf{f}^{v,n} \left(\mathbf{x}^h \left(\alpha_p^h \right) \right) \right)_i \rightarrow \left(\mathbf{f}^{v,n} \left(\mathbf{x}^h \left(\alpha^h \right) \right) \right)_i \text{ in } \mathbb{R}, \text{ as } p \rightarrow \infty,$$

for each $i, j = 1, \dots, n$. From here and (5.29) we get

$$\mathbf{u}_\varepsilon^{v,n} \left(\mathbf{x}^h \left(\alpha_p^h \right) \right) \rightarrow \mathbf{u}_\varepsilon^{v,n} \left(\mathbf{x}^h \left(\alpha^h \right) \right) \text{ in } \mathbb{R}^n, \text{ as } p \rightarrow \infty. \quad (5.34)$$

Using the map from the reference element r , the global shape function reads as follows:

$$\xi_i^h \left(\mathbf{x}^h \right) = \sum_{e \in E_i^h} \sum_{j: \mathcal{G}^e(j)=i} \xi_j^e \left(\mathbf{x}^e \right) = \sum_{e \in E_i^h} \sum_{j: \mathcal{G}^e(j)=i} \mathbf{S}^e \left(\mathbf{x}^e \right) \cdot \widehat{\xi}_j^r,$$

which together with (5.33) and with Assumption 5.3 yield

$$\xi_i^h \left(\mathbf{x}^h \left(\alpha_p^h \right) \right) \rightarrow \xi_i^h \left(\mathbf{x}^h \left(\alpha^h \right) \right) \text{ in } \mathbb{R}^n, \text{ as } p \rightarrow \infty. \quad (5.35)$$

Combining (5.28), (5.34), and (5.35), we have completed the proof, i.e.,

$$\mathbf{u}_\varepsilon^{v,h} \left(\alpha_p^h \right) \rightarrow \mathbf{u}_\varepsilon^{v,h} \left(\alpha^h \right) \text{ in } \mathcal{U}^h, \text{ as } p \rightarrow \infty.$$

□

Lemma 5.11. *Let $\varepsilon > 0$ be a regularization parameter. Let $\{h_n\}_{n=1}^\infty \subset \mathbb{R}$ be a sequence of positive discretization parameters such that $h_n \rightarrow 0_+$, as $n \rightarrow \infty$. Let $\{\Omega^{h_n}\}_{n=1}^\infty$, $\Omega^{h_n} \subset \Omega$, be a sequence of subdomains satisfying $\Omega^{h_n} \nearrow \Omega$, as $n \rightarrow \infty$. Further, let $\alpha \in \mathcal{U}$ be a shape and $\{\alpha^{h_n}\}_{n=1}^\infty \subset \mathcal{U}$, $\alpha^{h_n} \in \mathcal{U}^{h_n}$, be a sequence of discretized shapes such that*

$$\alpha^{h_n} \rightarrow \alpha \text{ in } \mathcal{U}, \text{ as } n \rightarrow \infty.$$

Then for each $v = 1, \dots, n_v$

$$\mathbf{X}_{\nu_1}^{h_n} \left(\mathbf{u}_\varepsilon^{v,h_n} \left(\alpha^{h_n} \right) \right) \rightarrow \mathbf{u}_\varepsilon^v \left(\alpha \right) \text{ in } \mathbf{H}_0 \left(\mathbf{B}; \Omega \right), \text{ as } n \rightarrow \infty,$$

where $\mathbf{u}_\varepsilon^{v,h_n} \left(\alpha^{h_n} \right)$ is the solution to $(W_\varepsilon^{v,h_n} \left(\alpha^{h_n} \right))$ and $\mathbf{u}_\varepsilon^v \left(\alpha \right)$ is the solution to $(W_\varepsilon^v \left(\alpha \right))$.

Proof. It is enough to prove that the assumption (4.42) is fulfilled and the rest is then fairly the same as the proof of Theorem 4.2, while (4.43) is replaced by the assumption (5.27).

Given an arbitrary $\mathbf{x} \in \Omega \setminus (\partial\Omega_0(\alpha) \cup \partial\Omega_1(\alpha))$, it follows from (5.6) that either $\mathbf{x} \in \Omega_0(\alpha)$ or $\mathbf{x} \in \Omega_1(\alpha)$. Thus, by (5.7) either

$$\mathbf{D}_\alpha(\mathbf{x}) = \mathbf{D}_0 \text{ or } \mathbf{D}_\alpha(\mathbf{x}) = \mathbf{D}_1,$$

respectively. Having $\Omega^{h_n} \nearrow \Omega$, $\alpha^{h_n} \rightrightarrows \alpha$, as $n \rightarrow \infty$, and due to (5.23), there exists $n_0(\mathbf{x}) \in \mathbb{N}$ such that for each $n \in \mathbb{N}$, $n \geq n_0(\mathbf{x})$ either $\mathbf{x} \in \Omega_0^{h_n}(\alpha^{h_n})$ or $\mathbf{x} \in \Omega_1^{h_n}(\alpha^{h_n})$, respectively. Therefore, either

$$\mathbf{D}_{\alpha^{h_n}}(\mathbf{x}) = \mathbf{D}_0 \text{ or } \mathbf{D}_{\alpha^{h_n}}(\mathbf{x}) = \mathbf{D}_1,$$

respectively. Thus, we have verified the assumption (4.42), i.e., for any $i, j = 1, \dots, \nu_2$:

$$\left| d_{\alpha^{h_n}, i, j}^{h_n}(\mathbf{x}) - d_{\alpha, i, j}(\mathbf{x}) \right| \rightarrow 0 \text{ a.e. in } \Omega, \text{ as } n \rightarrow \infty,$$

where $\mathbf{D}_{\alpha^{h_n}}(\mathbf{x}) := \left(d_{\alpha^{h_n}, i, j}^{h_n}(\mathbf{x}) \right)_{i, j} \in \mathbb{R}^{\nu_2 \times \nu_2}$ and $\mathbf{D}_\alpha(\mathbf{x}) := \left(d_{\alpha, i, j}(\mathbf{x}) \right)_{i, j} \in \mathbb{R}^{\nu_2 \times \nu_2}$. □

5.4.3 Discretized optimization problem

The regularized and discretized cost functional is

$$\mathcal{J}_\varepsilon^h(\alpha^h) := \mathcal{I}\left(\alpha^h, \mathbf{B}\left(\mathbf{X}_{\nu_1}^h\left(\mathbf{u}_\varepsilon^{1,h}(\alpha^h)\right)\right), \dots, \mathbf{B}\left(\mathbf{X}_{\nu_1}^h\left(\mathbf{u}_\varepsilon^{n_\nu,h}(\alpha^h)\right)\right)\right), \quad \alpha^h \in \mathcal{U}^h, \quad (5.36)$$

where $\mathbf{X}_{\nu_1}^h : \mathbf{H}_0(\mathbf{B}; \Omega^h)^h \mapsto \mathbf{H}_0(\mathbf{B}; \Omega)$ is due to (4.33) and Lemma 4.3. The relevant setting of the shape optimization problem reads as follows:

$$\left. \begin{array}{l} \text{Find } \alpha_\varepsilon^{h*} \in \mathcal{U}^h: \\ \mathcal{J}_\varepsilon^h(\alpha_\varepsilon^{h*}) \leq \mathcal{J}_\varepsilon^h(\alpha^h) \quad \forall \alpha^h \in \mathcal{U}^h \end{array} \right\} \quad (P_\varepsilon^h)$$

Theorem 5.5. *Let $\varepsilon > 0$ and $h > 0$. Then there exists $\alpha_\varepsilon^{h*} \in \mathcal{U}^h$ that is a solution to (P_ε^h) .*

Proof. Taking any $\varepsilon > 0$ and $h > 0$, the proof is fairly the same as the one of Theorem 5.2, where we use the symbol $\mathcal{J}_\varepsilon^h$ instead of \mathcal{J} , and Lemma 5.10 instead of Lemma 5.3. \square

Theorem 5.6. *Let $\varepsilon > 0$ be a fixed regularization parameter. Let $\{h_n\}_{n=1}^\infty \subset \mathbb{R}$ be a sequence of positive discretization parameters such that $h_n \rightarrow 0_+$, as $n \rightarrow \infty$, and let $\alpha_\varepsilon^{h_n*} \in \mathcal{U}^{h_n}$ denote the corresponding solutions to the problems $(P_\varepsilon^{h_n})$. Then there exist a subsequence $\{h_{n_k}\}_{k=1}^\infty \subset \{h_n\}_{n=1}^\infty$ and a shape $\alpha_\varepsilon^* \in \mathcal{U}$ such that*

$$\alpha_\varepsilon^{h_{n_k}*} \rightarrow \alpha_\varepsilon^* \text{ in } \mathcal{U}, \text{ as } k \rightarrow \infty,$$

holds and, moreover, α_ε^ is a solution to the problem (P_ε) .*

Proof. By Theorem 5.5, for each $\varepsilon > 0$ and $h_n > 0$ there exists $\alpha_\varepsilon^{h_n*} \in \mathcal{U}^{h_n}$, a solution to $(P_\varepsilon^{h_n})$. By Lemma 5.1, there exist a subsequence of shapes $\{\alpha_\varepsilon^{h_{n_k}*}\}_{k=1}^\infty \subset \{\alpha_\varepsilon^{h_n*}\}_{n=1}^\infty$ and a shape $\alpha_\varepsilon^* \in \mathcal{U}$ such that

$$\alpha_\varepsilon^{h_{n_k}*} \rightarrow \alpha_\varepsilon^* \text{ in } \mathcal{U}, \text{ as } k \rightarrow \infty. \quad (5.37)$$

Let $\alpha \in \mathcal{U}$ be an arbitrary shape. By Lemma 5.8, there exists a sequence $\{\alpha^{h_{n_k}}\}_{k=1}^\infty, \alpha^{h_{n_k}} \in \mathcal{U}^{h_{n_k}}$ such that

$$\alpha^{h_{n_k}} \rightarrow \alpha \text{ in } \mathcal{U}, \text{ as } k \rightarrow \infty. \quad (5.38)$$

Then, due to the definition of $(P_\varepsilon^{h_{n_k}})$ for any $k \in \mathbb{N}$ we have

$$\mathcal{J}_\varepsilon^{h_{n_k}}(\alpha_\varepsilon^{h_{n_k}*}) \leq \mathcal{J}_\varepsilon^{h_{n_k}}(\alpha^{h_{n_k}}). \quad (5.39)$$

Using (5.37), (5.38), Lemma 5.11, and the continuity of \mathcal{I} , both the left and right-hand side of (5.39) converge

$$\mathcal{J}_\varepsilon^{h_{n_k}}(\alpha_\varepsilon^{h_{n_k}*}) \rightarrow \mathcal{J}(\alpha_\varepsilon^*) \text{ and } \mathcal{J}_\varepsilon^{h_{n_k}}(\alpha^{h_{n_k}}) \rightarrow \mathcal{J}(\alpha), \text{ as } k \rightarrow \infty.$$

Therefore, we have proven that for any $\alpha \in \mathcal{U}$

$$\mathcal{J}(\alpha_\varepsilon^*) \leq \mathcal{J}(\alpha).$$

\square

Finally, we introduce the regularized and discretized cost functional $\tilde{\mathcal{J}}_\varepsilon^h : \Upsilon \mapsto \mathbb{R}$ by

$$\tilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}) := \mathcal{J}_\varepsilon^h\left(\pi_\omega^h(F(\mathbf{p}))\right), \quad \mathbf{p} \in \Upsilon.$$

Then, the regularized and discretized optimization problem reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{p}_\varepsilon^{h*} \in \Upsilon: \\ \tilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}_\varepsilon^{h*}) \leq \tilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}) \quad \forall \mathbf{p} \in \Upsilon \end{array} \right\} \quad (\tilde{P}_\varepsilon^h)$$

Since Υ is a compact set and $\pi_\omega^h \circ F : \Upsilon \mapsto \mathcal{U}^h$ is a continuous mapping, we can state and prove the existence theorem for $(\tilde{P}_\varepsilon^h)$ similarly to Theorem 5.5. We can also state the convergence theorem, the proof of which is even simpler than the one of Theorem 5.6, as the set of admissible design parameters Υ is not changed by discretization.

Remark 5.1. *In cases of complex geometries, as those in Chapter 7, Assumption 5.3 is a serious bottleneck of this discretization approach. For small discretization parameters and large changes in the design we cannot guarantee that the perturbed elements still satisfy some regularity condition. They might be even flipped. In this case, we have to re-mesh the geometry and solve the optimization problem again, but now on a grid of different topology. Then certainly the cost functional is not continuous any more and the just introduced convergence theory cannot be applied. Nevertheless, in literature this approach is still the most frequently used one as far as a finite element discretization is concerned. In practice, after we get an optimized shape we should compare the value of a very fine discretized cost functional for the optimized design with that value for the initial one. If we can see a progress then the optimization surely did a good job. Some solutions to this obstacle are discussed in Conclusion.*

Chapter 6

Numerical methods for shape optimization

In this chapter, which is the heart of the thesis, we will focus on the Newton–type algorithms for smooth discretized shape optimization problems. At the beginning, we will revise the discretized setting and we will show its smoothness, i.e., the continuity of both the cost and constraint functions up to their second derivatives. Further, we will recall a quasi–Newton algorithm that uses the first–derivatives only. Then, we will discuss the first–order sensitivity analysis methods. We will derive a robust, but still efficient, algorithm based on the algebraic approach of the first–order shape sensitivity analysis, and we will implement it into an object–oriented software framework. At the end, we will introduce a multilevel optimization approach.

An extensive literature on the gradient– or Newton–type optimization methods has been written. Let us refer to NOCEDAL AND WRIGHT [148], FLETCHER AND REEVES [63], FLETCHER [61, 62], SVANBERG [203, 204], DENNIS AND SCHNABEL [55], GILL, MURRAY, AND WRIGHT [66], GROSSMANN AND TERNO [72], CEA [39], CLARKE [48], HAGER, HEARN, AND PARDALOS [81], HESTENSEN [88, 89], POLAK [161], POLAK AND RIBIÈRE [162], BOGGS AND TOLLE [23], CONN, GOULD, AND TOINT [49], MÄKELÄ AND NEITTAANMÄKI [130], ZOWE, KOČVARA, AND BENDSØE [222]. There are many essential monographs and papers dealing with the sensitivity analysis in shape optimization. Let us mention HAUG, CHOI, AND KOMKOV [86], HASLINGER AND NEITTAANMÄKI [85], HASLINGER AND MÄKINEN [83], ZOLESIO [220], SOKOŁOWSKI AND ZOLESIO [196], SOKOŁOWSKI AND ZOCHOWSKI [195], PETERSSON [157], SIMON [193], LAPORTE AND TALLEC [118], DELFOUR AND ZOLESIO [54], HANSEN, ZIU, AND OLHOFF [82], BROCKMAN [34], GRIEWANK [70], MÄKINEN [131], NEITTAANMÄKI AND SALMENJOKI [145].

6.1 The discretized optimization problem revisited

Throughout this chapter, we will consider the optimization problem $(\tilde{P}_\varepsilon^h)$, introduced in Chapter 5. Recall that for a given $\varepsilon > 0$ and $h > 0$ the problem reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{p}_\varepsilon^{h*} \in \Upsilon: \\ \tilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}_\varepsilon^{h*}) \leq \tilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}) \quad \forall \mathbf{p} \in \Upsilon \end{array} \right\}, \quad (\tilde{P}_\varepsilon^h)$$

where $\tilde{\mathcal{J}}_\varepsilon^h : \Upsilon \mapsto \mathbb{R}$ denotes the discretized and regularized cost functional and $\Upsilon \subset \mathbb{R}^{n_\Upsilon}$, $n_\Upsilon \in \mathbb{N}$, is the set of admissible design parameters.

6.1.1 Constraint function

Let us rewrite the admissible set as follows:

$$\Upsilon := \{\mathbf{p} \in \mathbb{R}^{n_\Upsilon} \mid \mathbf{v}(\mathbf{p}) \leq \mathbf{0}\},$$

where $\mathbf{v} : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_v}$, $n_v \in \mathbb{N}$, is the *constraint function*.

Assumption 6.1. We assume that $\mathbf{v} \in [C^2(\mathbb{R}^{n_\Upsilon})]^{n_v}$.

6.1.2 Design-to-shape mapping

Let us take a deep look into the structure of the cost functional in order to prove its *smoothness* and to derive the first-order derivatives with respect to the design variables. First, we define the shape parameterization function $\boldsymbol{\alpha}^h := (\alpha_1^h, \dots, \alpha_{n_{\alpha^h}}^h) : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_{\alpha^h}}$, $n_{\alpha^h} := n_{\mathbf{x}_\omega^h} \in \mathbb{N}$, by

$$\alpha_i^h(\mathbf{p}) := (\pi_\omega^h \circ F)(\mathbf{p}) \equiv [F(\mathbf{p})](\mathbf{x}_{\omega,i}^h) \quad \text{for } i = 1, \dots, n_{\alpha^h}, \quad (6.1)$$

where $F : \Upsilon \mapsto \mathcal{U}$ is due to (5.5) and $\pi_\omega^h : \mathcal{U} \mapsto P^1(\mathcal{T}_\omega^h)$ is defined by (5.22).

6.1.3 Shape-to-mesh mapping

Here, we revisit the dependence of the discretization grid nodes on the shape control nodes, i.e., we introduce a function $\mathbf{x}^h : \mathbb{R}^{n_{\alpha^h}} \mapsto \mathbb{R}^{mn_{\mathbf{x}^h}}$ the components of which correspond to the grid nodal coordinates (4.11), where $n_{\mathbf{x}^h} \in \mathbb{N}$ denotes the number of nodes in the discretization $\mathcal{T}^h(\boldsymbol{\alpha}^h)$ of the domain Ω^h . The function $\mathbf{x}^h(\cdot)$ maps the control shape coordinates onto the remaining grid nodal coordinates by means of solving an auxiliary discretized linear elasticity problem in terms of grid displacements $\Delta \mathbf{x}^h$ with a nonhomogeneous Dirichlet boundary condition that corresponds to given shape displacements $\boldsymbol{\alpha}^h$, and with prescribed zero displacements on $\partial\Omega^h$ and on such inner interfaces that are not allowed to move. The zero displacements are, for example, prescribed along the boundary of a subdomain containing nonzero sources \mathbf{f}^h . The shape-to-mesh mapping is as follows:

$$\mathbf{x}^h(\boldsymbol{\alpha}^h) := \mathbf{x}_0^h + \Delta \mathbf{x}^h(\boldsymbol{\alpha}^h) + \mathcal{M}^h \cdot \boldsymbol{\alpha}^h, \quad \text{where } \mathbf{K}^h(\mathbf{x}_0^h) \cdot \Delta \mathbf{x}^h(\boldsymbol{\alpha}^h) = \mathbf{b}^h(\boldsymbol{\alpha}^h), \quad (6.2)$$

in which the vector $\mathbf{x}_0^h \in \mathbb{R}^{mn_{\mathbf{x}^h}}$ contains the initial grid nodal coordinates and is independent of $\boldsymbol{\alpha}^h$, where further $\mathbf{K}^h(\mathbf{x}_0^h) \in \mathbb{R}^{(mn_{\mathbf{x}^h}) \times (mn_{\mathbf{x}^h})}$ is a nonsingular symmetric stiffness matrix, $\mathbf{b}^h(\boldsymbol{\alpha}^h) \in \mathbb{R}^{mn_{\mathbf{x}^h}}$ is a right-hand side vector linearly dependent on $\boldsymbol{\alpha}^h \in \mathbb{R}^{n_{\alpha^h}}$, and where finally $\mathcal{M}^h \in \mathbb{R}^{(mn_{\mathbf{x}^h}) \times n_{\alpha^h}}$ is a rectangular permutation matrix that identically maps the shape nodal coordinates onto the corresponding grid nodal coordinates. Both \mathbf{K}^h and \mathbf{b}^h arise from the finite element discretization of the auxiliary linear elasticity problem. For finite elements in elasticity, we refer to ZIENKIEWICZ [217]. The matrix \mathcal{M}^h might also involve some symmetry assumptions on the geometry, as we will state later in Chapter 7. Solving the equation (6.2) takes approximately the same computational effort as solving one state problem. Nevertheless, the mapping is very general, which fits to our intent in developing a robust and efficient numerical method for shape optimization.

6.1.4 Multistate problem

Concerning the multistate problem, we recall that we arrive at solving the following n_v linear systems of algebraic equations

$$\mathbf{A}_\varepsilon^n(\mathbf{x}^h) \cdot \mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h) = \mathbf{f}^{v,n}(\mathbf{x}^h), \quad v = 1, \dots, n_v, \quad (6.3)$$

where both the system matrix and the right-hand side vectors are assembled by means of Algorithm 1

$$\begin{aligned} \left(\mathbf{A}_\varepsilon^n(\mathbf{x}^h) \right)_{i,j} &:= \sum_{e \in E_i^h \cap E_j^h} \sum_{k,l=1}^{n_e} a_{\varepsilon, \mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)), \\ \left(\mathbf{f}^{v,n}(\mathbf{x}^h) \right)_i &:= \sum_{e \in E_i^h} \sum_{k=1}^{n_e} f^{v,e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)) \end{aligned} \quad (6.4)$$

for $i, j = 1, \dots, n$, where E_i^h denotes the set of elements neighbouring with e_i , see (4.20), and $\mathbf{x}^e \in \mathbb{R}^{m(m+1)}$ is the vector of coordinates of the element domain corners, see (4.25), which is also included in \mathbf{x}^h by means of the mapping \mathcal{H}^e , see (4.26). Components of the solution to (6.3) are denoted by

$$\mathbf{u}_\varepsilon^{v,n} := \left(u_{\varepsilon,1}^{v,n}, \dots, u_{\varepsilon,n}^{v,n} \right) \in \mathbb{R}^n, \quad v = 1, \dots, n_v.$$

Using the map from the reference element r , the element contributions to the bilinear form and linear functional, respectively, see also (4.29) and (4.30), are

$$\begin{aligned} a_{\varepsilon, \mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)) &:= \\ &:= \int_{K^r} \left(\mathbf{S}_B^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_k^r(\widehat{\mathbf{x}})) \right) \cdot \left(\mathbf{D}^e \cdot \left(\mathbf{S}_B^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_l^r(\widehat{\mathbf{x}})) \right) \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}} + \\ &+ \varepsilon \int_{K^r} \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_k^r(\widehat{\mathbf{x}}) \right) \cdot \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_l^r(\widehat{\mathbf{x}}) \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}}, \\ f^{v,e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)) &:= \int_{K^r} \mathbf{f}^{v,e} \cdot \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_k^r(\widehat{\mathbf{x}}) \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}}, \end{aligned} \quad (6.5)$$

where $k, l = 1, \dots, n_e$ and where we consider (5.26) and Assumption 5.4. Then,

$$\mathbf{u}_\varepsilon^{v,h}(\mathbf{x}^h; \mathbf{x}) = \sum_{i=1}^n u_{\varepsilon,i}^{v,n}(\mathbf{x}^h) \boldsymbol{\xi}_i^h(\mathbf{x}^h; \mathbf{x}), \quad v = 1, \dots, n_v, \quad \mathbf{x}^h \in \mathbb{R}^{mn_{\mathbf{x}^h}}, \quad \mathbf{x} \in \overline{\Omega^h}, \quad (6.6)$$

is the solution to the state problem $(W_\varepsilon^{v,h}(\alpha^h))$, where $\boldsymbol{\xi}_i^h(\mathbf{x}^h; \mathbf{x})$ denote the global shape functions. Moreover, for $e \in E^h$ we introduce the element solution vector by

$$\mathbf{u}_\varepsilon^{v,n,e} := \left(u_{\varepsilon,1}^{v,n,e}, \dots, u_{\varepsilon,n_e}^{v,n,e} \right) \in \mathbb{R}^{n_e}, \quad \text{where} \quad u_{\varepsilon,i}^{v,n,e} := u_{\varepsilon, \mathcal{G}^e(i)}^{v,n} \quad \text{for } i = 1, \dots, n_e.$$

As we look for $\mathbf{B}(\mathbf{u}_\varepsilon^{v,h})$ rather than for $\mathbf{u}_\varepsilon^{v,h}$, we further elementwise evaluate the following block column vector

$$\mathbf{B}_\varepsilon^{v,n}(\mathbf{x}^h) := \left[\mathbf{B}_\varepsilon^{v,n,e_1}(\mathbf{x}^h), \dots, \mathbf{B}_\varepsilon^{v,n,e_{n_{\Omega^h}}}(\mathbf{x}^h) \right] := \mathbf{B}(\mathbf{x}^h, \mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h)) \in \mathbb{R}^{\nu_{2n_{\Omega^h}}} \quad (6.7)$$

for $v = 1, \dots, n_v$, where for $i = 1, \dots, n_{\Omega^h}$ the corresponding element vector is defined by

$$\mathbf{B}_\varepsilon^{v,n,e_i}(\mathbf{x}^h) := \mathbf{B}(\mathbf{x}^{e_i}, \mathbf{u}_\varepsilon^{v,n,e_i}(\mathbf{x}^h)),$$

and where $\mathbf{x}^h \in \mathbb{R}^{mn_{\mathbf{x}^h}}$ contains all the grid nodes, $\mathbf{x}^e \in \mathbb{R}^{m_{n_e}}$ contains the grid nodes related to the element $e \in E^h$, where further

$$\begin{aligned} \mathbf{B}(\mathbf{x}^h, \mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h))|_{K_i} &:= \mathbf{B}(\mathbf{x}^{e_i}, \mathbf{u}_\varepsilon^{v,n,e_i}(\mathbf{x}^h)) := \mathbf{B}_\mathbf{x}(\mathbf{u}_\varepsilon^{v,h}(\mathbf{x}^h; \mathbf{x}))|_{K_i} = \\ &= \sum_{j=1}^{n_e} u_{\varepsilon, \mathcal{G}^{e_i(j)}}^{v,n}(\mathbf{x}^{e_i}) \mathbf{B}_\mathbf{x}(\boldsymbol{\xi}_j^{e_i}(\mathbf{x}^{e_i}; \mathbf{x})) = \\ &= \sum_{j=1}^{n_e} u_{\varepsilon, j}^{v,n,e_i}(\mathbf{x}^{e_i}) \mathbf{S}_\mathbf{B}^{e_i}(\mathbf{x}^{e_i}) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j(\widehat{\mathbf{x}})) \quad \text{for } i = 1, \dots, n_{\Omega^h}, \end{aligned} \quad (6.8)$$

where $\mathbf{x} := \mathbf{R}^{e_i} \cdot \widehat{\mathbf{x}} + \mathbf{r}^{e_i} \in \overline{K_i}$, and where $e_i \in E^h$ is the element related to K_i . Recall that since we employ the lowest, i.e., first-order finite elements, the function $\mathbf{B}_\mathbf{x}(\mathbf{u}_\varepsilon^{v,h}(\mathbf{x}^h; \mathbf{x}))$ is elementwise constant.

6.1.5 Cost functional

Now, we revisit the cost functional (5.36) from the algebraic point of view. In addition, it depends on the vector of grid nodal coordinates \mathbf{x}^h as follows:

$$\widetilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}) := \mathcal{I}^h(\boldsymbol{\alpha}^h(\mathbf{p}), \mathbf{x}^h(\boldsymbol{\alpha}^h), \mathbf{B}_\varepsilon^{1,n}(\mathbf{x}^h), \dots, \mathbf{B}_\varepsilon^{n_v,n}(\mathbf{x}^h)),$$

where $\boldsymbol{\alpha}^h := \boldsymbol{\alpha}^h(\mathbf{p})$ is a vector of shape control coordinates, where for $v = 1, \dots, n_v$ $\mathbf{B}_\varepsilon^{v,n}(\mathbf{x}^h) \in \mathbb{R}^{\nu_2 n_{\Omega^h}}$ is given by (6.7) and (6.8), and where $\mathcal{I}^h : \mathbb{R}^{n_{\boldsymbol{\alpha}^h}} \times \mathbb{R}^{mn_{\mathbf{x}^h}} \times [\mathbb{R}^{\nu_2 n_{\Omega^h}}]^{n_v} \mapsto \mathbb{R}$ is the revised cost functional which is for $\mathbf{p} \in \Upsilon$ and for $\mathbf{x}^h := \mathbf{x}^h(\boldsymbol{\alpha}^h(\mathbf{p}))$ defined by

$$\begin{aligned} \mathcal{I}^h(\boldsymbol{\alpha}^h(\mathbf{p}), \mathbf{x}^h, \mathbf{B}_\varepsilon^{1,n}(\mathbf{x}^h), \dots, \mathbf{B}_\varepsilon^{n_v,n}(\mathbf{x}^h)) &:= \\ &:= \mathcal{I}(\pi_\omega^h(F(\mathbf{p})), \mathbf{X}_{\nu_1}^h(\mathbf{B}(\mathbf{u}_\varepsilon^{1,h})), \dots, \mathbf{X}_{\nu_1}^h(\mathbf{B}(\mathbf{u}_\varepsilon^{n_v,h}))), \end{aligned}$$

in which $\pi_\omega^h : \mathcal{U} \mapsto \mathcal{U}^h$ is defined by (5.22), $F : \Upsilon \mapsto \mathcal{U}$ is due to (5.5), $\mathbf{X}_{\nu_1}^h : \mathbf{H}_0(\mathbf{B}; \Omega^h) \mapsto \mathbf{H}_0(\mathbf{B}; \Omega)$ is due to (4.33) and Lemma 4.3, and where $\mathbf{u}_\varepsilon^{v,h}$ is the solution (6.6).

The complete evaluation of the cost functional proceeds as follows:

$$\begin{aligned} \mathcal{P} \xrightarrow{\pi_\omega^h \circ F} \boldsymbol{\alpha}^h \xrightarrow{\mathbf{K}^h \cdot \Delta \mathbf{x}^h = \mathbf{b}^h(\boldsymbol{\alpha}^h)} \mathbf{x}^h \xrightarrow{\text{FEM}} \mathbf{A}_\varepsilon^n, \mathbf{f}^{v,n} \xrightarrow{\mathbf{A}_\varepsilon^n \cdot \mathbf{u}_\varepsilon^{v,n} = \mathbf{f}^{v,n}} \mathbf{u}_\varepsilon^{v,n} \xrightarrow{\mathbf{B}(\mathbf{x}^h, \mathbf{u}_\varepsilon^{v,n})} \\ \xrightarrow{\mathbf{B}(\mathbf{x}^h, \mathbf{u}_\varepsilon^{v,n})} \mathbf{B}_\varepsilon^{v,n} \xrightarrow{\mathcal{I}^h(\boldsymbol{\alpha}^h, \mathbf{x}^h, \mathbf{B}_\varepsilon^{1,n}, \dots, \mathbf{B}_\varepsilon^{n_v,n})} \widetilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}). \end{aligned} \quad (6.9)$$

The cost functional $\widetilde{\mathcal{J}}_\varepsilon^h$ is compounded of the following submappings:

- $\boldsymbol{\alpha}^h$ which is the discretized shape parameterization,
- $\mathbf{K}^h \cdot \Delta \mathbf{x}^h = \mathbf{b}^h(\boldsymbol{\alpha}^h)$, see (6.2), that maps the shape control nodal coordinates $\boldsymbol{\alpha}^h$ onto the remaining nodal coordinates \mathbf{x}^h in the grid,

- FEM which assembles the system matrix A_ε^n and the right-hand side vectors $\mathbf{f}^{1,n}, \dots, \mathbf{f}^{n_v,n}$ by means of the finite element method, as described in Algorithm 1,
- $A_\varepsilon^n \cdot \mathbf{u}_\varepsilon^{v,n} = \mathbf{f}^{v,n}$ that solve the n_v linear systems of algebraic equations,
- $\mathbf{B}_\varepsilon^{v,n}$ which is a block column vector whose individual vectors represent the elementwise constant functions $\mathbf{B}_x(\mathbf{u}_\varepsilon^{v,h}(\mathbf{x}))$, see in Algorithm 2, and
- \mathcal{I}^h which calculates the cost functional.

6.1.6 Smoothness of the cost functional

To prove the smoothness of $\tilde{\mathcal{J}}_\varepsilon^h$, we need the smoothness of all the submappings.

Assumption 6.2. We assume that for each $h > 0$ such that $h \leq \bar{h}$ the following hold:

- $\forall \mathbf{x} \in \bar{\omega} : [F(\cdot)](\mathbf{x}) \in C^2(\Upsilon)$,
- $\mathbf{K}^h \in \mathbb{R}^{(mn_{\mathbf{x}^h}) \times (mn_{\mathbf{x}^h})}$ is nonsingular,
- $\mathbf{b}^h(\boldsymbol{\alpha}^h) \in [C^2(\mathbb{R}^{n_{\boldsymbol{\alpha}^h})}]^{mn_{\mathbf{x}^h}}$,
- $\forall e \in E^h : \mathbf{R}^e(\mathbf{x}^e) \in [C^2(\mathbb{R}^{m(m+1)})]^{m \times m}$,
- $\forall e \in E^h : \mathbf{S}_B^e(\mathbf{x}^e) \in [C^2(\mathbb{R}^{m(m+1)})]^{\nu_2 \times \nu_2}$,
- $\forall e \in E^h : \mathbf{S}^e(\mathbf{x}^e) \in [C^2(\mathbb{R}^{m(m+1)})]^{\nu_1 \times \nu_1}$, and
- $\mathcal{I}^h(\boldsymbol{\alpha}^h, \mathbf{B}_\varepsilon^{1,n}, \dots, \mathbf{B}_\varepsilon^{n_v,n}) \in C^2(\mathbb{R}^{n_{\boldsymbol{\alpha}^h}} \times (\mathbb{R}^{\nu_2 n_{\Omega^h}})^{n_v})$.

Lemma 6.1. Under Assumptions 5.1, 5.3, 5.4, and 6.2, for any $h > 0$ such that $h \leq \bar{h}$ it holds that

$$\tilde{\mathcal{J}}_\varepsilon^h \in C^2(\Upsilon).$$

Proof. We will step-by-step use Assumption 6.2 and apply Lemma 3.4 to prove the smoothness of the individual submappings.

Let $h > 0$ be given such that $h \leq \bar{h}$. By Assumption 6.2, for each $i = 1, \dots, n_{\mathbf{x}_\omega^h}$ we have $[F(\cdot)](\mathbf{x}_{\omega,i}^h) \in C^2(\Upsilon)$, therefore,

$$\boldsymbol{\alpha}^h \in (C^2(\Upsilon))^{n_{\boldsymbol{\alpha}^h}}. \quad (6.10)$$

Again by Assumption 6.2 and by Lemma 3.3, $\det(\mathbf{K}^h) \neq 0$ and

$$[\mathbf{K}^h]^{-1} := \frac{1}{\det(\mathbf{K}^h)} \widetilde{\mathbf{K}^h},$$

holds, where $\widetilde{\mathbf{K}^h}$ denotes the adjoint matrix, which was defined by (3.13). Then, due to the latter and by Assumption 6.2, we get

$$\mathbf{x}^h(\boldsymbol{\alpha}^h) = [\mathbf{K}^h]^{-1} \cdot \mathbf{b}^h(\boldsymbol{\alpha}^h) \in [C^2(\mathbb{R}^{n_{\boldsymbol{\alpha}^h})}]^{mn_{\mathbf{x}^h}}. \quad (6.11)$$

Now, (6.10), (6.11), and Lemma 3.4 yield

$$\mathbf{x}^h \circ \boldsymbol{\alpha}^h \in [C^2(\Upsilon)]^{mn_{\mathbf{x}^h}}.$$

Let us further prove the smoothness of the solutions $\mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h)$ to the discretized multistate problem (6.3). Let $\varepsilon > 0$ and $v = 1, \dots, n_v$ be arbitrary. By Assumption 6.2, for each $e \in E^h$ we get

$$\mathbf{R}^e \in \left[C^2\left(\mathbb{R}^{m(m+1)}\right) \right]^{\widehat{m} \times \widehat{m}}, \quad \mathbf{S}_B^e \in \left[C^2\left(\mathbb{R}^{m(m+1)}\right) \right]^{\nu_2 \times \nu_2}, \quad \mathbf{S}^e \in \left[C^2\left(\mathbb{R}^{m(m+1)}\right) \right]^{\nu_1 \times \nu_1}.$$

Then, also due to the definition (3.12), $\det(\mathbf{R}^e) \in C^2(\mathbb{R}^{m(m+1)})$. From Assumption 5.3 it follows that the element K^e must not flip, so the determinant does not change its sign, i.e.,

$$|\det(\mathbf{R}^e)| \in C^2(\mathbb{R}^{m(m+1)}).$$

Now let us look at the element contributions (6.5). Having Assumptions 5.1 and 5.4, only the reference shape functions $\widehat{\boldsymbol{\xi}}_k^r$ and $\widehat{\boldsymbol{\xi}}_l^r$ depend on the integration variable $\widehat{\mathbf{x}}$. After some matrix–vector multiplications we get the following structure of the element bilinear form and linear functional, respectively,

$$\begin{aligned} a_{\varepsilon, \mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)) &= \sum_{i=1}^N c_{\varepsilon, i}^e(\mathbf{x}^e) \int_{K^r} F_i(\widehat{\mathbf{x}}) d\widehat{\mathbf{x}}, \\ f^{v, e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)) &= \sum_{i=1}^M d_i^{v, e}(\mathbf{x}^e) \int_{K^r} G_i(\widehat{\mathbf{x}}) d\widehat{\mathbf{x}}, \end{aligned}$$

where both $c_{\varepsilon, i}^e(\mathbf{x}^e), d_i^{v, e}(\mathbf{x}^e) \in C^2(\mathbb{R}^{m(m+1)})$, since they arise as summations and multiplications of the entries of $\mathbf{S}_B^e, \mathbf{D}^e, \mathbf{S}^e$, and $\mathbf{f}^{v, e}$ multiplied then by $|\det(\mathbf{R}^e)|$, and where both $F_i(\widehat{\mathbf{x}})$ and $G_i(\widehat{\mathbf{x}})$ are common for all $e \in E^h$. Thus, it follows that both the element bilinear form and element linear functional are smooth, i.e.,

$$a_{\varepsilon, \mathbf{x}^e}^e(\boldsymbol{\xi}_k^e(\mathbf{x}^e), \boldsymbol{\xi}_l^e(\mathbf{x}^e)) \in \left[C^2\left(\mathbb{R}^{m(m+1)}\right) \right]^{n_e \times n_e}, \quad f^{v, e}(\boldsymbol{\xi}_k^e(\mathbf{x}^e)) \in \left[C^2\left(\mathbb{R}^{m(m+1)}\right) \right]^{n_e}. \quad (6.12)$$

Now we employ Assumption 5.3, which assures that the topology of the discretization \mathcal{T}^h is fixed. Hence, neither E_i^h nor E_j^h in (6.3) depends on \mathbf{x}^h . From (6.12) it follows that

$$\mathbf{A}_\varepsilon^n(\mathbf{x}^h) \in \left[C^2(\mathbb{R}^{mn_{\mathbf{x}^h}}) \right]^{n \times n} \quad \text{and} \quad \mathbf{f}^{v, n}(\mathbf{x}^h) \in \left[C^2(\mathbb{R}^{mn_{\mathbf{x}^h}}) \right]^n,$$

a consequence of which is

$$\det\left(\mathbf{A}_\varepsilon^n(\mathbf{x}^h)\right) \in C^2(\mathbb{R}^{mn_{\mathbf{x}^h}}).$$

Lemma 5.9 provides us the existence of the solution $\mathbf{u}_\varepsilon^{v,n}(\mathbf{x}^h)$ to (6.3). Hence, there exist the inverse matrix $[\mathbf{A}_\varepsilon^n(\mathbf{x}^h)]^{-1}$. Then, by Lemma 3.3, $\det(\mathbf{A}_\varepsilon^n(\mathbf{x}^h)) \neq 0$ and

$$\left[\mathbf{A}_\varepsilon^n(\mathbf{x}^h) \right]^{-1} := \frac{1}{\det(\mathbf{A}_\varepsilon^n(\mathbf{x}^h))} \widetilde{\mathbf{A}}_\varepsilon^n(\mathbf{x}^h) \in \left[C^2(\mathbb{R}^{mn_{\mathbf{x}^h}}) \right]^{n \times n}$$

holds, where $\widetilde{\mathbf{A}}_\varepsilon^n(\mathbf{x}^h)$ denotes the adjoint matrix, which was defined by (3.13). From (3.15) we get

$$\mathbf{u}_\varepsilon^{v, n}(\mathbf{x}^h) = \left[\mathbf{A}_\varepsilon^n(\mathbf{x}^h) \right]^{-1} \cdot \mathbf{f}^{v, n}(\mathbf{x}^h) \in \left[C^2(\mathbb{R}^{mn_{\mathbf{x}^h}}) \right]^n \quad \text{for } v = 1, \dots, n_v. \quad (6.13)$$

The symbol $B_\varepsilon^{v,n}$ is calculated by (6.7) and (6.8). Since there appear only summations and multiplications of the components of $\mathbf{u}_\varepsilon^{v,n,e}(\mathbf{x}^h)$ and $\mathbf{S}_B^e(\mathbf{x}^e)$ with the constant vectors $\mathbf{B}_{\hat{\mathbf{x}}}(\hat{\boldsymbol{\xi}}_j^r(\hat{\mathbf{x}}))$, we can use (6.13) and Lemma 3.4 which yield

$$\mathbf{B}_\varepsilon^{v,n}(\mathbf{x}^h) := \mathbf{B}(\mathbf{x}^h, \mathbf{u}^{v,n}(\mathbf{x}^h)) \in [C^2(\mathbb{R}^{mn_{\mathbf{x}^h}})]^{\nu_2 n_{\Omega^h}} \quad \text{for } v = 1, \dots, n_v. \quad (6.14)$$

Finally, we compound the submappings $\boldsymbol{\alpha}^h : \Upsilon \mapsto \mathbb{R}^{n_{\alpha^h}}$, $\mathbf{x}^h : \mathbb{R}^{n_{\alpha^h}} \mapsto \mathbb{R}^{mn_{\mathbf{x}^h}}$, $\mathbf{B}_\varepsilon^{v,n} : \mathbb{R}^{mn_{\mathbf{x}^h}} \mapsto \mathbb{R}^{\nu_2 n_{\Omega^h}}$, and $\mathcal{I}^h : \mathbb{R}^{n_\alpha} \times \mathbb{R}^{mn_{\mathbf{x}^h}} \times (\mathbb{R}^{\nu_2 n_{\Omega^h}})^{n_v} \mapsto \mathbb{R}$. First, Assumption 6.2 yields

$$\mathcal{I}^h \in C^2(\mathbb{R}^{n_{\alpha^h}} \times \mathbb{R}^{mn_{\mathbf{x}^h}} \times (\mathbb{R}^{\nu_2 n_{\Omega^h}})^{n_v}).$$

Then, using the latter, (6.10), (6.14), and applying Lemma 3.4, we have proven the statement

$$\tilde{\mathcal{J}}_\varepsilon^h := \mathcal{I}^h \circ \left[\boldsymbol{\alpha}^h \times (\mathbf{x}^h \circ \boldsymbol{\alpha}^h) \times \left(\mathbf{B}_\varepsilon^{1,n} \circ \mathbf{x}^h \circ \boldsymbol{\alpha}^h \right) \times \dots \times \left(\mathbf{B}_\varepsilon^{n_v,n} \circ \mathbf{x}^h \circ \boldsymbol{\alpha}^h \right) \right] \in C^2(\Upsilon).$$

□

Convention 6.1. *Just for the purposes of this chapter let us skip in our notation the discretization parameter h , the superscript n in (6.3), and the regularization parameter ε that will be each fixed for the moment. If not stated otherwise, all the symbols in the sequel will be considered as discretized ones, even if they were previously reserved for the continuous setting. Hence, we consider the following discrete optimization problem with inequality constraints*

$$\left. \begin{array}{l} \text{Find } \mathbf{p}^* := \arg \min_{\mathbf{p} \in \mathbb{R}^{n_\Upsilon}} \tilde{\mathcal{J}}(\mathbf{p}) \\ \text{subject to } \mathbf{v}(\mathbf{p}) \leq \mathbf{0} \end{array} \right\}, \quad (\mathcal{P})$$

where $\tilde{\mathcal{J}} : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}$ and $\mathbf{v}(\mathbf{p}) := (v_1(\mathbf{p}), \dots, v_{n_v}(\mathbf{p})) : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_v}$. The problem (\mathcal{P}) is governed by the following multistate problem

$$\mathbf{A}(\mathbf{x}) \cdot \mathbf{u}^v(\mathbf{x}) = \mathbf{f}^v(\mathbf{x}) \quad (P^v(\mathbf{x}))$$

for $v = 1, \dots, n_v$.

6.2 Newton-type optimization methods

Since we do not usually have any rigorous analysis locating the global solution \mathbf{p}^* , it is hardly possible to solve the problem (\mathcal{P}) in a suitable computational time and with a suitable precision at the same time, when having only some ten design variables. The algorithms looking for a global minimizer are of an exponential order of complexity with respect to the number n_Υ of design variables. On the other hand, the Newton-type algorithms search for a local minimizer only, but the computational time is quadratically proportional to the distance of the initial design from the closest local minimizer. This is due to the fact that we can precisely provide derivatives of the cost function $\tilde{\mathcal{J}}$ as well as of the constraint function \mathbf{v} with respect to the design variables \mathbf{p} . Here, we will restrict ourselves to developing efficient methods that calculate derivatives for shape optimization. We refer to NOCEDAL AND WRIGHT [148] for a detail overview of optimization methods for a large variety of problems.

6.2.1 Quadratic programming subproblem

The Newton–type algorithms are based on an approximation of the original nonlinear optimization problem by a quadratic or a sequence of quadratic optimization subproblems, which are also referred to as *quadratic programming* subproblems. In general, a quadratic programming problem reads as follows:

$$\left. \begin{array}{l} \text{Find } \mathbf{p}^* := \arg \min_{\mathbf{p} \in \mathbb{R}^{n_\Upsilon}} Q(\mathbf{p}) \\ \text{subject to } \mathcal{L}(\mathbf{p}) \leq \mathbf{0} \end{array} \right\}, \quad (\mathcal{QP})$$

where $Q : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}$ denotes a quadratic function and $\mathcal{L} : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_\nu}$ denotes a linear vector function.

Basically, there are two approaches to the approximation of the problem (\mathcal{P}) by a subproblem (\mathcal{QP}) . In both of them the input $\mathbf{p}_0 \in \mathbb{R}^{n_\Upsilon}$ denotes initial design parameters. The first approach is called a *line search* approach where we look for an optimal Newton direction \mathbf{s}_{QP}^* being the solution to the following subproblem

$$\left. \begin{array}{l} \text{Find } \mathbf{s}_{\text{QP}}^* := \arg \min_{\mathbf{s} \in \mathbb{R}^{n_\Upsilon}} \left\{ \left[Q(\tilde{\mathcal{J}}, \mathbf{p}_0) \right] (\mathbf{s}) \right\} \\ \text{subject to } [\mathcal{L}(\mathbf{v}, \mathbf{p}_0)] (\mathbf{s}) \leq \mathbf{0} \end{array} \right\}, \quad (\mathcal{QP}_1(\mathbf{p}_0))$$

where $\mathbf{s} := \mathbf{p} - \mathbf{p}_0$ stands for a directional vector from the initial design \mathbf{p}_0 to the current one \mathbf{p} , $Q(\tilde{\mathcal{J}}, \mathbf{p}_0)$ stands for quadratic Taylor's expansion, see Theorem 3.3, of the function $\tilde{\mathcal{J}}$ at the point \mathbf{p}_0 while skipping the constant term $\tilde{\mathcal{J}}(\mathbf{p}_0)$

$$\left[Q(\tilde{\mathcal{J}}, \mathbf{p}_0) \right] (\mathbf{s}) := \mathbf{grad}(\tilde{\mathcal{J}}(\mathbf{p}_0)) \cdot \mathbf{s} + \frac{1}{2} \mathbf{s} \cdot \left(\mathbf{Hess}(\tilde{\mathcal{J}}(\mathbf{p}_0)) \cdot \mathbf{s} \right), \quad \mathbf{s} \in \mathbb{R}^{n_\Upsilon}, \quad (6.15)$$

in which $\mathbf{Hess}(\tilde{\mathcal{J}}(\mathbf{p}_0)) \in \mathbb{R}^{n_\Upsilon \times n_\Upsilon}$ denotes the *Hessian matrix* whose entries are as follows:

$$\left[\mathbf{Hess}(\tilde{\mathcal{J}}(\mathbf{p}_0)) \right]_{i,j} := \frac{\partial^2 \tilde{\mathcal{J}}(\mathbf{p}_0)}{\partial p_i \partial p_j}, \quad i, j = 1, \dots, n_\Upsilon,$$

and where $\mathcal{L}(\tilde{\mathcal{J}}, \mathbf{p}_0)$ denotes linear Taylor's expansion, see Theorem 3.3, of the vector function \mathbf{v} at the point \mathbf{p}_0

$$[\mathcal{L}(\mathbf{v}, \mathbf{p}_0)] (\mathbf{s}) := \mathbf{v}(\mathbf{p}_0) + \mathbf{Grad}(\mathbf{v}(\mathbf{p}_0)) \cdot \mathbf{s}, \quad \mathbf{s} \in \mathbb{R}^{n_\Upsilon}, \quad (6.16)$$

in which the matrix $\mathbf{Grad}(\mathbf{v}(\mathbf{p}_0)) \in \mathbb{R}^{n_\Upsilon \times n_\nu}$ denotes the following gradient matrix

$$\mathbf{Grad}(\mathbf{v}(\mathbf{p}_0)) := [\mathbf{grad}(v_1(\mathbf{p}_0)), \dots, \mathbf{grad}(v_{n_\nu}(\mathbf{p}_0))].$$

The optimal direction \mathbf{s}_{QP}^* is then an input to the following one–dimensional optimization problem, the *line search problem*

$$\left. \begin{array}{l} \text{Find } \alpha_{\text{QP}}^* := \arg \min_{\alpha > 0} \left\{ \tilde{\mathcal{J}}(\mathbf{p}_0 + \alpha \mathbf{s}_{\text{QP}}^*) \right\} \\ \text{subject to } \mathbf{v}(\mathbf{p}_0 + \alpha \mathbf{s}_{\text{QP}}^*) \leq \mathbf{0} \end{array} \right\}, \quad (\mathcal{LS}(\mathbf{p}_0, \mathbf{s}_{\text{QP}}^*))$$

and

$$\mathbf{p}_{\text{QP}}^* := \mathbf{p}_0 + \alpha_{\text{QP}}^* \mathbf{s}_{\text{QP}}^*$$

is the solution.

The second approach is called a *trust region* method. It supposes that the quadratic subproblem approximates the original problem well, but just in a given neighbourhood of \mathbf{p}_0 . Hence, given an initial point \mathbf{p}_0 and a trust region diameter $d > 0$, we solve the following quadratic subproblem

$$\left. \begin{array}{l} \text{Find } \mathbf{p}_{\text{QP}}^* := \arg \min_{\mathbf{p} \in \mathbb{R}^{n_T}} \left\{ \left[\mathcal{Q}(\tilde{\mathcal{J}}, \mathbf{p}_0) \right] (\mathbf{p} - \mathbf{p}_0) \right\} \\ \text{subject to } \left[\mathcal{L}(\mathbf{v}, \mathbf{p}_0) \right] (\mathbf{p} - \mathbf{p}_0) \leq \mathbf{0} \\ \|\mathbf{p} - \mathbf{p}_0\| \leq \frac{d}{2} \end{array} \right\}, \quad (\mathcal{QP}_2(\mathbf{p}_0, d))$$

where \mathcal{Q} and \mathcal{L} are respectively given by (6.15) and (6.16).

6.2.2 Sequential quadratic programming

The problem (\mathcal{QP}) is usually solved sequentially such that the optimal solution \mathbf{p}_{QP}^* is used as an initial design for the next quadratic subproblem. This is also referred to as *sequential quadratic programming* (SQP). Its two simplest versions that use the line search or the trust region approach, respectively, are sketched in Algorithm 3 or in Algorithm 4, cf. NOCEDAL AND WRIGHT [148, p. 532].

Algorithm 3 Sequential quadratic programming using the line search method

Given \mathbf{p}_0
 $k := 0$
while a convergence test is not satisfied **do**
 Solve $(\mathcal{QP}_1(\mathbf{p}_k)) \rightsquigarrow \mathbf{s}_{\text{QP}}^*$
 Solve $(\mathcal{LS}(\mathbf{p}_k, \mathbf{s}_{\text{QP}}^*)) \rightsquigarrow \alpha_{\text{QP}}^*$
 $\mathbf{p}_{\text{QP}}^* := \mathbf{p}_0 + \alpha_{\text{QP}}^* \mathbf{s}_{\text{QP}}^*$
 $\mathbf{p}_{k+1} := \mathbf{p}_{\text{QP}}^*$
 $k := k + 1$
end while
 $\mathbf{p}^* := \mathbf{p}_k$

Algorithm 4 Sequential quadratic programming using the trust region method

Given \mathbf{p}_0 and $d_0 > 0$
 $k := 0$
while a convergence test is not satisfied **do**
 Solve $(\mathcal{QP}_2(\mathbf{p}_k, d_k)) \rightsquigarrow \mathbf{p}_{\text{QP}}^*$
 $\mathbf{p}_{k+1} := \mathbf{p}_{\text{QP}}^*$
 Update $d_k \rightsquigarrow d_{k+1}$
 $k := k + 1$
end while
 $\mathbf{p}^* := \mathbf{p}_k$

Let us note that there are many aspects to deal with, as to find a proper convergence criterion or to modify the quadratic subproblem when it does not admit a solution, which is the case if the Hessian matrix or its certain invariant is not positive definite. Here, we want to mention the

BFGS modification, see FLETCHER [62], named after its authors Broyden, Fletcher, Goldfarb, and Shanno. It is originally based on the idea of DAVIDON [52, 53]. The method was a revolutionary improvement of the SQP algorithm. At each iteration, it requires to evaluate only the gradient of the objective and constraint functions, while the Hessian matrix is iteratively built up by measuring changes in the gradients. For $k \geq 0$, $k \in \mathbb{N}$, the BFGS formula is the following, cf. NOCEDAL AND WRIGHT [148, p. 25],

$$\mathbf{H}_{k+1} := \mathbf{H}_k - \frac{(\mathbf{H}_k \cdot \mathbf{s}_k) \cdot (\mathbf{H}_k \cdot \mathbf{s}_k)}{\mathbf{s}_k \cdot (\mathbf{H}_k \cdot \mathbf{s}_k)} + \frac{\mathbf{y}_k \cdot \mathbf{y}_k^T}{\mathbf{y}_k \cdot \mathbf{s}_k}, \quad (6.17)$$

where \mathbf{H}_k and \mathbf{H}_{k+1} are two successive approximations of the Hessian matrices $\text{Hess}(\tilde{\mathcal{J}}(\mathbf{p}_k))$ and $\text{Hess}(\tilde{\mathcal{J}}(\mathbf{p}_{k+1}))$, respectively, and where

$$\mathbf{s}_k := \mathbf{p}_{k+1} - \mathbf{p}_k \quad \text{and} \quad \mathbf{y}_k := \text{grad}(\tilde{\mathcal{J}}(\mathbf{p}_{k+1})) - \text{grad}(\tilde{\mathcal{J}}(\mathbf{p}_k)).$$

The SQP method with the BFGS update is classified as a *quasi-Newton method*.

6.3 The first-order sensitivity analysis methods

Recall that by the first-order sensitivity analysis we mean calculation of gradients of the cost and constraint functions. There are three kinds of the sensitivity analysis methods, namely, a numerical differentiation, an automatic differentiation, and semianalytical methods, while others are certain modifications and/or combinations of them.

The most frequently used is the *numerical differentiation*. It is usually based on formulas for the central difference. Given a function $f \in C^1(\bar{\Omega})$, $\Omega \subset \mathbb{R}^n$, and a point $\mathbf{x} := (x_1, \dots, x_i, \dots, x_n) \in \Omega$, then, using linear Taylor's expansion, see Theorem 3.3, we can derive the following first central difference formula

$$\frac{\partial f(\mathbf{x})}{\partial x_i} \approx \frac{f(x_1, \dots, x_i + \delta, \dots, x_n) - f(x_1, \dots, x_i - \delta, \dots, x_n)}{2\delta},$$

where $\delta > 0$. The approximation error decreases with δ^2 until a computer round-off error becomes significant. Therefore, we have to choose δ such that neither the approximation nor round-off error is large. Another possibility is using a numerical differentiation formula of a higher m -th order, $m \in \mathbb{N}$, for which the approximation error decreases with δ^{m+1} . However, evaluating the gradient approximation needs $2mn$ evaluations of f , which is in the case of shape optimization very time consuming. Hence, we have to balance between the time issue and the precision. The advantages of the method are robustness and easy implementation.

The *automatic differentiation*, see GRIEWANK [70], differentiate the function f symbolically. The input of the method is a routine that evaluates $f(\mathbf{x})$ and the output is again a routine which now evaluates $\text{grad}(f(\mathbf{x}))$. An implementation of the method is very difficult, since it involves a syntax recognizing, and it relies on the programming language that f is coded in. Nowadays, there are free software packages available. The method is robust and precise up to the computer round-off error, but it is too much both time and memory consuming in case of shape optimization, since the routine for solving the linear system which arises from the finite element discretization is also differentiated symbolically.

Here, we will focus on the *semianalytical methods*, cf. HAUG, CHOI, AND KOMKOV [86], that bases on the *algebraic approach* to sensitivity analysis, cf. HASLINGER AND NEITTAANMÄKI [85]. The methods respect the structure of the shape optimization problem, in which solution to a linear system of algebraic equations is involved. The cost functional is a compound map and its gradient is then a product of the gradients of the individual submappings. Most difficult to evaluate is differentiation of the solution to the linear system with respect to nodal coordinates of the discretization grid. This is performed by solution to other linear systems with the original but transposed system matrix and with new right-hand side vectors. The method is precise up to the numerical error of the linear system solver. The computational time roughly corresponds to the computation of the function f . The method is not robust, as it covers just the shape optimization problems, nevertheless, some other classes of optimization problems, e.g., the topology optimization, have a similar structure. Thus, an extension of the method is straightforward. The semianalytical methods might also be combined with both the numerical and automatic differentiation.

6.3.1 Sensitivities of the cost and constraint functions

Consider the discretized shape optimization problem (\mathcal{P}) . The key point to an efficient implementation of the method is making use of the structure of the cost function $\tilde{\mathcal{J}}$. Recall that our constraint function is state-independent, i.e., it does not depend upon the solution to the governing state problem. Therefore, the evaluation of the gradient of the constraint function with respect to the design variables is simple. For $i = 1, \dots, n_v$ and $j = 1, \dots, n_\gamma$ the gradient is as follows:

$$\mathbf{grad}(v_i(\mathbf{p})) := \left(\frac{\partial v_i(\mathbf{p})}{\partial p_1}, \dots, \frac{\partial v_i(\mathbf{p})}{\partial p_{n_\gamma}} \right) \in \mathbb{R}^{n_\gamma}, \quad (6.18)$$

where $\mathbf{p} := (p_1, \dots, p_{n_\gamma}) \in \mathbb{R}^{n_\gamma}$ stands for a vector of design variables and where the constraint function is denoted by $\mathbf{v}(\mathbf{p}) := (v_1(\mathbf{p}), \dots, v_{n_v}(\mathbf{p})) \in \mathbb{R}^{n_v}$. The evaluation of the cost functional proceeds as depicted in (6.9). Let us express a partial derivative of the cost functional with respect to a design variable. Using the chain rule (3.17) for differentiation of a compound function, for $i = 1, \dots, n_\gamma$ we get

$$\begin{aligned} \frac{\partial \tilde{\mathcal{J}}(\mathbf{p})}{\partial p_i} &= \frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v})}{\partial p_i} = \\ &= \sum_{o=1}^{n_\alpha} \left\{ \frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v})}{\partial \alpha_j} + \sum_{k=1}^{n_x} \sum_{l=1}^m \left[\frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v})}{\partial x_{k,l}} + \right. \right. \\ &\quad \left. \left. + \sum_{v=1}^{n_v} \sum_{e \in E} \sum_{j=1}^{v_2} \frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v})}{\partial B_j^{v,e}} \left(\frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{k,l}} + \right. \right. \right. \\ &\quad \left. \left. \left. + \sum_{p=1}^{n_e} \frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial u_p^{v,e}} \frac{\partial u_p^{v,e}(\mathbf{x})}{\partial x_{k,l}} \right) \right] \sum_{o=1}^{n_\alpha} \frac{\partial x_{k,l}(\boldsymbol{\alpha})}{\partial \alpha_o} \right\} \frac{\partial \alpha_o(\mathbf{p})}{\partial p_i}, \end{aligned} \quad (6.19)$$

where

- $\mathbf{p} := (p_1, \dots, p_{n_\gamma}) \in \mathbb{R}^{n_\gamma}$ denotes a design vector,
- $\boldsymbol{\alpha}(\mathbf{p}) := (\alpha_1(\mathbf{p}), \dots, \alpha_{n_\alpha}(\mathbf{p})) \in \mathbb{R}^{n_\alpha}$ denotes control coordinates of the shape and for $o = 1, \dots, n_\alpha, i = 1, \dots, n_\gamma$ it holds that

$$\frac{\partial \alpha_o(\mathbf{p})}{\partial p_i} = \frac{\partial [F(\mathbf{p})](\mathbf{x}_{\omega,o})}{\partial p_i},$$

- $\mathbf{x}(\boldsymbol{\alpha}) := [\mathbf{x}_1(\boldsymbol{\alpha}), \dots, \mathbf{x}_{n_{\mathbf{x}}}(\boldsymbol{\alpha})] \in \mathbb{R}^{mn_{\mathbf{x}}}$ denotes a block column vector consisting of all the grid nodes, where for $o = 1, \dots, n_{\boldsymbol{\alpha}}$

$$\frac{\partial \mathbf{x}(\boldsymbol{\alpha})}{\partial \alpha_o} := \frac{\partial \Delta \mathbf{x}(\boldsymbol{\alpha})}{\partial \alpha_o} + \begin{pmatrix} [\mathcal{M}]_{1,o} \\ \vdots \\ [\mathcal{M}]_{n_{\mathbf{x}},o} \end{pmatrix}, \quad \text{where } \mathbf{K}(\mathbf{x}_0) \cdot \frac{\partial \Delta \mathbf{x}(\boldsymbol{\alpha})}{\partial \alpha_o} = \frac{\partial b(\boldsymbol{\alpha})}{\partial \alpha_o},$$

- $\mathbf{x}^e(\boldsymbol{\alpha}) := [\mathbf{x}_1^e(\boldsymbol{\alpha}), \dots, \mathbf{x}_{m+1}^e(\boldsymbol{\alpha})] \in \mathbb{R}^{m(m+1)}$ denotes a block column vector consisting of the corners of the element domain K^e , $e \in E$,
- $\mathbf{x}_l^e(\boldsymbol{\alpha}) := (x_{l,1}^e(\boldsymbol{\alpha}), \dots, x_{l,m}^e(\boldsymbol{\alpha})) \in \mathbb{R}^m$ denotes coordinates of the l -th corner of the element domain K^e , $e \in E$,
- $\mathbf{u}^v(\mathbf{x}) := (u_1^v(\mathbf{x}), \dots, u_n^v(\mathbf{x})) \in \mathbb{R}^n$ denotes the solution to v -th state problem, i.e., to the system of linear algebraic equations ($P^v(\mathbf{x})$),
- $\mathbf{u}^{v,e}(\mathbf{x}) := (u_1^{v,e}(\mathbf{x}), \dots, u_{n_e}^{v,e}(\mathbf{x})) \in \mathbb{R}^{n_e}$ denotes the solution of the problem ($P^v(\mathbf{x})$) associated to an element $e \in E$ in such a way that

$$u_j^{v,e}(\mathbf{x}) = u_{\mathcal{G}^e(j)}^v(\mathbf{x}),$$

- $\mathbf{B}^v(\mathbf{x}) := [\mathbf{B}^{v,e_1}(\mathbf{x}), \dots, \mathbf{B}^{v,e_{n_{\Omega}}}(\mathbf{x})] \in \mathbb{R}^{\nu_2 n_{\Omega}}$ denotes a block column vector resulting after the application of the operator \mathbf{B} to the solution of the problem ($P^v(\mathbf{x})$),
- $\mathbf{B}^{v,e}(\mathbf{x}) := (B_1^{v,e}(\mathbf{x}), \dots, B_{\nu_2}^{v,e}(\mathbf{x})) \in \mathbb{R}^{\nu_2}$ denotes the value of $\mathbf{B}(\mathbf{u}^v(\mathbf{x}))$ over the element domain K^e , $e \in E$, such that

$$\mathbf{B}^{v,e}(\mathbf{x}) := \mathbf{B}(\mathbf{x}^e, \mathbf{u}^{v,e}) := \sum_{j=1}^{n_e} u_j^{v,e}(\mathbf{x}) \mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_j^r(\widehat{\mathbf{x}})),$$

and

- $\widetilde{\mathcal{J}}(\mathbf{p}) := \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v}) \in \mathbb{R}$ denotes the value of the cost functional.

6.3.2 State sensitivity

The main computational effort in the formula (6.19) is connected with the bracket term, which is the sensitivity of $B(\mathbf{x}, \mathbf{u}^v(\mathbf{x}))$ with respect to the grid nodal coordinates \mathbf{x} , i.e., with the derivatives

$$\frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e}(\mathbf{x}))}{\partial x_{k,l}} := \frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{k,l}} + \sum_{p=1}^{n_e} \frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial u_p^{v,e}} \frac{\partial u_p^{v,e}(\mathbf{x})}{\partial x_{k,l}}, \quad (6.20)$$

where for $k = 1, \dots, n_{\mathbf{x}}$, $l = 1, \dots, m$, and for $p = 1, \dots, n_e$

$$\begin{aligned} \frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{k,l}} &:= 0, \quad \text{if } \forall z \in \{1, \dots, m+1\} : \mathcal{H}^e(z) \neq k, \\ \frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{k,l}} &:= \sum_{p=1}^{n_e} u_p^{v,e}(\mathbf{x}) \left[\frac{\partial \mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e)}{\partial x_{z,l}^e} \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_p^r(\widehat{\mathbf{x}})) \right]_j, \quad \text{if } \mathcal{H}^e(z) = k, \\ \frac{\partial B_j^{v,e}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial u_p^{v,e}} &:= \left[\mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_p^r(\widehat{\mathbf{x}})) \right]_j. \end{aligned}$$

In (6.20) it remains to express the derivative $\partial \mathbf{u}^v(\mathbf{x})/\partial x_{k,l}$. To this goal, let us differentiate the state equation ($P^v(\mathbf{x})$), where $v = 1, \dots, n_v$, with respect to the l -th coordinate of a node $\mathbf{x}_k \in \bar{\Omega}$, where $k = 1, \dots, n_x$. We arrive at the following linear system of equations

$$\mathbf{A}(\mathbf{x}) \cdot \frac{\partial \mathbf{u}^v(\mathbf{x})}{\partial x_{k,l}} = \frac{\partial \mathbf{f}^v(\mathbf{x})}{\partial x_{k,l}} - \frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_{k,l}} \cdot \mathbf{u}^v(\mathbf{x}) \quad (6.21)$$

which is solved for $\partial \mathbf{u}^v(\mathbf{x})/\partial x_{k,l}$, where

$$\frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_{k,l}} := \begin{pmatrix} \frac{\partial A_{1,1}(\mathbf{x})}{\partial x_{k,l}} & \dots & \frac{\partial A_{1,n}(\mathbf{x})}{\partial x_{k,l}} \\ \vdots & \ddots & \vdots \\ \frac{\partial A_{n,1}(\mathbf{x})}{\partial x_{k,l}} & \dots & \frac{\partial A_{n,n}(\mathbf{x})}{\partial x_{k,l}} \end{pmatrix}, \quad \frac{\partial \mathbf{f}^v(\mathbf{x})}{\partial x_{k,l}} := \begin{pmatrix} \frac{\partial f_1^v(\mathbf{x})}{\partial x_{k,l}} \\ \vdots \\ \frac{\partial f_n^v(\mathbf{x})}{\partial x_{k,l}} \end{pmatrix},$$

$$\frac{\partial \mathbf{u}^v(\mathbf{x})}{\partial x_{k,l}} := \begin{pmatrix} \frac{\partial u_1^v(\mathbf{x})}{\partial x_{k,l}} \\ \vdots \\ \frac{\partial u_n^v(\mathbf{x})}{\partial x_{k,l}} \end{pmatrix},$$

and where $\mathbf{A}(\mathbf{x}) := (A_{i,j}(\mathbf{x}))_{i,j=1}^n$, $\mathbf{f}^v(\mathbf{x}) := (f_i^v(\mathbf{x}))_{i=1}^n$, and $\mathbf{u}^v(\mathbf{x}) := (u_i^v(\mathbf{x}))_{i=1}^n$, respectively, are the system matrix, the right-hand side vector, and the solution to the state problem ($P^v(\mathbf{x})$). Due to Assumption 5.4, we can skip the term $\partial \mathbf{f}^v(\mathbf{x})/\partial x_{k,l}$. Hence, it remains to express the term $\partial \mathbf{A}(\mathbf{x})/\partial x_{k,l}$. From (6.4) and (6.5) it follows that

$$\frac{\partial A_{i,j}(\mathbf{x})}{\partial x_{k,l}} = \sum_{e \in E_i \cap E_j} \sum_{o,p=1}^{n_e} \frac{\partial a_{\mathbf{x}^e}^e(\boldsymbol{\xi}_o^e(\mathbf{x}^e), \boldsymbol{\xi}_p^e(\mathbf{x}^e))}{\partial x_{k,l}},$$

where

$$\begin{aligned} \frac{\partial a_{\mathbf{x}^e}^e(\boldsymbol{\xi}_o^e(\mathbf{x}^e), \boldsymbol{\xi}_p^e(\mathbf{x}^e))}{\partial x_{k,l}} &= 0, \quad \text{if } \forall z \in \{1, \dots, m+1\} : \mathcal{H}^e(z) \neq k, \\ \frac{\partial a_{\mathbf{x}^e}^e(\boldsymbol{\xi}_o^e(\mathbf{x}^e), \boldsymbol{\xi}_p^e(\mathbf{x}^e))}{\partial x_{k,l}} &= \\ &= \int_{K^r} \left(\frac{\partial \mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e)}{\partial x_{z,l}^e} \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_o^r) \right) \cdot \left(\mathbf{D}^e \cdot \left(\mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_p^r) \right) \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}} + \\ &+ \int_{K^r} \left(\mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_o^r) \right) \cdot \left(\mathbf{D}^e \cdot \left(\frac{\partial \mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e)}{\partial x_{z,l}^e} \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_p^r) \right) \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}} + \\ &+ \int_{K^r} \left(\mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_o^r) \right) \cdot \left(\mathbf{D}^e \cdot \left(\mathbf{S}_{\mathbf{B}}^e(\mathbf{x}^e) \cdot \mathbf{B}_{\widehat{\mathbf{x}}}(\widehat{\boldsymbol{\xi}}_p^r) \right) \right) \frac{\partial |\det(\mathbf{R}^e(\mathbf{x}^e))|}{\partial x_{z,l}^e} d\widehat{\mathbf{x}} + \\ &+ \varepsilon \int_{K^r} \left(\frac{\partial \mathbf{S}^e(\mathbf{x}^e)}{\partial x_{z,l}^e} \cdot \widehat{\boldsymbol{\xi}}_o^r \right) \cdot \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_p^r \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}} + \\ &+ \varepsilon \int_{K^r} \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_o^r \right) \cdot \left(\frac{\partial \mathbf{S}^e(\mathbf{x}^e)}{\partial x_{z,l}^e} \cdot \widehat{\boldsymbol{\xi}}_p^r \right) |\det(\mathbf{R}^e(\mathbf{x}^e))| d\widehat{\mathbf{x}} + \\ &+ \varepsilon \int_{K^r} \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_o^r \right) \cdot \left(\mathbf{S}^e(\mathbf{x}^e) \cdot \widehat{\boldsymbol{\xi}}_p^r \right) \frac{\partial |\det(\mathbf{R}^e(\mathbf{x}^e))|}{\partial x_{z,l}^e} d\widehat{\mathbf{x}}, \quad \text{if } \mathcal{H}^e(z) = k. \end{aligned} \quad (6.22)$$

Note that none of $\partial \mathbf{A}(\mathbf{x})/\partial x_{k,l}$ is evaluated itself. In Section 6.3.4 we will rather assemble the vector

$$[\mathbf{G}(\mathbf{x}, \mathbf{u}^v(\mathbf{x}))]^T \cdot \mathbf{v} \in \mathbb{R}^{mn_{\mathbf{x}}},$$

where

$$\mathbf{G}(\mathbf{x}, \mathbf{u}^v(\mathbf{x})) := \left[-\frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_{1,1}} \cdot \mathbf{u}^v(\mathbf{x}), \dots, -\frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_{n_{\mathbf{x}},m+1}} \cdot \mathbf{u}^v(\mathbf{x}) \right] \in \mathbb{R}^{n \times (mn_{\mathbf{x}})}, \quad (6.23)$$

and where $\mathbf{v} \in \mathbb{R}^n$.

6.3.3 Semianalytical methods

We employ matrix notation and, using (6.20), (6.21), (6.23), and the symmetry of $\mathbf{A}(\mathbf{x})$, we rewrite (6.19) as follows:

$$\begin{aligned} \underbrace{\text{grad}(\tilde{\mathcal{J}}(\mathbf{p}))}_{n_{\Gamma} \times 1} &= \underbrace{\text{Grad}(\alpha(\mathbf{p}))}_{n_{\Gamma} \times n_{\alpha}} \cdot \left\{ \underbrace{\text{grad}_{\alpha}(\mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v}))}_{n_{\alpha} \times 1} + \right. \\ &+ \underbrace{\text{Grad}(\mathbf{x}(\alpha))}_{n_{\alpha} \times mn_{\mathbf{x}}} \cdot \left[\underbrace{\text{grad}_{\mathbf{x}}(\mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v}))}_{mn_{\mathbf{x}} \times 1} + \sum_{v=1}^{n_v} \left(\underbrace{\text{Grad}_{\mathbf{x}}(B(\mathbf{x}, \mathbf{u}^v))}_{(mn_{\mathbf{x}}) \times (\nu_2 n_{\Omega})} + \right. \right. \\ &\left. \left. + \underbrace{\mathbf{G}(\mathbf{x}, \mathbf{u}^v)^T}_{(mn_{\mathbf{x}}) \times n} \cdot \underbrace{\mathbf{A}(\mathbf{x})^{-1}}_{n \times n} \cdot \underbrace{\text{Grad}_{\mathbf{u}^v}(B(\mathbf{x}, \mathbf{u}^v))}_{n \times (\nu_2 n_{\Omega})} \right) \cdot \underbrace{\text{grad}_{B^v}(\mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v}))}_{(\nu_2 n_{\Omega}) \times 1} \right] \left. \right\}, \end{aligned} \quad (6.24)$$

in which matrix (or vector) size is written under the brackets, and where the gradients are

$$\text{grad}(\tilde{\mathcal{J}}(\mathbf{p})) := \begin{pmatrix} \frac{\partial \tilde{\mathcal{J}}(\mathbf{p})}{\partial p_1} \\ \vdots \\ \frac{\partial \tilde{\mathcal{J}}(\mathbf{p})}{\partial p_{n_{\Gamma}}} \end{pmatrix}, \quad \text{grad}_{\alpha}(\mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v})) := \begin{pmatrix} \frac{\partial \mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v})}{\partial \alpha_1} \\ \vdots \\ \frac{\partial \mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v})}{\partial \alpha_{n_{\alpha}}} \end{pmatrix},$$

$$\text{Grad}(\alpha(\mathbf{p})) := [\text{grad}(\alpha_1(\mathbf{p})), \dots, \text{grad}(\alpha_{n_{\alpha}}(\mathbf{p}))] := \begin{bmatrix} \frac{\partial \alpha_1(\mathbf{p})}{p_1} & \cdots & \frac{\partial \alpha_{n_{\alpha}}(\mathbf{p})}{p_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \alpha_1(\mathbf{p})}{p_{n_{\Gamma}}} & \cdots & \frac{\partial \alpha_{n_{\alpha}}(\mathbf{p})}{p_{n_{\Gamma}}} \end{bmatrix},$$

$$\begin{aligned} \text{Grad}(\mathbf{x}(\alpha)) &:= [\text{Grad}(\mathbf{x}_1(\alpha)), \dots, \text{Grad}(\mathbf{x}_{n_{\mathbf{x}}}(\alpha))] := \\ &:= \begin{bmatrix} \left(\frac{\partial x_{1,1}(\alpha)}{\partial \alpha_1} & \cdots & \frac{\partial x_{1,m}(\alpha)}{\partial \alpha_1} \right) & \cdots & \left(\frac{\partial x_{n_{\mathbf{x}},1}(\alpha)}{\partial \alpha_1} & \cdots & \frac{\partial x_{n_{\mathbf{x}},1}(\alpha)}{\partial \alpha_1} \right) \\ \vdots & & & \ddots & & & \vdots \\ \left(\frac{\partial x_{1,1}(\alpha)}{\partial \alpha_{n_{\alpha}}} & \cdots & \frac{\partial x_{1,m}(\alpha)}{\partial \alpha_{n_{\alpha}}} \right) & \cdots & \left(\frac{\partial x_{n_{\mathbf{x}},1}(\alpha)}{\partial \alpha_{n_{\alpha}}} & \cdots & \frac{\partial x_{n_{\mathbf{x}},1}(\alpha)}{\partial \alpha_{n_{\alpha}}} \right) \end{bmatrix}, \end{aligned}$$

$$\text{grad}_{\mathbf{x}}(\mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v})) := \begin{pmatrix} \text{grad}_{\mathbf{x}_1}(\mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v})) \\ \vdots \\ \text{grad}_{\mathbf{x}_{n_{\mathbf{x}}}}(\mathcal{I}(\alpha, \mathbf{x}, B^1, \dots, B^{n_v})) \end{pmatrix},$$

in which for $k = q, \dots, n_x$

$$\mathbf{grad}_{\mathbf{x}_k}(\mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})) := \begin{pmatrix} \frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})}{\partial x_{k,1}} \\ \vdots \\ \frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})}{\partial x_{k,m}} \end{pmatrix},$$

where further

$$\mathbf{Grad}_{\mathbf{x}}(B(\mathbf{x}, \mathbf{u}^v)) := \begin{bmatrix} \mathbf{Grad}_{\mathbf{x}_1}(B(\mathbf{x}^{e_1}, \mathbf{u}^{v,e_1})) & \dots & \mathbf{Grad}_{\mathbf{x}_1}(B(\mathbf{x}^{e_{n_\Omega}}, \mathbf{u}^{v,e_{n_\Omega}})) \\ \vdots & \ddots & \vdots \\ \mathbf{Grad}_{\mathbf{x}_{n_x}}(B(\mathbf{x}^{e_1}, \mathbf{u}^{v,e_1})) & \dots & \mathbf{Grad}_{\mathbf{x}_{n_x}}(B(\mathbf{x}^{e_{n_\Omega}}, \mathbf{u}^{v,e_{n_\Omega}})) \end{bmatrix},$$

in which for $k = 1, \dots, n_x$

$$\mathbf{Grad}_{\mathbf{x}_k}(B(\mathbf{x}^{e_1}, \mathbf{u}^{v,e_1})) = \mathbf{0}, \quad \text{if } \forall z \in \{1, \dots, m+1\} : \mathcal{H}^e(z) \neq k,$$

$$\mathbf{Grad}_{\mathbf{x}_k}(B(\mathbf{x}^e, \mathbf{u}^{v,e})) = \begin{pmatrix} \frac{\partial B_1(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{z,1}^e} & \dots & \frac{\partial B_{\nu_2}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{z,1}^e} \\ \vdots & \vdots & \vdots \\ \frac{\partial B_1(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{z,m}^e} & \dots & \frac{\partial B_{\nu_2}(\mathbf{x}^e, \mathbf{u}^{v,e})}{\partial x_{z,m}^e} \end{pmatrix}, \quad \text{if } \mathcal{H}^e(z) = k,$$

where further

$$\mathbf{Grad}_{\mathbf{u}^v}(B(\mathbf{x}, \mathbf{u}^v)) := [\mathbf{Grad}_{\mathbf{u}^v}(B(\mathbf{x}^{e_1}, \mathbf{u}^{v,e_1})), \dots, \mathbf{Grad}_{\mathbf{u}^v}(B(\mathbf{x}^{e_{n_\Omega}}, \mathbf{u}^{v,e_{n_\Omega}}))],$$

in which for $e \in E$

$$\mathbf{Grad}_{\mathbf{u}^v}(B(\mathbf{x}^e, \mathbf{u}^{v,e})) := [\mathbf{grad}_{\mathbf{u}^v}(B_1(\mathbf{x}^e, \mathbf{u}^{v,e})), \dots, \mathbf{grad}_{\mathbf{u}^v}(B_{\nu_2}(\mathbf{x}^e, \mathbf{u}^{v,e}))],$$

and where finally

$$\mathbf{grad}_{B^v}(\mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})) := \begin{pmatrix} \mathbf{grad}_{B^{v,e_1}}(\mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})) \\ \vdots \\ \mathbf{grad}_{B^{v,e_{n_\Omega}}}(\mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})) \end{pmatrix},$$

in which for $e \in E$

$$\mathbf{grad}_{B^{v,e}}(\mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})) := \begin{pmatrix} \frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})}{\partial B_1^{v,e}} \\ \vdots \\ \frac{\partial \mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, B^1, \dots, B^{n_v})}{\partial B_{\nu_2}^{v,e}} \end{pmatrix}.$$

Now, all the art is how to evaluate the expression (6.24) efficiently. Basically, there are two possibilities. Either we proceed from left to right, then it is called the *direct method*, or the other way round, which is called the *adjoint method*. The main computational effort is in calculating the state sensitivity. In case of the direct method, we would solve $n_v n_\Upsilon$ systems consisting of n linear equations, while, in case of the adjoint method, we have to solve just n_v systems of n linear equations. This is why we prefer the latter. Let us note that if the constraint function v were state dependent, the adjoint method would arrive at solving $n_v(1 + n_v)$ systems of n linear equations.

6.3.4 Adjoint method

The method bases on evaluating the expression (6.24) from right to left such that not all the individual gradients are calculated but rather so-called *adjoint variables* are assembled. Algorithm 5 describes the method. There, the symbols $\gamma \in \mathbb{R}^{n_\gamma}$, $\tau, \tau_1, \tau_2 \in \mathbb{R}^{mn_x}$, $\lambda, \delta \in \mathbb{R}^n$, and $\theta \in \mathbb{R}^{n_\alpha}$ stand for the adjoint variables.

Algorithm 5 Adjoint method

```

Given  $\mathbf{p}, \alpha, \mathbf{x}, \mathbf{A}(\mathbf{x}), \mathbf{u}^v$ , and  $\mathbf{B}^v$  for  $v = 1, \dots, n_v$ 
Evaluate  $\mathcal{I}_\alpha := \mathbf{grad}_\alpha(\mathcal{I}(\alpha, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v}))$ 
Evaluate  $\mathcal{I}_x := \mathbf{grad}_x(\mathcal{I}(\alpha, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v}))$ 
 $\tau := \mathbf{0}$ 
for  $v := 1, \dots, n_v$  do
  Evaluate  $\mathcal{I}_{B^v} := \mathbf{grad}_{B^v}(\mathcal{I}(\alpha, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v}))$ 
  Assemble  $\tau_1 := \mathbf{Grad}_x(\mathbf{B}(\mathbf{x}, \mathbf{u}^v)) \cdot \mathcal{I}_{B^v}$ 
  Assemble  $\lambda := \mathbf{Grad}_{u^v}(\mathbf{B}(\mathbf{x}, \mathbf{u}^v)) \cdot \mathcal{I}_{B^v}$ 
  Solve  $\mathbf{A}(\mathbf{x}) \cdot \delta = \lambda \rightsquigarrow \delta$ 
  Assemble  $\tau_2 := \mathbf{G}(\mathbf{x}, \mathbf{u}^v)^T \cdot \delta$ 
   $\tau := \tau + \tau_1 + \tau_2$ 
end for
 $\tau := \tau + \mathcal{I}_x$ 
Assemble  $\theta := \mathbf{Grad}(\mathbf{x}(\alpha)) \cdot \tau$ 
 $\theta := \theta + \mathcal{I}_\alpha$ 
Assemble  $\gamma := \mathbf{Grad}(\alpha(\mathbf{p})) \cdot \theta$ 
 $\mathbf{grad}(\tilde{\mathcal{J}}(\mathbf{p})) := \gamma$ 

```

Only the gradients of \mathcal{I} have to be provided by the user. All the other parts are more or less independent. The particular assembling procedures are depicted in Algorithms 7–9. On the CD there are enclosed the corresponding MATLAB [208] routines used for optimal shape design in 2-dimensional magnetostatics.

Algorithm 6 Adjoint method: the shape part (Assemble γ)

```

Given  $\mathbf{p}, \alpha$ , and  $\mathbf{v}$ 
 $\gamma := \mathbf{0}$ 
for  $i := 1, \dots, n_\gamma$  do
  for  $j := 1, \dots, n_\alpha$  do
     $\gamma_i := \gamma_i + (\partial [F(\mathbf{p})](\mathbf{x}_{\omega,j}) / \partial p_i) \cdot v_j$ 
  end for
end for

```

6.3.5 An object-oriented software library

Here, we present an efficient implementation of the adjoint method for optimal shape design in an object-oriented framework, see Fig. 6.1. Our main aim is a maximal reusability of the individual components when solving various shape optimization problems. For the details we also refer to LUKÁŠ, MÜHLHUBER, AND KUHN [125].

Algorithm 7 Adjoint method: the grid part (Assemble θ)

Given K , α , and τ
 Solve $K^T \cdot \beta = \tau \rightsquigarrow \beta$
for $o := 1, \dots, n_\alpha$ **do**
 $\theta_o := \beta \cdot (\partial b(\alpha) / \partial \alpha_o)$
 for $j := 1, \dots, n_x$ **do**
 $\theta_o := \theta_o + \mathcal{M}_{j,o} \cdot \tau_j$
 end for
end for

Algorithm 8 Adjoint method: FEM preprocessor part (Assemble τ_2)

Given x , u^v , and δ
 $\tau_2 := 0$
for $i := 1, \dots, n_\Omega$ **do**
 for $z := 1, \dots, m + 1$ **do**
 for $l := 1, \dots, m$ **do**
 $k := m (\mathcal{H}^{e_i}(z) - 1) + l$
 for $o, p := 1, \dots, n_e$ **do**
 if $o = p$ or $(\mathcal{G}^{e_i}(o) \notin \mathcal{I}_0^h$ and $\mathcal{G}^{e_i}(p) \notin \mathcal{I}_0^h)$ **then**
 Evaluate $[\tau_2]_k := [\tau_2]_k - \delta_{\mathcal{G}^{e_i}(o)} \cdot \left(\partial \alpha_{x^{e_i}}^{e_i}(\xi_o^{e_i}(x^{e_i}), \xi_p^{e_i}(x^{e_i})) / \partial x_{z,l}^{e_i} \right) \cdot u_{\mathcal{G}^{e_i}(p)}^v$
 end if
 end for
 end for
 end for
end for

The library supports routines for evaluating the cost and constraint functions and their gradients with respect to the shape design parameters. The library can be used with any gradient- or Newton-like optimization algorithm, as SQP with BFGS (BFGS-SQP), see (6.17). The library uses 3 external modules: a mesh generator, a finite element method (FEM) module, and a solver of linear algebraic systems of equations. The mesh generator runs just once at the very beginning and it discretizes the domain Ω for the initial design \mathbf{p}_0 . It provides initial grid nodes $\mathbf{x}_0 := \mathbf{x}(\alpha(\mathbf{p}_0))$ and the discretization \mathcal{T} . Having some grid nodes \mathbf{x} , the FEM preprocessor assembles the matrix $\mathbf{A}(\mathbf{x})$ and the right-hand side vector $\mathbf{f}^v(\mathbf{x})$ for each state $v = 1, \dots, n_v$. Then, the solver of linear systems provides the solution u^v to the FEM postprocessor that assembles the solution B^v . The FEM module is moreover supposed to assemble the corresponding gradient-vector multiplications in Algorithm 5, namely τ_1 , τ_2 , and λ . The efficiency of the library strongly depends on the linear system solver. We have used the software tools developed by KUHN, LANGER, AND SCHÖBERL [117] at the University Linz in Austria, where the conjugate gradient method, cf. GOLUB AND VAN LOAN [69], with a multigrid preconditioning, cf. HACKBUSCH [78], is involved.

Now, let us explain how the optimization proceeds in terms of Fig. 6.1. Given an initial vector \mathbf{p}_0 of design parameters and a discretization parameter h , the BFGS-SQP algorithm starts its run while at the same time the mesh generator provides the initial grid \mathbf{x}_0 and the grid topological information \mathcal{T} . Then, a quadratic programming subproblem is going to be solved, see also Algorithms 3 and 4, which requires evaluation of $\tilde{\mathcal{J}}(\mathbf{p}_0)$, $\mathbf{v}(\mathbf{p}_0)$, $\text{grad}(\tilde{\mathcal{J}}(\mathbf{p}_0))$, and $\text{Grad}(\mathbf{v}(\mathbf{p}_0))$.

Algorithm 9 Adjoint method: FEM postprocessor part (Assemble τ_1 , Assemble λ)

Given \mathbf{x} , \mathbf{u}^v , and \mathcal{I}_{B^v}

$\tau_1 := \mathbf{0}$

$\lambda := \mathbf{0}$

for $i := 1, \dots, n_\Omega$ **do**

for $z := 1, \dots, m + 1$ **do**

for $l := 1, \dots, m$ **do**

$k := m(\mathcal{H}^{e_i}(z)) + l$

for $p := 1, \dots, n_e$ **do**

 Evaluate $\mathbf{c} := u_{\mathcal{G}^{e_i(p)}}^v \cdot \left(\partial \mathbf{S}_{\mathbf{B}}^{e_i}(\mathbf{x}^{e_i}) / \partial x_{z,l}^{e_i} \right) \cdot \mathbf{B}_{\hat{\mathbf{x}}}(\hat{\boldsymbol{\xi}}_p^r(\hat{\mathbf{x}}))$

 Evaluate $\mathbf{d} := \mathbf{S}_{\mathbf{B}}^{e_i}(\mathbf{x}^{e_i}) \cdot \mathbf{B}_{\hat{\mathbf{x}}}(\hat{\boldsymbol{\xi}}_p^r(\hat{\mathbf{x}}))$

for $j := 1, \dots, \nu_2$ **do**

$t := \nu_2(i - 1) + j$

$[\tau_1]_k := [\tau_1]_k + c_j [\mathcal{I}_{B^v}]_t$

$\lambda_{\mathcal{G}^{e_i(p)}} := \lambda_{\mathcal{G}^{e_i(p)}} + d_j [\mathcal{I}_{B^v}]_t$

end for

end for

end for

end for

end for

The evaluation of $\mathbf{v}(\mathbf{p}_0)$ is straightforward. The evaluation of $\tilde{\mathcal{J}}(\mathbf{p}_0)$ proceeds as depicted in (6.9), where the design-to-shape mapping, shape-to-mesh mapping, FEM preprocessor, solver of linear systems, FEM postprocessor, and computation of $\tilde{\mathcal{J}}$ modules take control of the run consecutively. Computing $\mathbf{Grad}(\mathbf{v}(\mathbf{p}_0))$ is again quite straightforward, since it is basically an analytical formula, which is evaluated in the module “Computation of $\mathbf{Grad}(\mathbf{v}(\mathbf{p}_0))$ ”, see Fig. 6.1. The most computationally expensive part is (together with the evaluation of the cost functional) the evaluation of $\mathbf{grad}(\tilde{\mathcal{J}}(\mathbf{p}_0))$. This evaluation follows Algorithms 5–9, where the input data flows as depicted in Fig. 6.1. Finally, the next iteration of the BFGS–SQP algorithm is performed and the procedure repeats for a new vector \mathbf{p} of design parameters until a terminate criterion is fulfilled.

From the programming point of view, the only part which has always to be coded by the user is the module that computes the cost functional $\mathcal{I}(\boldsymbol{\alpha}, \mathbf{x}, \mathbf{B}^1, \dots, \mathbf{B}^{n_v})$ and the constraint function $\mathbf{v}(\mathbf{p})$. Thus, we have minimized the programming effort that is necessary for solving a new shape optimization problem just to the specification of the problem itself.

6.3.6 A note on using the automatic differentiation

The semianalytical methods turned out to be most effective for the shape optimization, since they make much use of the problem structure. Nevertheless, some parts of the algorithm still remain to be automatized using, e.g., the automatic differentiation (AD) method. We have in mind the module that calculates the gradients of \mathcal{I} and \mathbf{v} with respect to the input variables $\boldsymbol{\alpha}$, \mathbf{x} , \mathbf{B}^v , and \mathbf{p} , respectively. In fact, the routine calculating these gradients is just an analytical differentiation of the routine that calculates the functions \mathcal{I} and \mathbf{v} . In this case, we avoid the main obstacle of using AD, namely, differentiation of the linear system solver. Another possible use of AD might be an automatic generation of routines that calculate sensitivities of the local contributions (6.22) to the system matrix \mathbf{A} .

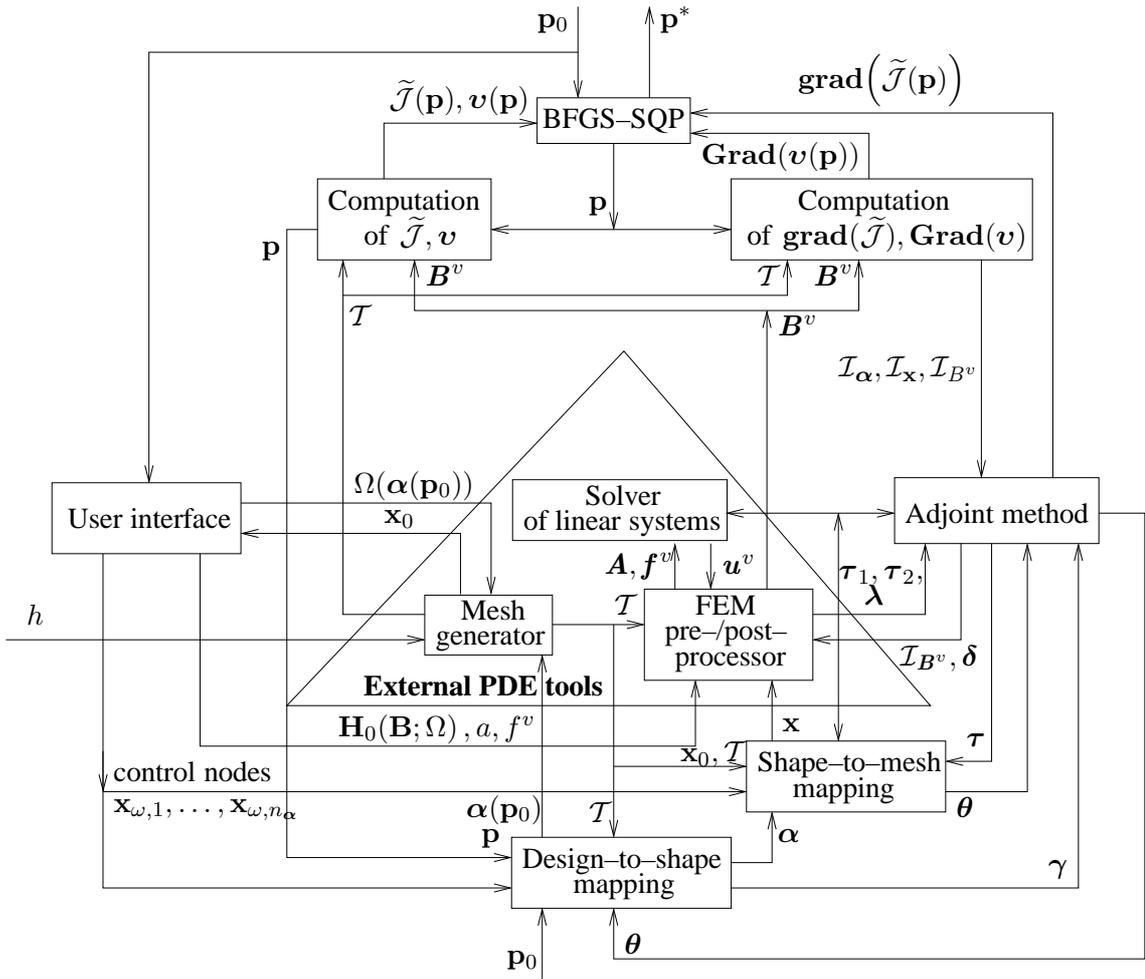


Figure 6.1: Structure of the library (data flow diagram)

6.4 Multilevel optimization approach

Here we introduce a rather new optimization approach. It has been inspired by techniques used in multigrid methods. This research was initiated by Prof. Ulrich Langer whose group at the University of Linz, Austria, has achieved world-leading results in scientific computing using multigrid methods, cf. APEL AND SCHÖBERL [10], HAASE ET AL. [73, 75, 76], HAASE AND LANGER [74], HAASE AND LINDNER [77], JUNG AND LANGER [102], KUHN, LANGER, AND SCHÖBERL [117], SCHINNERL, LANGER, AND LERCH [183], or SCHINNERL ET AL. [184]. We want to establish a hierarchy of discretizations to our continuous shape optimization problem (\tilde{P}) such that the optimized design achieved at a coarse level is used as the initial design at a next finer discretization level. The first results can be found in LUKÁŠ [123] and in LUKÁŠ [128].

In this section, we employ the full notation with both the regularization parameter ε and the discretization parameter h

$$\left. \begin{array}{l} \text{Find } \mathbf{p}_\varepsilon^{h*} \in \Upsilon: \\ \tilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}_\varepsilon^{h*}) \leq \tilde{\mathcal{J}}_\varepsilon^h(\mathbf{p}) \quad \forall \mathbf{p} \in \Upsilon \end{array} \right\}, \quad (\tilde{P}_\varepsilon^h)$$

where $\tilde{\mathcal{J}}_\varepsilon^h : \Upsilon \mapsto \mathbb{R}$ denotes the discretized and regularized cost functional and

$$\Upsilon := \{\mathbf{p} \in \mathbb{R}^{n_\Upsilon} \mid \mathbf{v}(\mathbf{p}) \leq \mathbf{0}\}$$

is the set of admissible design parameters, where $\mathbf{v} : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_v}$, $n_v \in \mathbb{N}$.

By the *classical optimization approach* we mean the standard technique when, given a fixed regularization parameter ε , a fixed discretization parameter h , both of which are small enough, given an initial vector \mathbf{p}_0 of design parameters, the optimization algorithm proceeds just once ending up with the optimized design $\mathbf{p}_\varepsilon^{h*}$, see Algorithm 10.

Algorithm 10 Classical optimization approach

Given ε , h and \mathbf{p}_0
 Discretize $(\tilde{P}) \rightsquigarrow (\tilde{P}_\varepsilon^h)$
 Solve $(\tilde{P}_\varepsilon^h)$ with the initial design $\mathbf{p}_0 \rightsquigarrow \mathbf{p}_\varepsilon^{h*}$
 $\mathbf{p}_\varepsilon^{h*}$ is the optimized design

By the *multilevel or hierarchical optimization approach* we mean that, given an initial design $\mathbf{p}_{1,0}$, we first regularize and discretize the problem (P) at the first level with a rather large values of ε_1 and h_1 , and the optimization algorithm proceeds ending up with a coarse optimized design $\mathbf{p}_{\varepsilon_1}^{h_1*}$. Then, we refine both the regularization and the discretization parameters and run the optimization algorithm with smaller values of ε_2 and h_2 while using the design $\mathbf{p}_{\varepsilon_1}^{h_1*}$ as the initial one at this second level. We end up with a finer optimized design $\mathbf{p}_{\varepsilon_2}^{h_2*}$, and so further. The approach is described in Algorithm 11.

Algorithm 11 Hierarchical optimization approach

Given ε_1 , h_1 , and $\mathbf{p}_{1,0}$
 $l := 1$
while $l > 1$ and a terminate criterion is not satisfied **do**
 Discretize (\tilde{P}) at the level $l \rightsquigarrow (\tilde{P}_{\varepsilon_l}^{h_l})$
 Solve $(\tilde{P}_{\varepsilon_l}^{h_l})$ with the initial design $\mathbf{p}_{l,0} \rightsquigarrow \mathbf{p}_{\varepsilon_l}^{h_l*}$
 Refine ε_l , $h_l \rightsquigarrow \varepsilon_{l+1}$, h_{l+1}
 $\mathbf{p}_{l+1,0} := \mathbf{p}_{\varepsilon_l}^{h_l*}$
 $l := l + 1$
end while
 $\mathbf{p}_{\varepsilon_{l-1}}^{h_{l-1}*}$ is the optimized design

The hierarchical approach in shape optimization has turned out to be much more effective than the classical one whenever the coarse optimized design $\mathbf{p}_{\varepsilon_0}^{h_0*}$ approximates the true one rather well. The crucial part of the algorithm is the refinement step. The updated values of ε_{l+1} and h_{l+1} must not be too smaller than those of ε_l and h_l , since SQP would take many iterations. On the other hand, if the refinement is rather coarse, i.e., the values of ε_{l+1} and h_{l+1} are comparable to ε_l and h_l , there is hardly any progress in the SQP algorithm and the hierarchical approach takes many iterations. In Section 7.4 we provide some numerical experiments without any use of multigrid yet. The idea, which we want to investigate in the future, is that the refinement strategy should benefit from the a posteriori finite element error analysis and from multigrid techniques. The first papers in this context have appeared just recently, see RAMM, MAUTE, AND SCHWARZ [164], and SCHLEUPEN, MAUTE, AND RAMM [185]. In the paper by SCHERZER [182] a multilevel approach is used for solving nonlinear ill-posed problems.

Another improvement can be done, when applying the multilevel approach on the level of mathematical modelling. It means that in those application where we can reduce the problem complexity by neglecting a dimension or some physical phenomena we can first solve the discretized reduced problem and then use the result as the initial design for the more complex problem. A typical example might be solving a shape optimization problem governed first by 2d linear magnetostatic state problem, then, prolong the optimized design dimensionally and use it as the initial design for shape optimization governed by 3d linear magnetostatics, and finally use the resulting shape as the initial design for shape optimization governed by 3d nonlinear magnetostatic state problem. In Section 7.4, we will give a numerical test of the 2d/3d dimensional step.

Chapter 7

An application and numerical experiments

At the beginning of this chapter, we will introduce a physical application, being of a practical use, which will lead to a problem of optimal shape design of electromagnets. Then, we will introduce mathematical settings of the shape optimization problems governed by linear magnetostatic multistate problems in both two- and three-dimensional cases. We will utilize the abstract theory introduced in Chapters 3–5 such that we will specify all the symbols introduced in Chapter 5, validate the corresponding assumptions, and the related existence as well as convergence theorems will then follow. We will also easily check Assumptions 6.1 and 6.2, both introduced in Chapter 6, which will justify us to use the SQP algorithm. Then, we will present shapes resulting from numerical calculations. We will also present numerical experiments with the multilevel optimization approach and with the adjoint method, which were both introduced in Chapter 6. We will compare them to the classical optimization approach and to the numerical differentiation, respectively. Some of the optimized shapes were manufactured and we are provided with physical measurements of the magnetic field. At the end, we will discuss magnetic field improvements in terms of the cost functional with respect to the original design.

The results of this chapter can be also found in PIŠTORA, POSTAVA, AND ŠEBESTA [160], KOPŘIVA ET AL. [111], LUKÁŠ [119, 120, 121, 122, 123, 128], LUKÁŠ ET AL. [124], and in LUKÁŠ, MÜHLHUBER, AND KUHN [125].

7.1 A physical problem

Let us consider two geometries of electromagnets: the *Maltese Cross* (MC) geometry and the *O-Ring* geometry that are both depicted in Fig. 7.1. Each consists of a ferromagnetic yoke and poles. There are 4 poles in case of the Maltese Cross and 8 ones in case of the O-Ring electromagnet. The poles are completed with coils which are pumped with direct electric currents. The electromagnets are used for measurements of *Kerr magneto-optic effects*, cf. ZVEDIN [223, p. 40]. They require the magnetic field as *homogeneous*, i.e., as constant as possible in a given normal direction. Let us note that the magneto-optic effects are investigated for applications in high capacity data storage media, like a development of new media materials for magnetic or compact discs recording. Let us also note that the electromagnets have been developed at the Institute of Physics, VŠB–Technical University of Ostrava, Czech Republic in the research group of Prof. Jaromír Pištora. Some instances have been already delivered to the following laboratories:

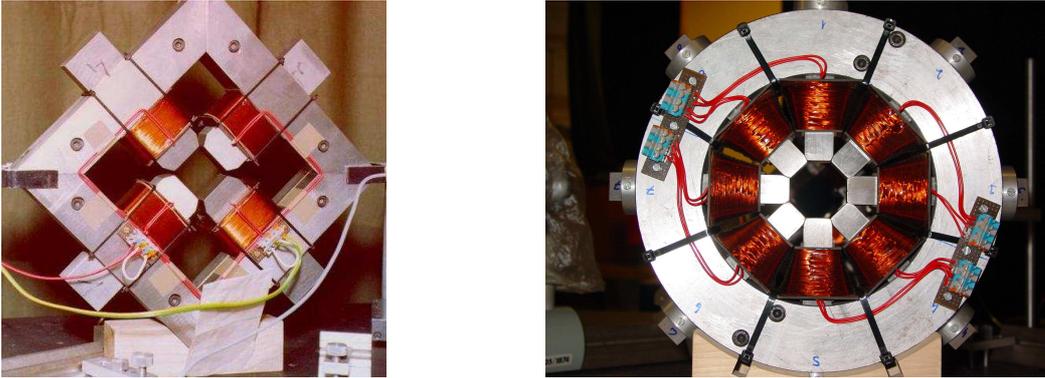


Figure 7.1: The Maltese Cross and O-Ring electromagnets

- Institute of Physics, Charles University Prague, Czech Republic,
- National Institute of Applied Sciences INSA in Toulouse, France,
- Department of Physics, Simon Fraser University in Vancouver, Canada,
- Department of Chemistry, Simon Fraser University in Vancouver, Canada,
- and University Paris VI., France.

First, we describe how the Kerr magneto-optic effect is measured. Let us consider, for instance, the Maltese Cross electromagnet, as in Fig. 7.1, and its cross-section, see Fig. 7.2. A sample of

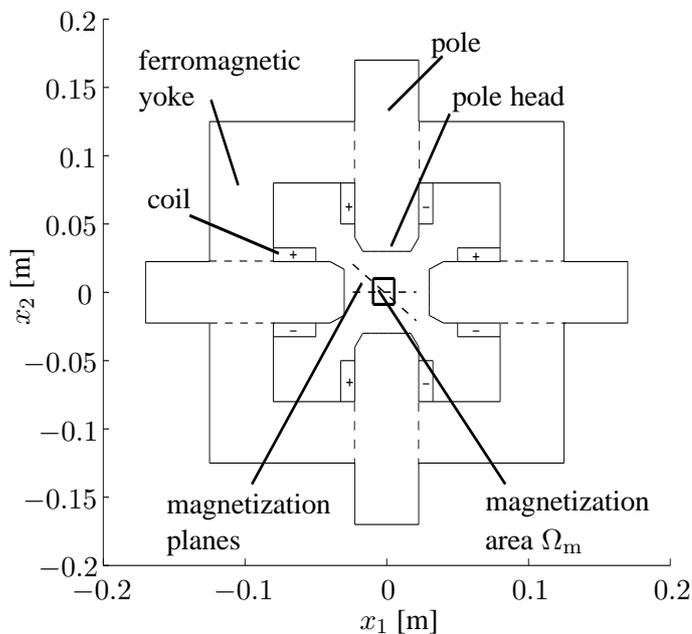


Figure 7.2: Cross-section of the Maltese Cross electromagnet

a magnetic material is placed into the *magnetization area* which is located in the middle among

the pole heads. In this area the magnetic field is homogeneous enough with respect to the normal vector of some *polarization plane*, see Fig. 7.2. We pass an optical (light) beam of a given polarization vector to the sample. There it reflects and components of the reflected polarization vector are measured in terms of the Kerr rotation and ellipticity, respectively. Briefly saying, we measure the polarization state of the reflected beam. The Kerr rotation means the difference between the angle of the main ellipticity axis of the reflected beam from that one before the reflection. Typical measured data is depicted in Fig. 7.3, which was measured by Ing. Igor Kopřiva at the Institute of Physics, VŠB–Technical University of Ostrava, see also KOPŘIVA ET AL. [111].

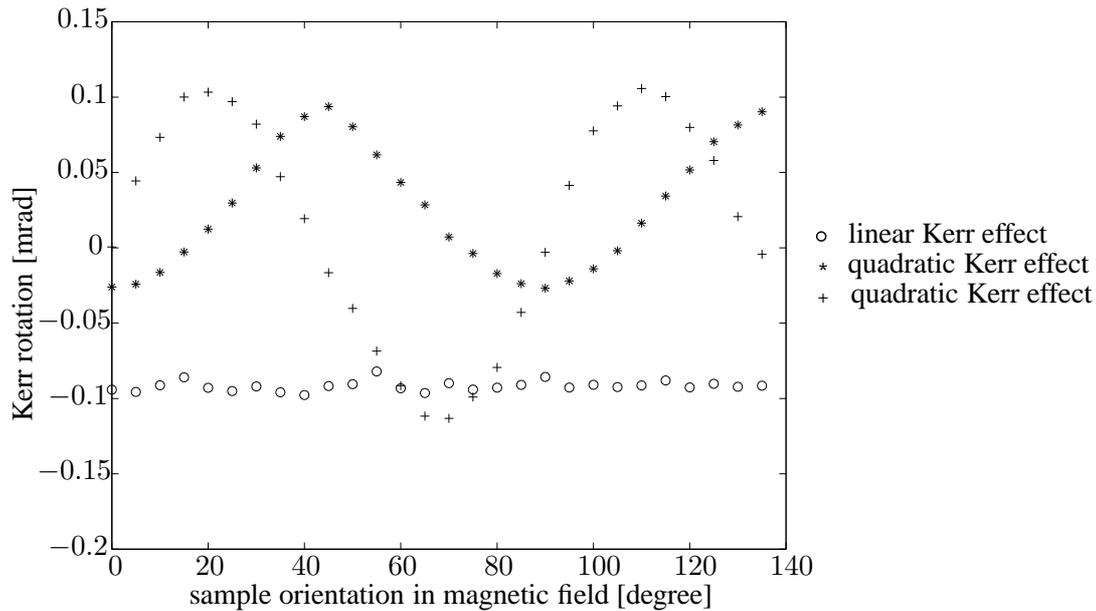


Figure 7.3: Dependence of magneto-optic effects on a sample rotation

From Fig. 7.3 we can see that the Kerr rotation depends on the orientation of the sample in the magnetic field, which is significant especially for the *quadratic Kerr effect*. This indicates anisotropic behaviour of the sample. Therefore, we should measure it in as many directions as possible. One has either to rotate the sample in the magnetic field, rotate the electromagnet while the sample is fixed, or rotate the magnetic field itself while both the sample and electromagnet are fixed. Certainly, the last variant is most preferred. The electromagnets have been developed such that they are capable to generate magnetic fields homogeneous in step-by-step different directions just by switching some currents in coils on or off, or by switching their senses. The more coils we have, the more directions the magnetic field can be oriented in. In case of the Maltese Cross or the O-Ring electromagnet, one can sequentially generate magnetic fields homogeneous in up to 8 or 16 directions, respectively. This will lead us to a multistate problem where only the current excitations, i.e., the right-hand sides differ.

Our aim is to improve the current geometries of electromagnets, see Fig. 7.1, in order to be better suited for measurements of the Kerr effect. The generated magnetic field should be strong and homogeneous enough in order to admit a magneto-optic effect. Unfortunately, these assumptions are contradictory and we have to balance them. From physical experience we know that the homogeneity of the magnetic field depends significantly on the shape of the pole heads. Hence, we aim at designing shapes of the pole heads in such a way that inhomogeneities of the

magnetic field are minimized, but the field itself is still strong enough.

7.2 Three-dimensional mathematical setting

Now, we introduce a complete 3d mathematical setting of the shape optimization problem. We will specify the abstract symbols and assumptions that were introduced in the previous text and that are also summarized in Section 6.1. As there are no principal differences between the Maltese Cross and the O-Ring electromagnet, we will describe both at once.

Convention 7.1. *In all what follows the dimensions will be given in meters except for Figs. 7.5–7.7 and 7.19, where they are in millimeters.*

7.2.1 Geometries of the electromagnets

The computational domain is fixed and in case of the Maltese Cross and the O-Ring it is, respectively, as follows:

$$\Omega := \left(-\frac{d_1}{2}, \frac{d_1}{2}\right) \times \left(-\frac{d_2}{2}, \frac{d_2}{2}\right) \times \left(-\frac{d_3}{2}, \frac{d_3}{2}\right)$$

and

$$\Omega := \left\{ \mathbf{x} := (x_1, x_2, x_3) \in \mathbb{R}^3 \mid (x_1)^2 + (x_2)^2 < (r)^2 \text{ and } |x_3| < \frac{d_3}{2} \right\},$$

where

$$d_1 := d_2 := 0.4 \text{ [m]}, \quad d_3 := 0.02 \text{ [m]}, \quad r := 0.2 \text{ [m]}.$$

The computational domain obviously fulfills Assumption 3.1 and $\Omega \in \mathcal{L}$.

We describe the geometrical models of the electromagnets. Referring to Fig. 7.4, the green

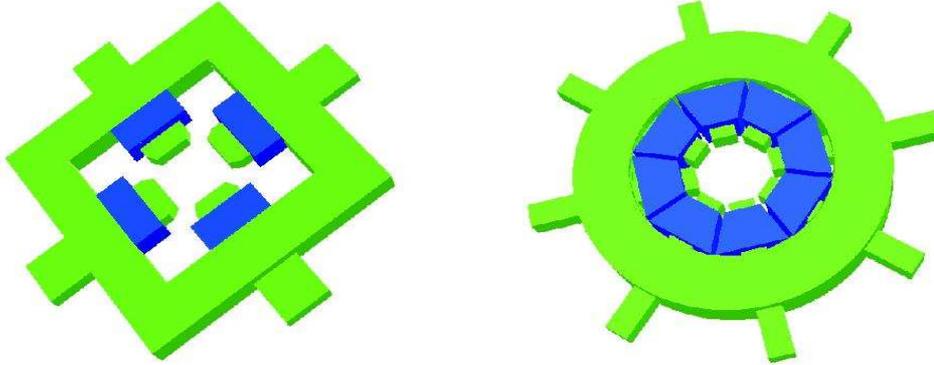


Figure 7.4: Geometrical models of the Maltese Cross and O-Ring electromagnets

parts are the ferromagnetic yoke and the poles. The blue parts are the coils. In Fig. 7.5 and in Fig. 7.6 we can see dimensions in millimeters for geometrical models of the Maltese Cross and of the O-Ring electromagnet, respectively. The symbol Ω_{yoke} stands for the domain occupied by the ferromagnetic yoke, the symbols Ω_{westp} , $\Omega_{\text{northwestp}}$, Ω_{northp} , $\Omega_{\text{northeastp}}$, Ω_{eastp} , $\Omega_{\text{southeastp}}$, Ω_{southp} , and $\Omega_{\text{southwestp}}$ denote the domains occupied by the particular poles, and the symbols Ω_{westc} , $\Omega_{\text{northwestc}}$, Ω_{northc} , $\Omega_{\text{northeastc}}$, Ω_{eastc} , $\Omega_{\text{southeastc}}$, Ω_{southc} , and $\Omega_{\text{southwestc}}$ denote the domains that are occupied by the corresponding coils. In Fig. 7.7 we can see the west pole of

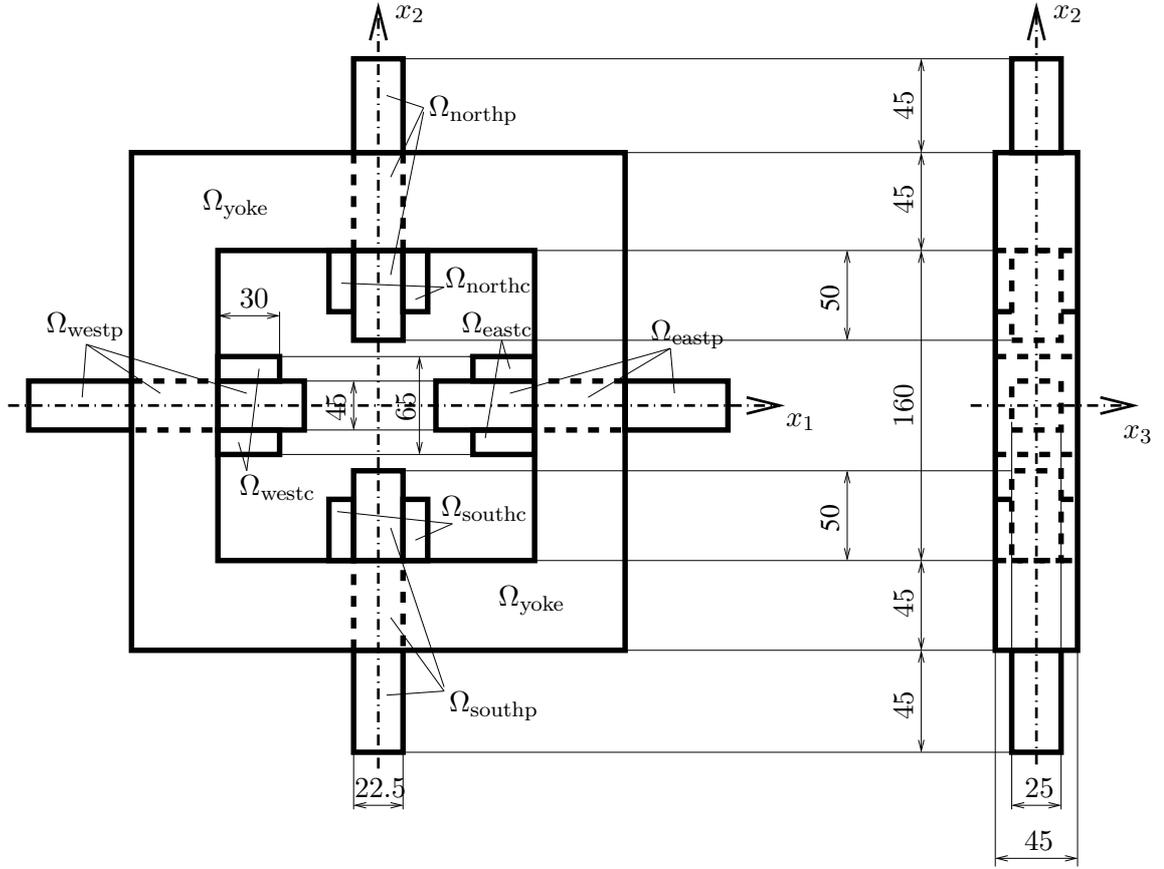


Figure 7.5: Drawing of the Maltese Cross electromagnet

the O-Ring in detail. The only geometrical parts being changed during the optimization will be shapes of the pole heads, see Fig. 7.7.

7.2.2 Set of admissible shapes

We assume all the shapes to be same and, moreover, symmetrical with respect to the two corresponding planes, e.g., with respect to $x_1 = 0$ and $x_3 = 0$ in case of the north pole head. Thus, from now on we will represent the shape of an arbitrary pole head by the shape of the north pole head. Due to the symmetry we consider only its quarter. Shape is then a continuous function defined over the domain

$$\omega := \left(0, \frac{d_{\text{pole},1}}{2}\right) \times \left(0, \frac{d_{\text{pole},3}}{2}\right),$$

where in case of the Maltese Cross electromagnet

$$d_{\text{pole},1} := 0.0225 \text{ [m]}, \quad d_{\text{pole},3} := 0.025 \text{ [m]},$$

and in case of the O-Ring

$$d_{\text{pole},1} := d_{\text{pole},3} := 0.02 \text{ [m]}.$$

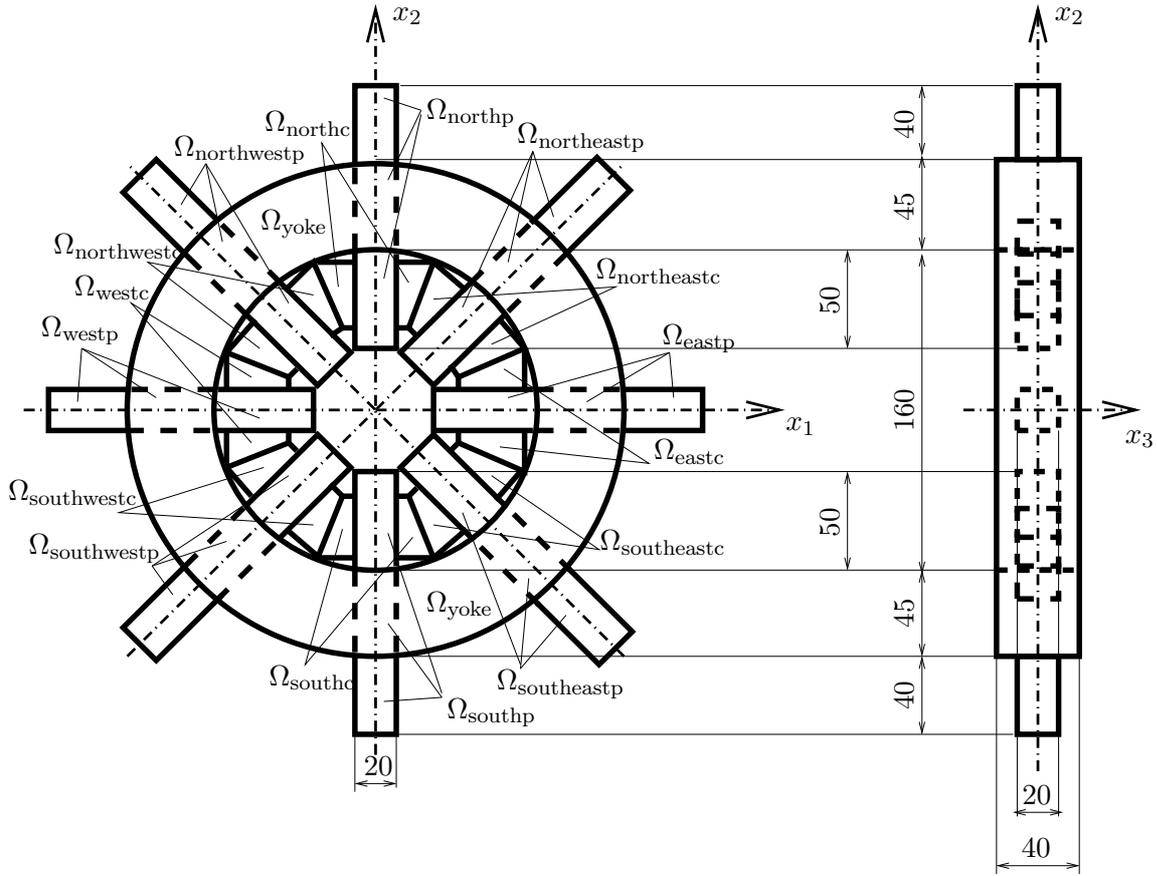


Figure 7.6: Drawing of the O-Ring electromagnet

The Lipschitz constant C_{16} in (5.1) corresponds to the maximal slope angle. We choose

$$C_{16} := \frac{3\pi}{8}.$$

The box constraints (5.2) are chosen such that the shape of the west pole head must not be either higher than the bottom of the north coil or penetrate with the neighbouring pole head. Therefore, we choose

$$\alpha_l := 0.012 \text{ [m]}, \quad \alpha_u := 0.05 \text{ [m]}$$

for the Maltese Cross and

$$\alpha_l := 0.028 \text{ [m]}, \quad \alpha_u := 0.05 \text{ [m]}$$

for the O-Ring. Then, the set \mathcal{U} of admissible shapes is given by (5.3) and Lemma 5.1 holds.

Since from the practical point of view we cannot manufacture any shape, we will restrict ourselves to those that are described by a Bézier patch of a fixed number of design parameters

$$n_\Upsilon := n_{\Upsilon,1} \cdot n_{\Upsilon,2}, \quad \text{where } n_{\Upsilon,1}, n_{\Upsilon,2} \in \mathbb{N}$$

and we choose

$$n_{\Upsilon,1} := 4, \quad n_{\Upsilon,2} := 3.$$

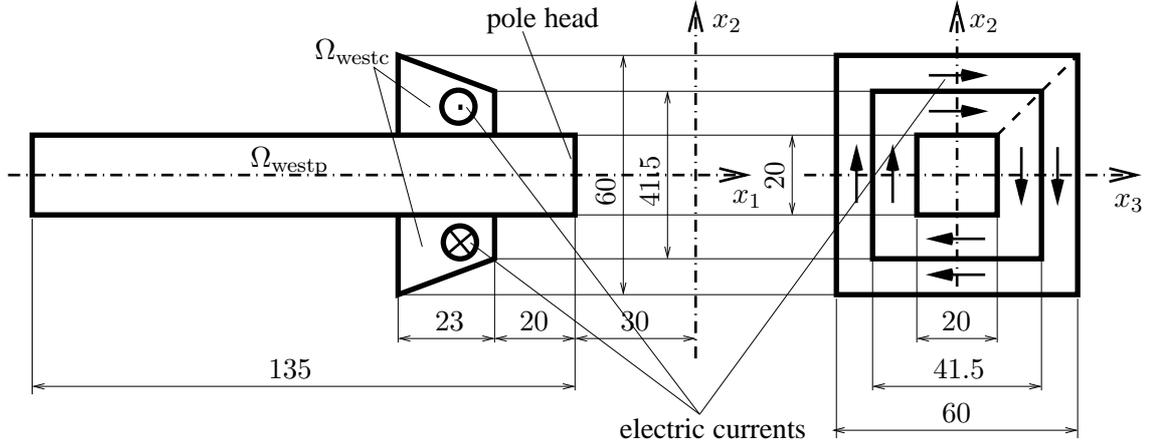


Figure 7.7: Detail drawing of the O-Ring west pole

Such shapes are by definition smooth enough. We decompose the domain ω into $(n_{\Upsilon,1} - 1)$ times $(n_{\Upsilon,2} - 1)$ regular rectangles whose $n_{\Upsilon,1}$ times $n_{\Upsilon,2}$ corners are

$$\mathbf{x}_{\omega,i,j} := \left(\frac{(i-1)d_{\text{pole},1}}{n_{\Upsilon,1}-1}, \frac{(j-1)d_{\text{pole},3}}{n_{\Upsilon,2}-1} \right) \quad \text{for } i = 1, \dots, n_{\Upsilon,1}, j = 1, \dots, n_{\Upsilon,2}.$$

The set Υ is defined as follows:

$$\Upsilon := \{ \mathbf{p} := (p_{1,1}, \dots, p_{1,n_{\Upsilon,2}}, \dots, p_{n_{\Upsilon,1},1}, \dots, p_{n_{\Upsilon,1},n_{\Upsilon,2}}) \in \mathbb{R}^{n_{\Upsilon}} \mid \alpha_1 \leq p_{i,j} \leq \alpha_u \}.$$

The mapping $F : \Upsilon \mapsto \mathcal{U}$, see also (5.5), is the following (tensor product) *Bézier mapping* that involves the symmetry

$$\begin{aligned} \alpha(x_1, x_3) &:= [F(x_1, x_3)](\mathbf{p}) := \\ &:= \sum_{i=1}^{n_{\Upsilon,1}} \sum_{j=1}^{n_{\Upsilon,2}} p_{i,j} \left[\beta_i^{2n_{\Upsilon,1}-1} \left(\frac{-2x_1 + d_{\text{pole},1}}{2d_{\text{pole},1}} \right) + \beta_i^{2n_{\Upsilon,1}-1} \left(\frac{2x_1 + d_{\text{pole},1}}{2d_{\text{pole},1}} \right) \right] \\ &\cdot \left[\beta_j^{2n_{\Upsilon,2}-1} \left(\frac{-2x_3 + d_{\text{pole},3}}{2d_{\text{pole},3}} \right) + \beta_j^{2n_{\Upsilon,2}-1} \left(\frac{2x_3 + d_{\text{pole},3}}{2d_{\text{pole},3}} \right) \right], \quad (x_1, x_3) \in \bar{\omega}, \end{aligned} \quad (7.1)$$

where for $n \in \mathbb{N}$, $i \in \mathbb{N}$, $i \leq n$, and $t \in \mathbb{R}$ such that $0 \leq t \leq 1$

$$\beta_i^n(t) := \frac{(n-1)!}{(i-1)!(n-i)!} t^{i-1} (1-t)^{n-i}, \quad (7.2)$$

which is called the *Bernstein polynom*. We can easily check that

$$\forall \mathbf{p} \in \Upsilon : [F(\cdot)](\mathbf{p}) \in \mathcal{U},$$

it means that both the relations (5.1) and (5.2) are fulfilled. An example of the mapping F is depicted in Fig. 7.8. Concerning (5.6), we perform mirroring of the shape α with respect to the planes $x_1 = 0$ and $x_3 = 0$ and, moreover, we copy this shape to all the remaining pole heads. In this way the shape α controls the decomposition of Ω into $\Omega_0(\alpha)$ that denotes the domain occupied by the coils or the air and into $\Omega_1(\alpha)$ which is the domain occupied by yoke and poles.

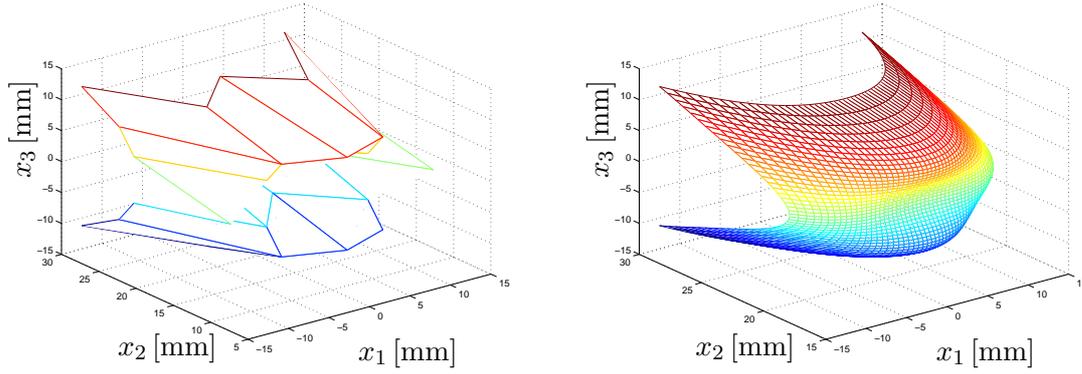


Figure 7.8: Bézier design parameters and the corresponding shape of the north pole head

7.2.3 Continuous multistate problem

Here, we are concerned with the 3–dimensional magnetostatics, i.e., with the differential operator

$$\mathbf{B} := \text{curl}.$$

Therefore, Assumptions 3.2–3.5 are satisfied by Lemmas 3.11–3.13 and by (3.31).

Now, we proceed through Section 5.2.2. We specify Assumption 5.1 by

$$\mathbf{D}_0 := \frac{1}{\mu_0}, \quad \mathbf{D}_1 := \frac{1}{\mu_1},$$

where $\mu_0 := 4\pi 10^{-7}$ [H.m⁻¹] and $\mu_1 := 5100\mu_0$ are the permeabilities of the air and the ferromagnetic parts, respectively.

Further, we consider

$$n^v := 2 \quad \text{and} \quad n^v := 3$$

variations of the current excitations in case of the Maltese Cross and the O–Ring electromagnet, respectively. In both the cases the right–hand side $\mathbf{f}^v(\mathbf{x})$ is calculated from the electric current I , from the number of turns n_I

$$I := 5 \text{ [A]} \quad \text{or} \quad I := 1.41 \text{ [A]}, \quad n_I := 600,$$

respectively, for the Maltese Cross or the O–Ring, and from the cross–section area through the coils, see also Figs. 7.5–7.7,

$$S_c := 0.03 \cdot 0.01 = 3 \cdot 10^{-4} \text{ [m}^2\text{]} \quad \text{or} \quad S_c := 0.023 \cdot \frac{0.02 + 0.01075}{2} = 4.6125 \cdot 10^{-4} \text{ [m}^2\text{]}.$$

The current densities $\mathbf{f}^v \equiv \mathbf{J}^v$ satisfy (5.2.2), i.e., they are divergence–free. The absolute value of $\mathbf{J}^v(\mathbf{x})$ is nonzero only in the subdomains $\Omega_{\text{westc}}, \dots, \Omega_{\text{southwestc}}$ where the direct electric currents are located

$$|\mathbf{J}^v(\mathbf{x})| = \frac{n_I I}{S_c}.$$

These subdomains are independent of the shape α , therefore, Assumption 5.2 is satisfied.

Now, we will describe the directions of $\mathbf{J}^v(\mathbf{x})$ for both the electromagnets and each variation v of the current excitation. During the description we will be referring to Fig. 7.7 and to Figs. 7.9–7.13. We say that two coils are *pumped (excited) in the same sense* if there is a magnetic circuit

that goes through both of them. Otherwise, the coils are *excited in the opposite sense*. In Fig. 7.9 the north and the south coils are excited in the same sense, the other two coils are switched off. In Fig. 7.10 all the coils are pumped by currents such that the west and the north one are in the same sense, the south and east one as well, but the west and the south one are in the opposite sense, as well as the north and the east one are. In Fig. 7.11 the north and the south coils are pumped in the same sense, the others are switched off. In Fig. 7.12 the situation is similar to Fig. 7.10 while the north–west, north–east, south–east, and south–west coils are switched off. Finally, in Fig. 7.13 the following 4 couples of coils are excited in the same sense: the south–west and south, the west and south–east, the north–west and east, and the north and north–east coil.



Figure 7.9: Magnetic flux lines for the *vertical current excitation* ($v := 1$) for the Maltese Cross

7.2.4 Continuous shape optimization problem

Now we shall specify the cost functional. Recall that we want to minimize inhomogeneities in the magnetic field in the area where the optical beam is magnetized such that the magnetic field is still strong enough. We will measure the inhomogeneities in the L^2 -norm which is, from the mathematical point of view, the most natural one. The magnetic field will not be allowed to decrease under some minimal magnitude which will be prescribed by a penalty term.

The magnetization area is for both the geometries

$$\Omega_m := [-0.005, 0.005] \times [-0.005, 0.005] \times [-0.005, 0.005] \text{ [m]}.$$

We choose the cost functional such that it measures differences of the magnetic flux density from its average value over the domain Ω_m . It is as follows:

$$\mathcal{I}(\mathbf{B}^1(\mathbf{x}), \dots, \mathbf{B}^{n_v}(\mathbf{x})) := \frac{1}{n_v} \sum_{v=1}^{n_v} [\varphi(\mathbf{B}^v(\mathbf{x})) + \rho \cdot \theta^v(\mathbf{B}^v(\mathbf{x}))], \quad (7.3)$$



Figure 7.10: Magnetic flux lines for the *diagonal current excitation* ($v := 2$) for the Maltese Cross

where

$$\mathbf{B}^v(\mathbf{x}) := \text{curl}_{\mathbf{x}}([\mathbf{u}^v(\alpha)](\mathbf{x})),$$

$$\varphi(\mathbf{B}^v(\mathbf{x})) := \frac{1}{\text{meas}(\Omega_m) (B_{\min}^{\text{avg},v})^2} \cdot \int_{\Omega_m} |\mathbf{B}^v(\mathbf{x}) - B_{\min}^{\text{avg},v}(\mathbf{B}^v(\mathbf{x})) \cdot \mathbf{n}_m^v|^2 d\mathbf{x}, \quad (7.4)$$

$$\theta^v(\mathbf{B}^v(\mathbf{x})) := (\max\{0, B_{\min}^{\text{avg},v} - B_{\min}^{\text{avg},v}(\mathbf{B}^v(\mathbf{x}))\})^2, \quad \rho := 10^6, \quad (7.5)$$

where $\mathbf{u}^v(\alpha)$ stands for the solution to $(W^v(\alpha))$ and where the following is the average magnetic flux density

$$B_{\min}^{\text{avg},v}(\mathbf{B}^v(\mathbf{x})) := \frac{1}{\text{meas}(\Omega_m)} \cdot \int_{\Omega_m} |\mathbf{B}^v(\mathbf{x}) \cdot \mathbf{n}_m^v| d\mathbf{x}. \quad (7.6)$$

Concerning the vectors \mathbf{n}_m^v , they are chosen as follows:

$$\mathbf{n}_m^v := \begin{cases} (0, 1, 0) & , v = 1 \\ (1/\sqrt{2}, 1/\sqrt{2}, 0) & , v = 2 \end{cases} \quad \text{or} \quad \mathbf{n}_m^v := \begin{cases} (0, 1, 0) & , v = 1 \\ (1/\sqrt{2}, 1/\sqrt{2}, 0) & , v = 2 \\ (\cos(\pi/8), -\sin(\pi/8), 0) & , v = 3 \end{cases}$$

in case of the Maltese Cross or the O-Ring, respectively. The minimal average magnetic flux densities are

$$B_{\min}^{\text{avg},1} := 0.1 \text{ [T]}, \quad B_{\min}^{\text{avg},2} := 0.15 \text{ [T]}$$

for both the geometries and, additionally, in case of O-Ring's super-diagonal excitation it is

$$B_{\min}^{\text{avg},3} := 0.3 \text{ [T]}.$$

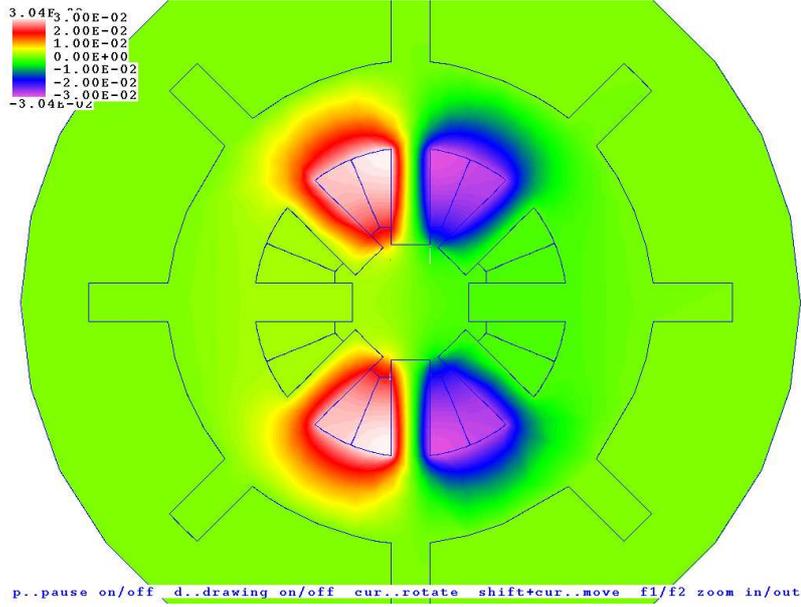


Figure 7.11: Magnetic flux lines for the *vertical current excitation* ($v := 1$) for the O-Ring

7.2.5 Regularization and finite element discretization

Once we choose a positive regularization parameter $\varepsilon > 0$, we have nothing more to specify concerning Section 5.3. Thus, we can proceed throughout Section 5.4. We choose a discretization parameter $h > 0$ such that

$$h \leq \bar{h},$$

where \bar{h} is the largest dimension in the geometry

$$\bar{h} := 0.4 \text{ [m]}.$$

To any discretization parameter h we associate a polyhedral domain Ω^h that is for the Maltese Cross $\Omega^h := \Omega$, while for the O-Ring it is like in Figs. 4.1, 7.11–7.13. Obviously, for both cases Assumption 4.3 is satisfied. Then we discretize the set of admissible shapes \mathcal{U} via a discretization \mathcal{T}_ω^h of the domain ω , as described in Section 5.4.1. Further, we discretize the polygonal computational domain Ω^h such that (5.24) holds. We provide the shape-to-mesh mapping by solving the auxiliary 3d discretized elasticity problem (6.2). Unfortunately, from a lot of numerical experiments we have learned that for slightly large shape deformations some elements flip. In this case we have to re-mesh the geometry, as noted in Remark 5.1.

We employ linear Nédelec tetrahedral elements that are described in Section 4.4.2. Therefore, Assumptions 4.1–4.2 and Assumptions 4.4–4.7 are satisfied whenever the discretization $\mathcal{T}^h(\alpha^h)$ satisfies the regularity condition (4.71).

For each $\alpha^h \in \mathcal{U}^h$ the permeability function is defined by (5.26). For any discretization parameter $h \leq \bar{h}$ the coil domains $\Omega_{\text{westc}}, \dots, \Omega_{\text{southwestc}}$ remain unchanged and their discretizations do not depend on α^h . This is guaranteed by the shape-to-mesh (elasticity) mapping (6.2) where we prescribe the homogeneous Dirichlet boundary condition on $\partial\Omega_{\text{westc}}, \dots, \partial\Omega_{\text{southwestc}}$. Therefore, Assumption 5.4 is true.

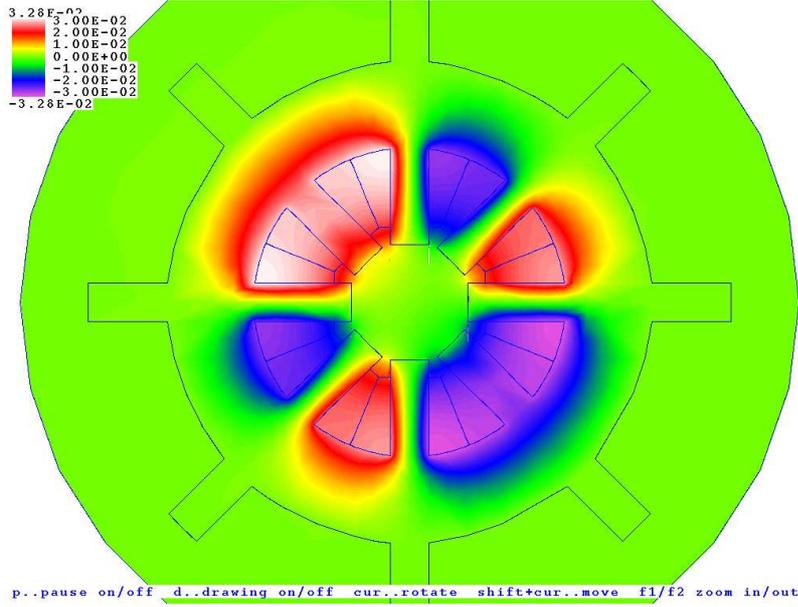


Figure 7.12: Magnetic flux lines for the *diagonal current excitation* ($v := 2$) for the O-Ring

Finally, the discretized (and regularized) cost functional is given by (5.36) and by the relations (7.3)–(7.6) which arrive at the following expressions

$$\mathcal{I}^h(\mathbf{B}_\varepsilon^{1,n}, \dots, \mathbf{B}_\varepsilon^{n_v,n}) := \frac{1}{n_v} \sum_{v=1}^{n_v} \left[\varphi^h(\mathbf{B}_\varepsilon^{v,n}) + \rho \cdot \theta^{v,h}(\mathbf{B}_\varepsilon^{v,n}) \right], \quad (7.7)$$

where $\mathbf{B}_\varepsilon^{v,n}$ is the elementwise constant magnetic field given by (6.7)–(6.8) and where

$$\varphi^h(\mathbf{B}_\varepsilon^{v,n}) := \frac{1}{\text{meas}(\Omega_m) (B_{\min}^{\text{avg},v})^2} \cdot \sum_{e \in E^h: K^e \subset \Omega_m} |\mathbf{B}_\varepsilon^{v,n,e} - B^{\text{avg},v,n}(\mathbf{B}_\varepsilon^{v,n,e}) \cdot \mathbf{n}_m^v|^2 \cdot \text{meas}(K^e), \quad (7.8)$$

$$\theta^{v,h}(\mathbf{B}_\varepsilon^{v,n}) := \left(\max \{0, B_{\min}^{\text{avg},v} - B^{\text{avg},v,n}(\mathbf{B}_\varepsilon^{v,n})\} \right)^2, \quad (7.9)$$

$$B^{\text{avg},v,n}(\mathbf{B}_\varepsilon^{v,n}) := \frac{1}{\text{meas}(\Omega_m)} \cdot \sum_{e \in E^h: K^e \subset \Omega_m} |\mathbf{B}_\varepsilon^{v,n,e} \cdot \mathbf{n}_m^v| \cdot \text{meas}(K^e). \quad (7.10)$$

In order to justify using a Newton-like optimization algorithm, we still have to satisfy Assumptions 6.1 and 6.2. Concerning Assumption 6.1, components of the constraint functional $\boldsymbol{\theta} : \mathbb{R}^{n_\Upsilon} \mapsto \mathbb{R}^{n_\theta}$, where $n_\theta := 2n_\Upsilon = 2n_{\Upsilon,1}n_{\Upsilon,2}$, are as follows:

$$\theta_k(\mathbf{p}) := \begin{cases} \alpha_1 - p_{i,j} & , i \leq n_{\Upsilon,1}, j \leq n_{\Upsilon,2} \\ p_{i-n_{\Upsilon,1}, j-n_{\Upsilon,2}} - \alpha_u & , i > n_{\Upsilon,1}, j > n_{\Upsilon,2} \end{cases} \quad \text{for } k = in_{\Upsilon,2} + j = 1, \dots, n_\theta$$

Hence, Assumption 6.1 is obviously satisfied. Now, we shall verify Assumption 6.2. The smoothness of the design-to-shape mapping F with respect to \mathbf{p} is easy to see from (7.1). Concerning the smoothness of the shape-to-mesh mapping \mathbf{x}^h , which is given by (6.2), it is well known that

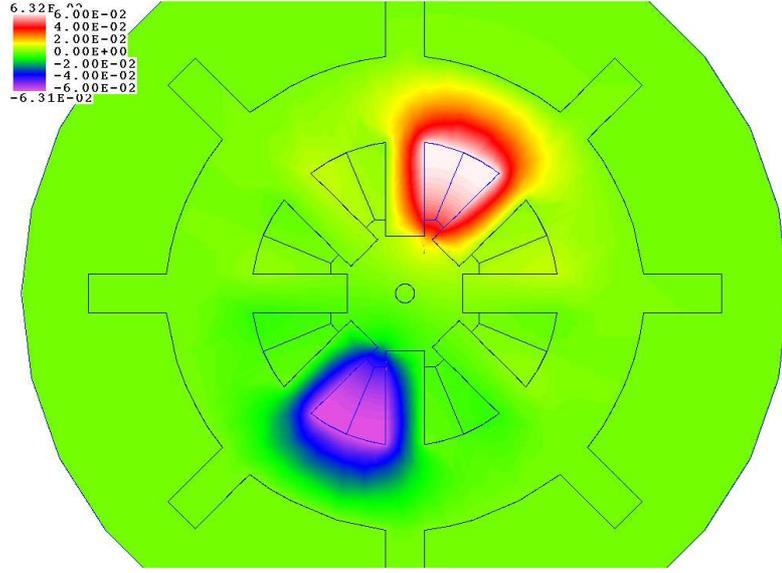


Figure 7.13: Magnetic flux lines for the *super-diagonal current excitation* ($v := 3$) for the O-Ring

the stiffness matrix $\mathbf{K}^h(\mathbf{x}_0)$ is nonsingular as far as we consider a Dirichlet boundary condition at a certain part of either the boundary $\partial\Omega^h$ or an interface $\partial\Omega_0^h(\alpha^h) \cap \partial\Omega_1^h(\alpha^h)$. Since $\mathbf{b}^h(\alpha^h)$ involves the following nonhomogeneous Dirichlet design interface boundary condition

$$\Delta \mathbf{x}^h = \mathcal{M}^h \cdot \alpha^h \text{ on } \Gamma_{\alpha^h},$$

where Γ_{α^h} denotes the design interface, then, \mathbf{x}^h , see (6.2), is smooth with respect to α^h . Further, due to (4.66), (4.68), and (4.69) we can see that each of $\mathbf{R}^e(\mathbf{x}^e)$, $\mathbf{S}^e(\mathbf{x}^e)$, and $\mathbf{S}_{\text{curl}}^e(\mathbf{x}^e)$, respectively, is smooth as far as K^e does not flip. The last item of Assumption 6.2 easily follows from (7.7)–(7.10).

7.3 Two-dimensional mathematical setting

Here, we reduce our mathematical model by neglecting the third dimension of the geometry, as described in Section 2.3. Thus, each 2-dimensional domain, denoted formally by Ω_{2d} , is created as the intersection of the related 3-dimensional domain Ω with the zero plane $\mathcal{Z} := \{\mathbf{x} \in \mathbb{R}^3 \mid x_3 = 0\}$ and by skipping the third component, i.e.,

$$\Omega_{2d} := \{\mathbf{x} := (x_1, x_2) \in \mathbb{R}^2 \mid (x_1, x_2, 0) \in \Omega\}.$$

We consider

$$\omega := \left(0, \frac{d_{\text{pole},1}}{2}\right).$$

The Lipschitz constant as well as the box constraints remain. The set of admissible shapes is given by (5.3).

Concerning Υ , we choose the following number of design parameters

$$n_{\Upsilon} := n_{\Upsilon,1} := 4,$$

which is the number of control Bézier nodes. The domain ω is decomposed into $(n_\Upsilon - 1)$ subintervals with the n_Υ nodes

$$\mathbf{x}_{\omega,i} := \frac{(i-1)d_{\text{pole},1}}{n_\Upsilon - 1} \quad \text{for } i = 1, \dots, n_\Upsilon.$$

The set Υ is as follows:

$$\Upsilon := \{\mathbf{p} := (p_1, \dots, p_{n_\Upsilon}) \in \mathbb{R}^{n_\Upsilon} \mid \alpha_l \leq p_i \leq \alpha_u\}.$$

The mapping $F : \Upsilon \mapsto \mathcal{U}$, which again involves the symmetry, reads

$$\alpha(x_1) := [F(x_1)](\mathbf{p}) := \sum_{i=1}^{n_\Upsilon} p_i \left[\beta_i^{2n_\Upsilon-1} \left(\frac{-2x_1 + d_{\text{pole},1}}{2d_{\text{pole},1}} \right) + \beta_i^{2n_\Upsilon-1} \left(\frac{2x_1 + d_{\text{pole},1}}{2d_{\text{pole},1}} \right) \right], \quad (7.11)$$

where $x_1 \in \bar{\omega}$ and β_i^n is given by (7.2). The mapping F is depicted in Fig. 7.14, where the red line connects the design parameters and the blue line is the resulting shape.

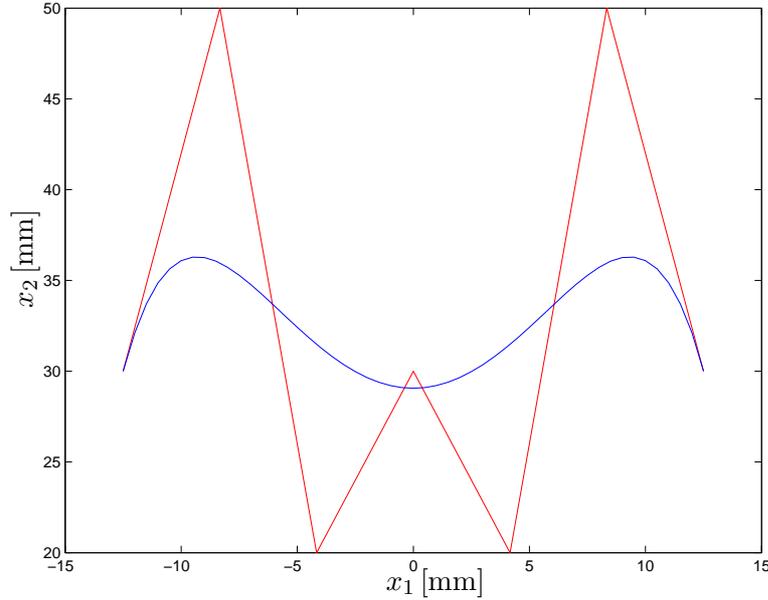


Figure 7.14: Bézier design parameters and the corresponding 2d shape of the north pole head

We concern the 2-dimensional magnetostatics with the differential operator

$$\mathbf{B} := \mathbf{grad}.$$

The space $H(\mathbf{grad}; \Omega_{2d})$ is equivalent to the space $H^1(\Omega_{2d})$, hence, Assumptions 3.2–3.5 are satisfied by Theorems 3.7–3.9 and by (3.27). Further, we can proceed throughout the rest of Section 7.2.4. We only recall that the compatibility condition (5.2.2) is satisfied, as $\text{Ker}(\mathbf{grad}; \Omega_{2d}) = \{0\}$.

As far as calculation of the 2d continuous cost functional is concerned, the magnetization area is

$$\Omega_m := [-0.005, 0.005] \times [-0.005, 0.005] \text{ [m]}$$

and the expressions (7.3)–(7.10) remain, where for $v = 1, \dots, n_v$ the vectors \mathbf{n}_m^v are as follows:

$$\mathbf{n}_m^v := \begin{cases} (0, 1) & , v = 1 \\ (1/\sqrt{2}, 1/\sqrt{2}) & , v = 2 \end{cases} \quad \text{or} \quad \mathbf{n}_m^v := \begin{cases} (0, 1) & , v = 1 \\ (1/\sqrt{2}, 1/\sqrt{2}) & , v = 2 \\ (\cos(\pi/8), -\sin(\pi/8)) & , v = 3 \end{cases}$$

in case of the Maltese Cross or the O–Ring electromagnet, respectively. The values of minimal magnetic flux densities remain as well.

Now, we do not need to introduce any regularization of the state problem, as the bilinear form is elliptic on the whole space $H_0(\mathbf{grad}; \Omega_{2d}) \equiv H_0^1(\Omega_{2d})$. Concerning the finite element discretization with a discretization parameter $h > 0$ such that $h \leq \bar{h}$, we approximate the domain Ω_{2d} by a polygonal domain Ω_{2d}^h , while for the Maltese Cross $\Omega_{2d}^h := \Omega_{2d}$ and for the O–Ring it is like in Figs. 7.11–7.13. Then, Assumption 4.3 holds. We use linear Lagrange elements that are described in Section 4.4.1. The discretization $\mathcal{T}^h(\alpha^h)$ satisfies the minimum angle condition (4.61) and, therefore, Assumptions 4.1–4.2 and Assumptions 4.4–4.7 are satisfied. The remaining specification of the 2d discretized shape optimization problem is as in Section 7.2.5. The only difference is that the smoothness of $\mathbf{R}^e(\mathbf{x}^e)$, $\mathbf{S}^e(\mathbf{x}^e)$, and $\mathbf{S}_{\mathbf{grad}}^e(\mathbf{x}^e)$ is now due to (4.56), (4.58), and (4.59), respectively.

7.4 Numerical results

In this section, we present numerical results for both 2d and 3d problems. In the optimization we employed the SQP algorithm with the BFGS update of the Hessian, see Section 6.2.2. For the calculation of gradients we used the first–order numerical differentiation. Moreover, we used a multilevel approach at 3 levels. The calculations were done using the scientific software tools Netgen, see SCHÖBERL [186], and Fepp, see KUHN, LANGER, AND SCHÖBERL [117], with an extension package for shape optimization, see LUKÁŠ, MÜHLHUBER, AND KUHN [125], which were all developed in the research project SFB F013 at the University Linz in Austria. All the calculations were run at the Department of Applied Mathematics, VŠB–Technical University Ostrava, Czech Republic, on a Linux PC machine with the processor Pentium III (1GHz) and 256MB of memory.

The optimized pole heads of the Maltese Cross electromagnet are depicted in Fig. 7.15 while the initial shape was a rectangle such that $\alpha_{\text{init}}(\mathbf{x}) := \alpha_u$. The 2d shape is described by 7 design

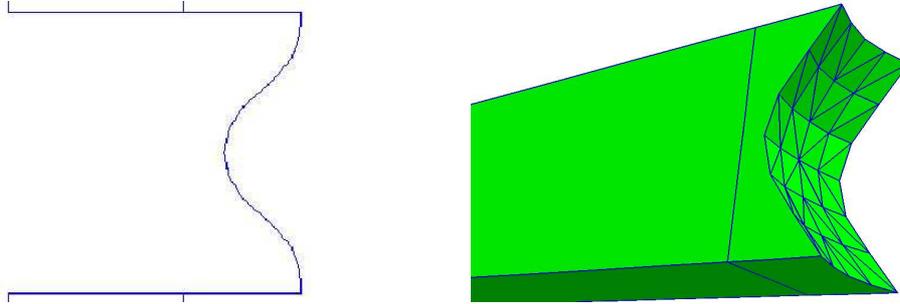


Figure 7.15: Optimized 2d and 3d pole heads of the Maltese Cross electromagnet

variables including the symmetry. Concerning discretization of the state problem, the discretiza-

tion parameters are

$$h := 0.05 \text{ [m]}, \quad h := 0.025 \text{ [m]}, \quad \text{and} \quad h := 0.0125 \text{ [m]}$$

at the first (coarsest), second, and third (finest) level, respectively. There are 12272 degrees of freedom at the last (3rd, finest) level. The optimization took 8 SQP iterations at the first (coarsest) level, 35 ones at the second level, and 25 ones at the third (finest) level, which was all done in 1 hour and 59 minutes, see also Fig. 7.17. The cost functional decreased from $1.97 \cdot 10^{-6}$ (1st level) to $1.49 \cdot 10^{-6}$ (3rd level). The 3d shape is determined by 12 design variables with the symmetry involved. The state problem at the finest level is discretized by 29541 degrees of freedom. Within the multilevel approach we made a step between the 2d and 3d model such that at the first level we had a coarse discretization of the 2d problem, at the second level we had a coarse discretization of the 3d problem, and at the last third level we had a fine discretization of the 3d problem. The calculation proceeded in 6, 50, and 37 SQP iterations at the respective levels, i.e., 93 SQP iterations in total, which took 29 hours and 46 minutes, see also Fig. 7.18. The cost functional decreased from $2.57 \cdot 10^{-6}$ (2nd level) to $7.32 \cdot 10^{-7}$ (3rd level).

The 2d optimized pole head of the O-Ring electromagnet is depicted in Fig. 7.16 while the initial shape was again a rectangle. The shape is described by 7 design variables including the

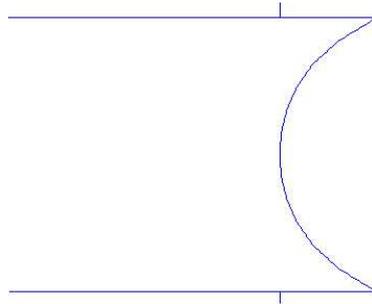


Figure 7.16: Optimized 2d pole head of the O-Ring electromagnet

symmetry. The state problem has 12005 degrees of freedom at the third (finest) level. The optimization took 19, 8, and 37 SQP iterations at the respective levels, which means 64 iterations in total and it was all done in 3 hours and 41 minutes. The cost functional decreased from $8.14 \cdot 10^{-4}$ (1st, coarsest level) to $2.87 \cdot 10^{-4}$ (3rd, finest level).

7.4.1 Testing the multilevel approach

Here, we present numerical tests of the multilevel optimization approach that was introduced in Section 6.4. We refer to Fig. 7.17, where we compare the multilevel approach with the classical one. We apply them to the 2d Maltese Cross optimization problem. From the last column in Fig. 7.17, we can see that the multilevel approach is much faster than the classical one. Using the multilevel approach, the calculation took about 2 hours while it took almost 7 hours, when using the classical approach.

In Fig. 7.18, a general multilevel approach is presented. It is tested on the 3d optimal shape design problem of the Maltese Cross electromagnet. At the first level ($h := 0.05 \text{ [m]}$), a coarse 2d optimization proceeds from the initial rectangular shape. The 2d coarse optimized shape from the first level is used as the initial guess at the second level, where a coarse ($h := 0.05 \text{ [m]}$), but now,

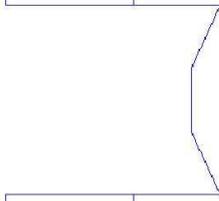
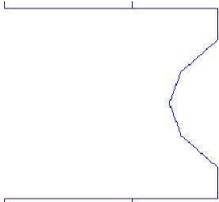
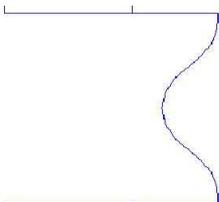
optimized designs	number of des. variables	number of unknowns	SQP iters.	CPU time
	2	1386	8	56s
	4	4705	47	27m 31s
		4970	53	30m 12s
	7	12272	72	1h 58m 55s
		12324	125	6h 44m 0s

Figure 7.17: Multilevel versus classical optimization approach

3d optimization is employed. This means that we have to prolong the 2d coarse optimized design into the third dimension by constant. This is the step that goes through the thick line in Fig. 7.18. Then, we proceed on, as we did in Fig. 7.17. We use the 3d coarse optimized design as the initial guess at the third level ($h := 0.025$ [m]). From the last line in Fig. 7.18, we can see that the whole calculation took almost 30 hours. We tried to compare this general multilevel approach with the classical one, but the calculation took more than 4 days and several re-meshings of the geometry had to be done. Unfortunately, in 4 days we were still not able to achieve the optimal solution, hence, the calculation was stopped.

7.4.2 Testing the adjoint method

Unfortunately, we have not finished the implementation of the adjoint method within the scientific software tool Fepp, see KUHN, LANGER, AND SCHÖBERL [117], yet. Despite of this fact, we provide a Matlab implementation, see LUKÁŠ [119], of the method, which is enclosed on the CD.

Let us consider a 2d academic optimization problem governed by linear magnetostatics. Its geometry is depicted in Fig. 7.19. Due to the symmetry and since we employ only one state problem, the computational domain is the top-left quarter

$$\Omega_{2d} := (-0.2, 0) \times (0, 0.1) \text{ [m]}.$$

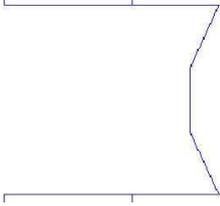
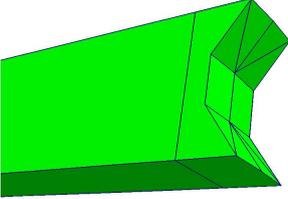
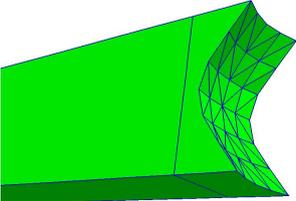
optimized designs	number of des. variables	number of unknowns	SQP iters.	CPU time
	2	2777	6	44s
	4	12086	47	7h 27m
	12	29541	93	29h 46m

Figure 7.18: A general multilevel optimization approach

The cost functional reads as follows:

$$\varphi(\mathbf{B}(\mathbf{x})) := \frac{1}{\text{meas}(\Omega_m) \|\mathbf{B}^{\text{req}}\|^2} \cdot \int_{\Omega_m} \|\mathbf{B}(\mathbf{x}) - \mathbf{B}^{\text{req}}\|^2 d\mathbf{x},$$

where we choose the required magnetic flux density

$$\mathbf{B}^{\text{req}} := (0.025, 0) \text{ [T]}.$$

The currents are located in the coil domains Ω_{westc} and Ω_{eastc} and the absolute value of the current density is

$$|J(\mathbf{x})| = 10^6 \text{ [A m}^{-2}\text{]}.$$

The box constraints are

$$\alpha_1 := -0.02 \text{ [m]}, \quad \alpha_u := 0.01 \text{ [m]}.$$

We discretize the problem with a discretization parameter

$$h := 0.01 \text{ [m]}.$$

The discretized grid and the solution to the state magnetostatic problem for the initial design are depicted in Fig. 7.20. Those for the optimized design are depicted in Fig. 7.21. The design is described by 4 variables, the state problem by 221 degrees of freedom. The cost functional has

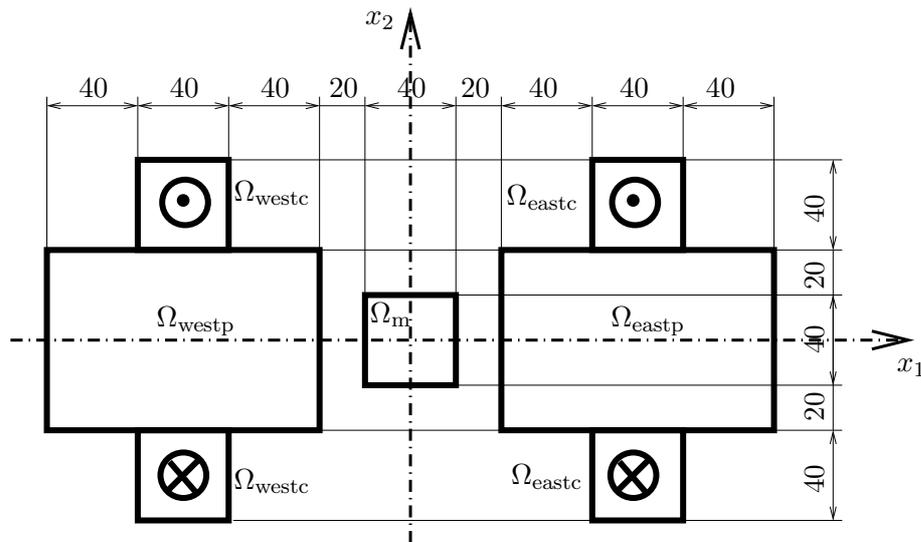


Figure 7.19: Geometry of the two-coils problem

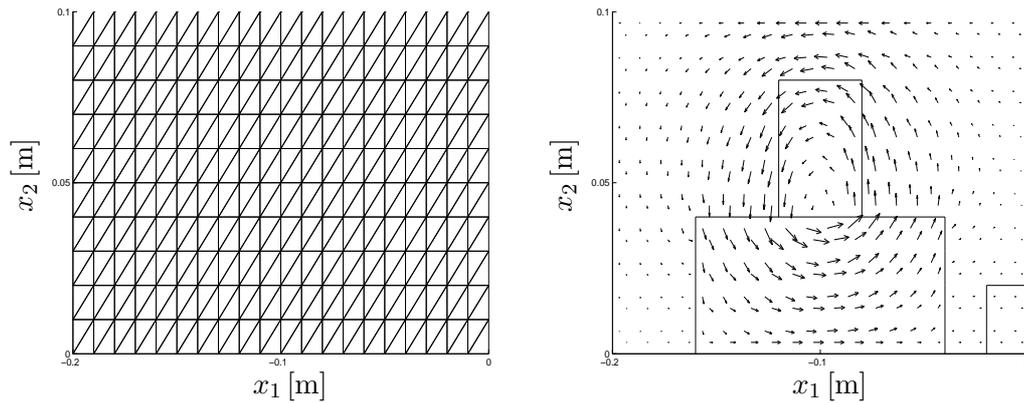


Figure 7.20: Initial design and the magnetic field of the two-coils problem

improved from 0.0077 to 0.0042. In Table 7.1 there is a comparison of the SQP method using the first-order numerical differentiation with the SQP method using the adjoint method for calculating gradients. Both the calculations took 4 SQP iterations. We can see that the SQP with the numerical differentiation needed 21 evaluations of the state problem while only 5 were needed by the adjoint method plus additional 4 evaluations of the adjoint state problem. In fact, the numerical differentiation took 5 evaluations of the cost functional plus additional 4 (number of design variables) times 4 (number of SQP iterations) evaluations, which give the total 21 evaluations. The cost functional was evaluated in about 18 seconds while the adjoint state problem in about 5 seconds. Enclosed there is a CD with the Matlab implementation, see also LUKÁŠ [119], where you can run

```
> optimization('n'); % numerical differentiation
> optimization('a'); % adjoint method
```

to see this comparison.

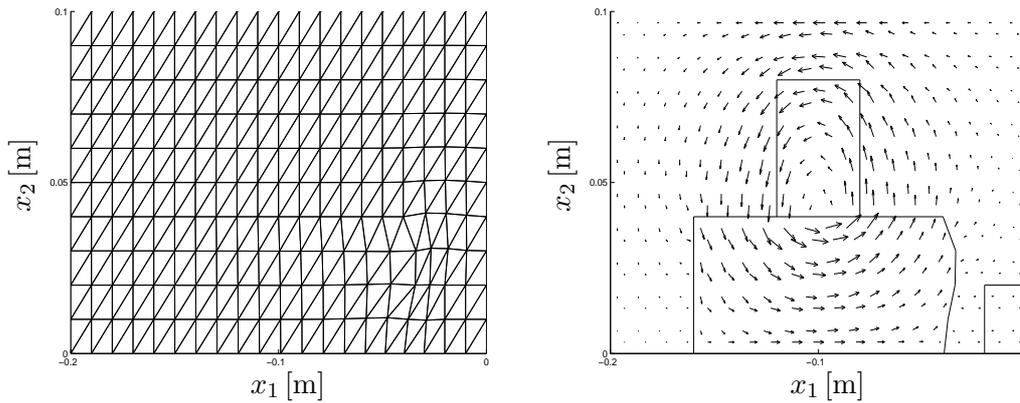


Figure 7.21: Optimized design and the magnetic field of the two-coils problem

	numerical differentiation	adjoint method
number of cost func. evals.	21	5
number of adjoint problem evals.	0	4
number of SQP iterations	4	4
CPU time	6min 26sec	1min 53sec

Table 7.1: Numerical differentiation versus the adjoint method

7.5 Manufacture and measurements

In my opinion, results of this section are the most highlight in this research. The calculated optimized shape was manufactured by the team of Prof. Pištora at the Institute of Physics, VŠB–Technical University Ostrava in the Czech Republic, and Dr. RNDr. Dalibor Ciprian measured the magnetic field for both the initial and optimized designs of the pole heads of the Maltese Cross electromagnet. These pole heads are depicted in Fig. 7.22.

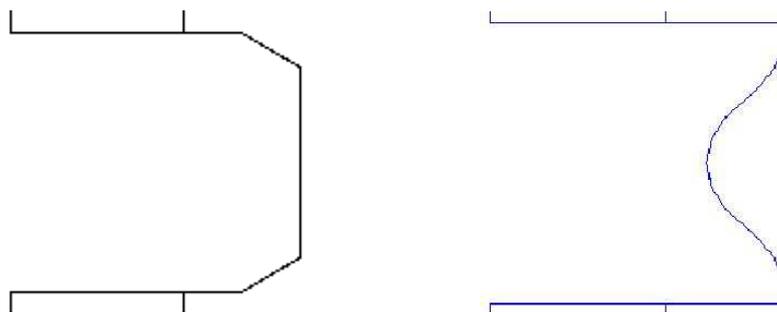


Figure 7.22: Initial and optimized 2d pole heads of the Maltese Cross electromagnet

In Fig. 7.23 there are distributions of the normal component of the magnetic flux density depicted. The blue solid line is the normal magnetic flux density along the magnetization plane for the diagonal excitation, see also Fig. 7.10, of the initial design, see Fig. 7.22. The red solid line is the normal magnetic flux density for the diagonal excitation of the optimized design. The blue and red dashed lines, respectively, are the normal magnetic flux densities along the magnetiza-

tion plane for the vertical excitation, see also Fig. 7.9, of the initial and optimized designs. In Fig. 7.23 we can see a significant improvement of the homogeneity of the magnetic field. The cost functional calculated from the measured data shows that it decreases 4.5-times. The cost functional calculated from the computer simulated magnetic field decreases only twice. The relative differences between the measured and the calculated magnetic fields are about 30%, which might be caused by saturation of the magnetic field in the corners. Employing a nonlinear governing magnetostatic state problem should improve also the mismatch of the magnetic fields. Nevertheless, the significant improvement of the cost functional shows that the optimization works well, no matter how big the nonlinearities in the direct magnetic field problem are.

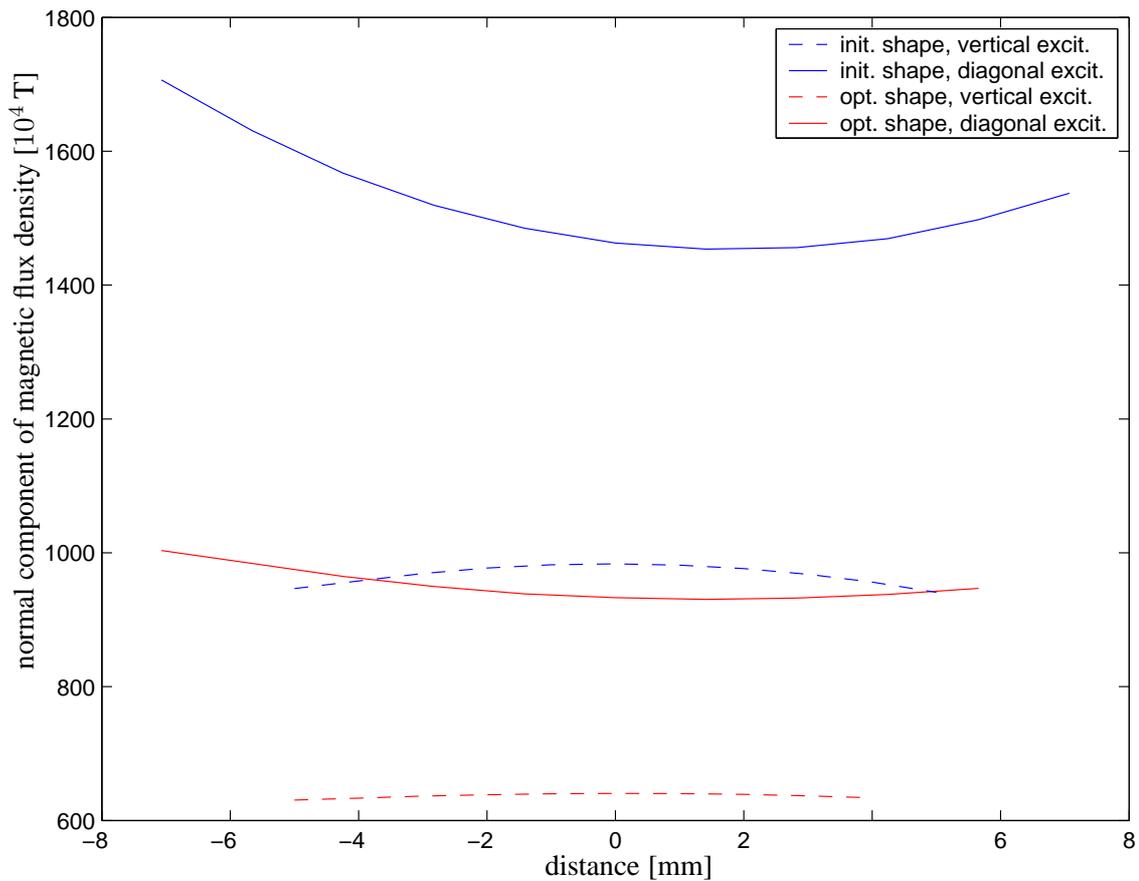


Figure 7.23: Magnetic field for the initial and optimized design of the MC electromagnet

Chapter 8

Conclusion

This thesis treated with the shape optimization in both two- and three-dimensional linear magnetostatics. The aim was to present a complete picture of the mathematical modelling process. We dealt with both the theoretical and computational aspects and demonstrated them on an application being of a practical purpose in the research on magneto-optic effects.

Let us summarize the main results obtained in the thesis.

- In Section 3.4 we developed an abstract theory for weak formulations of linear elliptic second-order boundary value problems (BVP).
- In Theorem 4.2 we proved the convergence of the solution of the finite element approximation to our abstract BVP while also dealing with an inner approximation of the original domain with the Lipschitz boundary by a sequence of domains with polyhedral (or polygonal) boundaries.
- In Sections 4.4.1 and 4.4.2 we concretized the abstract framework for the linear Lagrange and Nédélec elements on triangles or tetrahedra, respectively.
- In Chapter 5 we introduced an abstract shape optimization problem and its finite element approximation. We proved both the existence and convergence theorems while they rely on Lemma 5.3 and Theorem 4.2, respectively.
- In Section 6.3.4 there is the heart of the thesis. There we developed an efficient implementation of the adjoint method for the first-order sensitivity analysis. We provided also a Matlab implementation, which is enclosed on the CD.
- In Section 6.4 we introduced a multilevel optimization approach, which is a rather new technique. It is the first step towards adaptive optimization algorithms, as they have been recently presented in RAMM, MAUTE, AND SCHWARZ [164] and in SCHLEUPEN, MAUTE, AND RAMM [185]. The efficiency of our multilevel optimization approach was documented on numerical tests in Section 7.4.1.
- Finally, in Chapter 7 we presented a real-life application arising from the research on magneto-optic effects. We began with the physical description, went through the mathematical settings, and ended up with the manufacture of the optimized design and with the discussion of real improvements based on the physical measurements of the magnetic field.

In Chapter 5, we met one serious obstacle, see Remark 5.1, that the standard approximation theory does not completely cover problems of complex geometries. Namely, it is due to that we can hardly find a continuous mapping between the shape design nodes and the remaining nodes in the discretization grid. For fine discretizations and large changes in the design shape some elements flip. One possible outcome is in the use of the multilevel optimization techniques where on the fine grids the difference between the initial and optimized shapes is not that big. Another outcome might be when using composite finite elements that were developed for the treatment with complicated geometries in the papers by HACKBUSCH AND SAUTER [79, 80]. It is connected to an idea which was given to me in January 2002 by RNDr. Jan Chleboun, CSc. from the Mathematical Institute of the Czech Academy of Sciences. The idea is to use a fixed regular grid independent of the geometry and to resolve the fine details of the geometry within special elements that arise by the intersection of the geometry and the regular grid. This will move all the programming effort into the development of such special finite elements instead the shape-to-mesh mapping. We can also avoid this problem by using a boundary element discretization. From its matter, this is very suited for optimal shape design, as we need to handle only the boundary discretization. Nevertheless, construction of efficient multigrid solvers as well as using the method for nonlinear governing state problems are still topics of the current research.

Finally, let us draw the further directions of this research. They are mainly focused

- on development and rigorous analysis of the adaptive multilevel techniques in the shape optimization,
- on synergies among the inverse and shape optimization problems, namely, on the regularization techniques and numerical methods, e.g., the homogenization or level-set methods,
- on common aspects in the topology and shape optimization,
- on development of a user-friendly and well-documented scientific computing software tool for structural (both shape and topology) optimization,
- and on real-life applications in both electromagnetism and mechanics involving complex geometries and nonlinearities of the state problem, provided correct mathematical settings, i.e., the existence of a solution at least.

Bibliography

- [1] Ansys, <http://www.ansys.com>.
- [2] Fluent, <http://www.fluent.com>.
- [3] J. C. Adam, A. Gourdin-Serveniere, J. C. Nedelec, and P. A. Raviart, *Study of an implicit scheme for integrating Maxwell's equations*, Comput. Methods Appl. Mech. Eng. **22** (1980), 327–346.
- [4] R. A. Adams, *Sobolev spaces*, Academic Press, New York, 1975.
- [5] G. Allaire, *Homogenization and two-scale convergence*, SIAM J. Math. Anal. **23** (1992), 1482–1518.
- [6] ———, *Shape optimization by the homogenization method*, Springer, New York, Berlin, Heidelberg, 2002.
- [7] G. Allaire, E. Bonnetier, G. A. Francfort, and F. Jouve, *Shape optimization by the homogenization method*, Numer. Math. **76** (1997), 27–68.
- [8] G. Allaire, F. Jouve, and A.-M. Toader, *A level-set method for shape optimization*, C. R. Acad. Sci. Paris **334** (2002), 1125–1130.
- [9] C. Amrouche, C. Bernardi, M. Dauge, and V. Girault, *Vector potentials in three-dimensional non-smooth domains*, Math. Methods Appl. Sci. **21** (1998), 823–864.
- [10] T. Apel and J. Schöberl, *Multigrid methods for anisotropic edge refinement*, SIAM J. Numer. Anal. **40** (2002), no. 5, 1993–2006.
- [11] J. Arora, *Introduction to optimum design*, McGraw-Hill, New York, 1989.
- [12] J. P. Aubin, *Approximation of elliptic boundary-value problems*, Pure and Applied Mathematics, vol. 26, John Wiley & Sons, New York, 1972.
- [13] O. Axelsson, *Iterative solution methods*, Cambridge University Press, 1994.
- [14] I. Babuška and A. K. Aziz, *Survey lectures on the mathematical foundations of the finite element method*, The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, Academic Press, New York and London, 1972.
- [15] ———, *On the angle condition in the finite element method*, SIAM J. Numer. Anal. **13** (1976), 214–226.

- [16] P. K. Banerjee and R. Butterfield, *Boundary element methods in engineering science*, McGraw–Hill Book Company, 1981.
- [17] N. V. Banichuk, *Introduction to optimization of structures*, Springer, New York, 1990.
- [18] P. Di Barba, P. Navarra, A. Savini, and R. Sikora, *Optimum design of iron–core electromagnets*, IEEE Transactions on Magnetics **26** (1990), 646–649.
- [19] D. Begis and R. Glowinski, *Application de la méthode des éléments finis à la résolution d'un problème de domaine optimal*, Appl. Math. Optimization **2** (1975), 130–169.
- [20] M. P. Bendsøe, *Optimal shape design as a material distribution problem*, Struct. Multidisc. Optim. **1** (1989), 193–202.
- [21] _____, *Optimization of structural topology, shape and material*, Springer, Berlin, Heidelberg, 1995.
- [22] M. P. Bendsøe and O. Sigmund, *Topology optimization. Theory, methods and applications*, Springer, Berlin Heidelberg, 2003.
- [23] P. T. Boggs and J. W. Tolle, *Sequential quadratic programming*, Acta Numer. **4** (1995), 1–51.
- [24] A. D. Börner, *Mathematische Modelle für Optimum Shape Design Probleme in der Magnetostatik*, Ph.D. thesis, Universität Hamburg, Germany, 1985.
- [25] T. Borrvall, *Computational topology optimization of elastic continua by design restrictions*, Ph.D. thesis, Linköping University, Sweden, 2000.
- [26] A. Bossavit, *Computational electromagnetism. Variational formulations, complementarity, edge elements*, Orlando, FL: Academic Press, 1998.
- [27] D. Braess, *Finite elements*, Cambridge, 2001.
- [28] J. H. Bramble, *Multigrid methods*, Pitman Research Notes in Mathematics, vol. 294, John Wiley & Sons, 1993.
- [29] J. H. Bramble, J. E. Pasciak, and J. Xu, *Parallel multilevel preconditioners*, Math. Comput. **55** (1990), 1911–1922.
- [30] R. B. Brandtstätter, W. Ring, Ch. Magele, and K. R. Richter, *Shape design with great geometrical deformations using continuously moving finite element nodes*, IEEE Transactions on Magnetics **34** (1998), no. 5, 2877–2880.
- [31] C. A. Brebbia (ed.), *Topics in boundary element research, vol. 6 – electromagnetic applications*, Springer, Berlin, 1989.
- [32] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, Springer, 1994.
- [33] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, Springer, New York, 1991.

- [34] R. A. Brockman, *Geometric sensitivity analysis with isoparametric finite elements*, Commun. Appl. Numer. Methods **3** (1987), 495–499.
- [35] D. Bucur and J.-P. Zolesio, *Existence and stability of the optimum in shape optimization*, ZAMM **76** (1996), no. 2, 77–80.
- [36] M. Burger, *A level set method for inverse problems*, Inverse Probl. **17** (2001), no. 5, 1327–1355.
- [37] M. Burger and W. Mühlhuber, *Iterative regularization of parameter identification problems by sequential quadratic programming methods*, Inverse Probl. **18** (2002), no. 4, 943–969.
- [38] ———, *Numerical approximation of an SQP-type method for parameter identification*, SIAM J. Numer. Anal. **40** (2002), no. 5, 1775–1797.
- [39] J. Cea, *Optimization, théorie et algorithmes*, Dunod, Paris, 1971.
- [40] J. Cea, S. Garreau, P. Guillaume, and M. Masmoudi, *The shape and topological optimization connections*, Comput. Methods Appl. Mech. Eng. **188** (2000), 713–726.
- [41] G. Chen and J. Zhou, *Boundary element methods*, Computational Mathematics and Applications, Academic Press, Harcourt Brace Jovanovich, London, San Diego, New York, 1992.
- [42] G. Chen, J. Zhou, and R. C. McLean, *Boundary element method for shape (domain) optimization of linear–quadratic elliptic boundary control problem*, Boundary Control and Variation. Proceedings of the 5th working conference held in Sophia Antipolis, France, Inc. Lect. Notes Pure Appl. Math., vol. 163, 1994, pp. 27–72.
- [43] A. V. Cherkaev, *Variational methods for structural optimization*, Springer, New York, Berlin, Heidelberg, 2000.
- [44] J. Chleboun and R. Mäkinen, *Primal hybrid formulation of an elliptic equation in smooth optimal shape problems*, Adv. Math. Sci. Appl. **5** (1995), no. 1, 139–162.
- [45] P. Ciarlet, *The finite element method for elliptic problems*, North–Holland, 1978.
- [46] P. G. Ciarlet and J. Lions (eds.), *Handbook of numerical analysis*, North–Holland, Amsterdam, 1989.
- [47] D. Cioranescu and P. Donato, *An introduction to homogenization*, Oxford Lect. Series Math. Appl., vol. 17, Oxford University Press, 1999.
- [48] F. H. Clarke, *Nonsmooth analysis and optimization*, Wiley & Sons, 1983.
- [49] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust–region methods*, MPS–SIAM Series on Optimization, vol. 1, SIAM, Philadelphia, 2000.
- [50] M. Costabel and M. Dauge, *Singularities of Maxwell’s equations on polyhedral domains*, Numerics and applications of differential and integral equations (M. Bach, C. Constanda, and G. C. Hsiao et. al., eds.), Pitman Research Notes in Mathematics Series, vol. 379, Harlow: Adison Wesley Longman, 1998.

- [51] R. Courant and D. Hilbert, *Methods of mathematical physics*, Interscience Publishers, New York, 1962.
- [52] W. C. Davidon, *Variable metric method for minimization*, Technical report ANL-5990 (revised), Argonne National Laboratory, 1959.
- [53] ———, *Variable metric method for minimization*, SIAM J. Optim. **1** (1991), 1–17.
- [54] M. C. Delfour and J.-P. Zolesio, *Shape and geometries: Analysis, differential calculus, and optimization*, Advances in Design and Control, vol. 4, SIAM, 2001.
- [55] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice Hall, 1983.
- [56] P. Doktor, *On the density of smooth functions in certain subspaces of Sobolev spaces*, Comment. Math. Univ. Carol. **14** (1973), 609–622.
- [57] Z. Dostál, *Linear algebra*, Lecture Notes, VŠB–Technical University of Ostrava, 2000, in Czech.
- [58] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1996.
- [59] G. E. Farin, *Curves and surfaces for computer-aided geometric design: A practical guide*, Academic Press, 1996.
- [60] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics, vol. 3*, Addison–Wesley, 1966.
- [61] R. Fletcher, *Practical optimization methods*, 2nd edition, Wiley & Sons, 1987.
- [62] ———, *An optimal positive definite update for sparse Hessian matrices*, SIAM J. Optim. **5** (1995), 192–218.
- [63] R. Fletcher and C. M. Reeves, *Function minimization by conjugate gradients*, Comput. J. **7** (1964), 149–154.
- [64] J. Franců, *Monotone operators. A survey directed to applications to differential equations*, Appl. Math. Praha **35** (1990), 257–301.
- [65] P. L. George, *Automatic mesh generation*, John Wiley & Sons, 1993.
- [66] P. Gill, W. Murray, and M. Wright, *Practical optimization*, Academic Press, 1981.
- [67] V. Girault and P. Raviart, *Finite element methods for Navier–Stokes equations*, Springer, Berlin, 1986.
- [68] R. Glowinski, *Numerical methods for nonlinear variational problems*, Springer, New York, 1984.
- [69] G. H. Golub and C. F. van Loan, *Matrix computation*, 2nd ed., Johns Hopkins Series in the Mathematical Sciences, The Johns Hopkins University Press, 1989.
- [70] A. Griewank, *Evaluating derivatives, principles and techniques of algorithmic differentiation*, Frontiers in Applied Mathematics, vol. 19, SIAM, Philadelphia, 2000.

- [71] C. Grossmann and H.-G. Ross, *Numerik partieller Differentialgleichungen*, Teubner Studienbücher, Stuttgart, 1992.
- [72] C. Grossmann and J. Terno, *Numerik der Optimierung*, 2nd ed., Teubner Studienbücher, Stuttgart, 1997.
- [73] G. Haase, M. Kuhn, U. Langer, S. Reitzinger, and J. Schöberl, *Parallel Maxwell solvers*, Scientific computing in electrical engineering (Ursula van Rienen, Michael Günther, and Dirk Hecht, eds.), Lect. Notes Comput. Sci. Eng., vol. 18, 2001, pp. 71–78.
- [74] G. Haase and U. Langer, *On the use of multigrid preconditioners in the domain decomposition method*, Parallel algorithms for partial differential equations, Proc. 6th GAMM-Semin., Notes Numer. Fluid Mech., vol. 31, 1991, pp. 101–110.
- [75] G. Haase, U. Langer, A. Meyer, and S.V. Nepomnyaschikh, *Hierarchical extension operators and local multigrid methods in domain decomposition preconditioners*, East West J. Numer. Math. **31** (1994), no. 4–5, 269–272.
- [76] G. Haase, U. Langer, S. Reitzinger, and J. Schöberl, *Algebraic multigrid methods based on element preconditioning*, Int. J. Comput. Math. **78** (2001), no. 4, 575–598.
- [77] G. Haase and E. Lindner, *Advanced solving techniques in optimization of machine components*, Comput. Assist. Mech. Eng. Sci. **6** (1999), no. 3, 337–343.
- [78] W. Hackbusch, *Multi-grid methods and applications*, Springer, Berlin, 1985.
- [79] W. Hackbusch and S. A. Sauter, *Composite finite elements for problems containing small geometric details - part II: Implementation and numerical results*, Comput. Vis. Sci. **1** (1997), 15–25.
- [80] ———, *Composite finite elements for the approximation of PDEs on domains with complicated micro-structures*, Numer. Math. **75** (1997), 447–472.
- [81] W. W. Hager, D. W. Hearn, and P. M. Pardalos (eds.), *Large scale optimization*, Kluwer Academic Publishers, Dordrecht, 1994.
- [82] J. S. Hansen, Z. S. Liu, and N. Olhoff, *Shape sensitivity analysis using a fixed basis function finite element approach*, Struct. Multidisc. Optim. **21** (2001), 177–196.
- [83] J. Haslinger and R. A. E. Mäkinen, *Introduction to shape optimization. Theory, approximation, and computation*, Advances in Design and Control, vol. 7, SIAM, Philadelphia, 2003.
- [84] J. Haslinger, M. Miettinen, and D. Panagiotopoulos, *Finite element method for hemivariational inequalities. Theory, methods and applications*, Nonconvex Optimization and its Applications, vol. 35, Kluwer Academic Publishers, Dordrecht, 1999.
- [85] J. Haslinger and P. Neittaanmäki, *Finite element approximation for optimal shape, material and topology design*, 2nd ed., John Wiley & Sons Ltd., Chichester, 1997.
- [86] E. J. Haug, K. K. Choi, and V. Komkov, *Sensitivity analysis in structural design*, Mathematics in Science and Engineering, vol. 177, Academic Press, 1986.

- [87] H. A. Haus and J. R. Melcher, *Electromagnetic fields and energy*, Englewood Cliffs, Prentice Hall, 1989.
- [88] M. R. Hestensen, *Optimization theory*, John Wiley & Sons, 1975.
- [89] ———, *Conjugate direction methods in optimization*, Springer, 1980.
- [90] M. R. Hestensen and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Inst. Stand. Technol. **49** (1952), 409–436.
- [91] R. Hiptmair, *Multilevel preconditioning for mixed problems in three dimensions*, Ph.D. thesis, University of Augsburg, Germany, 1996.
- [92] ———, *Canonical construction of finite elements*, Math. Comput. **68** (1999), 1325–1346.
- [93] ———, *Finite elements in computational electromagnetism*, Acta Numer. (2002), 237–339.
- [94] R. Hiptmair and C. Schwab, *Natural boundary element methods for the electric field integral equation on polyhedra*, SIAM J. Numer. Anal. **40** (2002), no. 1, 66–86.
- [95] I. Hlaváček, J. Haslinger, J. Nečas, and J. Lovíšek, *Numerical solution of variational inequalities*, Springer, Berlin, 1988.
- [96] R. H. W. Hoppe, S. I. Petrova, and V. H. Schulz, *Numerical analysis and applications – topology optimization of conductive media described by Maxwell’s equations*, Lecture Notes Comput. Sci. (1988), 414–422.
- [97] ———, *3d structural optimization in electromagnetics*, Proceedings of Domain Decomposition Methods and Applications (Lyon, Oct. 9–12, 2000) (M. Garbey et al., ed.), 2001.
- [98] T. J. R. Hughes, *The finite element method*, Prentice–Hall, Englewood Cliffs, New Jersey, 1987.
- [99] N. Ida and J. P. A. Bastos, *Electromagnetics and calculation of fields*, Springer, 1997.
- [100] J. B. Jacobsen, N. Olhoff, and E. Rønholt, *Generalized shape optimization of three-dimensional structures using materials with optimum microstructures*, Mechanics of Materials **28** (1998), no. 1–4, 207–225.
- [101] C. Johnson, *Numerical solution of partial differential equations by the finite element method*, Cambridge University Press, 1990.
- [102] M. Jung and U. Langer, *Applications of multilevel methods to practical problems*, Surv. Math. Ind. **1** (1991), no. 3, 217–257.
- [103] ———, *Methode der finiten Elemente für Ingenieure: Eine Einführung in die numerischen Grundlagen und Computersimulation*, Teubner, Stuttgart, 2001.
- [104] J. Kačur, J. Nečas, J. Polák, and J. Souček, *Convergence of a method for solving the magnetostatic field in nonlinear media*, Appl. Math. Praha **13** (1968), 456–465.
- [105] A. M. Kalamkarov and A. G. Kolpakov, *Analysis, design and optimization of composite structures*, John Wiley & Sons, New York; London; Sydney, 1997.

- [106] M. Kaltenbacher, H. Landes, R. Lerch, and F. Lindinger, *A finite–element / boundary–element method for the simulation of coupled electrostatic–mechanical systems*, J. Physique III **7** (1997), 1975–1982.
- [107] B. Kawohl, O. Pironneau, L. Tartar, and J.-P. Zolesio (eds.), *Optimal shape design*, Lecture Notes in Mathematics, vol. 1740, Springer, Berlin, New York, Heidelberg, 2000.
- [108] N. Kikuchi, *Finite element methods in mechanics*, Cambridge University Press, 1986.
- [109] A. Kirsch, *An introduction to the mathematical theory of inverse problems*, Springer Series in Applied Mathematical Sciences, vol. 120, Springer, Berlin, 1996.
- [110] E. Kita and H. Tanie, *Topology and shape optimization of continuum structures using GA and BEM*, Struct. Multidisc. Optim. **17** (1999), no. 2–3, 130–139.
- [111] I. Kopřiva, D. Hrabovský, K. Postava, D. Ciprian, and J. Pištorá, *Anisotropy of the quadratic magneto-optical effects in a cubic crystal*, Proceedings of SPIE, vol. 4016, 2000, pp. 54–59.
- [112] A. Kost, *Numerische Methode in der Berechnung elektromagnetischer Felder*, Springer, Berlin, Heidelberg, New York, 1995.
- [113] M. Křížek, *On the maximum angle condition for linear tetrahedral elements*, SIAM J. Numer. Anal. **29** (1992), 513–520.
- [114] M. Křížek and P. Neittaanmäki, *On superconvergence techniques*, Acta Appl. Math. **9** (1987), 175–198.
- [115] M. Křížek and P. Neittaanmäki, *Mathematical and numerical modelling in electrical engineering*, Kluwer Academic Publishers, Dordrecht, 1996.
- [116] A. Kufner, O. John, and S. Fučík, *Function spaces*, Academia, Prague, 1977.
- [117] M. Kuhn, U. Langer, and J. Schöberl, *Scientific computing tools for 3d magnetic field problems*, The Mathematics of Finite Elements and Applications (MAFELAP X) (2000), 239–259.
- [118] E. Laporte and P. Le Tallec, *Numerical methods in sensitivity analysis and shape optimization*, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser, Boston, 2003.
- [119] D. Lukáš, *2–d shape optimization for magnetostatics*, Tutorial for a students project at the conference Industrial Mathematics and Mathematical Modelling (IMAMM) 2003, Rožnov pod Radhoštěm, Czech Republic, <http://lukas.am.vsb.cz/Teaching/IMAMM2003/>.
- [120] _____, *On solution to an optimal shape design problem in 3-dimensional magnetostatics*, Appl. Math. Praha, Submitted.
- [121] _____, *On the road – between Sobolev spaces and a manufacture of homogeneous electromagnets*, Transactions of the VŠB–TU of Ostrava, vol. 2, VŠB–TU Ostrava, To appear.
- [122] _____, *Shape optimization of homogeneous electromagnets*, Tech. report, Specialforschungsbereich SFB F013, University Linz, 2000.

- [123] ———, *Shape optimization of homogeneous electromagnets*, Scientific Computing in Electrical Engineering (Ursula van Rienen, Michael Günther, and Dirk Hecht, eds.), Lect. Notes Comp. Sci. Eng., vol. 18, Springer, 2001, pp. 145–152.
- [124] D. Lukáš, I. Kopřiva, D. Ciprian, and J. Pištorá, *Shape optimization of homogeneous electromagnets and their application to measurements of magneto-optic effects*, Records of COM-PUMAG 2001, vol. 4, 2001, pp. 156–157.
- [125] D. Lukáš, W. Mühlhuber, and M. Kuhn, *An object-oriented library for the shape optimization problems governed by systems of linear elliptic partial differential equations*, Transactions of the VŠB-TU of Ostrava, vol. 1, VŠB-TU Ostrava, 2001, pp. 115–128.
- [126] D. Lukáš, J. Vlček, Z. Hytka, and Z. Dostál, *Optimization of dislocation of magnets in a magnetic separator*, Transactions of the VŠB-TU of Ostrava, Electrical Engineering Series, vol. 1, VŠB-TU Ostrava, 1999, pp. 1–12.
- [127] D. Lukáš, *Shape optimization of a magnetic separator*, Master's thesis, VŠB - Technical University of Ostrava, Czech Republic, 1999, In Czech.
- [128] D. Lukáš, *Multilevel solvers for 3-dimensional optimal shape design with an application to magnetostatics*, Book of Abstracts of the 9th International Symposium on Microwave and Optical Technology, Ostrava, 2003, p. 121.
- [129] J. Lukeš and J. Malý, *Measure and integral*, MATFYZPRESS, 1995.
- [130] M. Mäkelä and P. Neittaanmäki, *Nonsmooth optimization: Analysis and algorithms with applications to the optimal control*, World Scientific Publishing Company, Singapore, 1992.
- [131] R. Mäkinen, *Finite element design sensitivity analysis for nonlinear potential problems*, Commun. Appl. Numer. Methods **6** (1986), 343–350.
- [132] A. Marrocco and O. Pironneau, *Optimum design with Lagrangian finite elements: Design of an electromagnet*, Comput. Methods Appl. Mech. Eng. **15** (1978), 277–308.
- [133] I. D. Mayergoyz, *Mathematical models of hysteresis*, Springer, Berlin, 1991.
- [134] V. G. Mazya, *Sobolev spaces*, Springer, Berlin, 1985.
- [135] B. Mohammadi and O. Pironneau, *Applied shape optimization for fluids*, Oxford University Press, Oxford, 2001.
- [136] P. Monk, *Analysis of a finite element method for Maxwell's equations*, SIAM J. Numer. Anal. **29** (1992), 714–729.
- [137] ———, *A comparison of three mixed methods for the time dependent Maxwell's equations*, SIAM J. Sci. Statist. Comput. **13** (1992), 1097–1122.
- [138] ———, *A finite element method for approximating the time-harmonic Maxwell's equations*, Numer. Math. **63** (1992), 243–261.
- [139] W. Mühlhuber, *Efficient solvers for optimal design problems with PDE constraints*, Ph.D. thesis, Johannes Kepler University Linz, Austria, 2002.

- [140] F. Murat and J. Simon, *Studies on optimal shape design problems*, Lecture Notes in Computational Science, vol. 41, Springer, Berlin, 1976.
- [141] J. Nečas, *Les méthodes directes en théorie des équations elliptiques*, Academia, Prague, 1967.
- [142] J. C. Nédélec, *Mixed finite elements in \mathbf{R}^3* , Numer. Math. **35** (1980), 315–341.
- [143] ———, *A new family of mixed finite elements in \mathbf{R}^3* , Numer. Math. **50** (1986), 57–81.
- [144] P. Neittaanmäki, M. Rudnicki, and A. Savini, *Inverse problems and optimal design in electricity and magnetism*, Oxford University Press, 1995.
- [145] P. Neittaanmäki and K. Salmenjoki, *Sensitivity analysis for optimal shape design problems*, Struct. Multidisc. Optim. **1** (1989), 241–251.
- [146] P. Neittaanmäki and J. Saranen, *Finite element approximation of electromagnetic fields in the three dimensional space*, Numer. Funct. Anal. Optim. **2** (1981), 487–506.
- [147] ———, *Finite element approximations of vector fields given by curl and divergence*, Math. Methods Appl. Sci. **3** (1981), 328–335.
- [148] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer, 1999.
- [149] J. T. Oden and L. F. Demkowicz, *Applied functional analysis*, CRC Press, 1996.
- [150] J. T. Oden and J. N. Reddy, *An introduction to the mathematical theory of finite elements*, John Wiley & Sons, 1976.
- [151] N. Olhoff, *Multicriterion structural optimization via bound formulation and mathematical programming*, Struct. Multidisc. Optim. **1** (1989), 11–17.
- [152] N. Olhoff and J. E. Taylor, *On structural optimization*, J. Appl. Math. **50** (1983), no. 4, 1139–1151.
- [153] A. R. Parkinson and R. J. Balling, *The OptdesX design optimization software*, Struct. Multidisc. Optim. **23** (2002), 127–139.
- [154] P. Pedersen (ed.), *Optimal design with advanced materials*, Elsevier, Amsterdam, The Netherlands, 1993.
- [155] G. H. Peichl and W. Ring, *Optimization of the shape of an electromagnet: Regularity results*, Adv. Math. Sci. and Appl. **8** (1998), 997–1014.
- [156] ———, *Asymptotic commutativity of differentiation and discretization in shape optimization*, Math. and Comput. Modelling **29** (1999), 19–37.
- [157] J. Petersson, *On continuity of the design-to-state mappings for trusses with variable topology*, Int. J. Eng. Sci. **39** (2001), no. 10, 1119–1141.
- [158] J. Petersson and J. Haslinger, *An approximation theory for optimum sheets in unilateral contact*, Q. Appl. Math. **56** (1998), no. 2, 309–332.
- [159] O. Pironneau, *Optimal shape design for elliptic systems*, Springer Series in Computational Physics, Springer, New-York, 1984.

- [160] J. Pištora, K. Postava, and R. Šebesta, *Optical guided modes in sandwiches with ultrathin metallic films*, *Journal of Magnetism and Magnetic Materials* **198–199** (1999), 683–685.
- [161] E. Polak, *Computational methods in optimization. A unified approach*, Academic Press, New York, 1971.
- [162] E. Polak and G. Ribière, *Note sur la convergence de méthodes de directions conjuguées*, *Rev. Fr. Inf. Rech. Oper.* **16** (1969), 35–43.
- [163] A. Ralston, *Základy numerické matematiky (Fundamentals of numerical mathematics)*, Academia Praha, 1978.
- [164] E. Ramm, K. Maute, and S. Schwarz, *Adaptive topology and shape optimization*, 1998, <http://citeseer.nj.nec.com/ramm98adaptive.html>.
- [165] J. Rasmussen, M. Damsgaard, S. T. Christensen, and E. Surma, *Design optimization with respect to ergonomic properties*, *Struct. Multidisc. Optim.* **24** (2002), 89–97.
- [166] J. Rasmussen, E. Lund, T. Birker, and N. Olhoff, *The CAOS system*, *Int. Ser. Numer. Math.* **110** (1993), 75–96.
- [167] P. A. Raviart and J. M. Thomas, *A mixed finite element method for second order elliptic problems*, *Lect. Notes Math.* **606** (1977), 292–315.
- [168] S. Reitzinger, *PEBBLES – user’s guide*, SFB ”Numerical and Symbolic Scientific Computing”, <http://www.sfb013.uni-linz.ac.at>.
- [169] _____, *Algebraic multigrid methods for large scale finite element equations*, Ph.D. thesis, Johannes Kepler University Linz, Austria, 2001.
- [170] K. Rektorys, *Survey of applicable mathematics*, Kluwer, Amsterdam, 1994.
- [171] A. Rietz and J. Petersson, *Simultaneous shape and thickness optimization*, *Struct. Multidisc. Optim.* **23** (2001), no. 1, 14–23.
- [172] W. Ritz, *Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik*, *J. Reine Angew. Math.* **135** (1909), 1–61.
- [173] G. I. N. Rozvany (ed.), *Shape and lay-out optimization of structural systems and optimality criteria methods*, CISM Lecture Notes no. 325, Springer, Wien, 1992.
- [174] G. I. N. Rozvany (ed.), *Optimization of large structural systems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.
- [175] _____, *Aims, scope, methods, history and unified terminology of computer-aided topology optimization in structural mechanics*, *Struct. Multidisc. Optim.* **21** (2001), 90–108.
- [176] W. Rudin, *Real and complex analysis*, McGraw Hill, New York, 1966.
- [177] _____, *Functional analysis*, McGraw Hill, New York, 1973.
- [178] _____, *Principles of mathematical analysis*, 3rd ed., International Series in Pure and Applied Mathematics, McGraw-Hill Book Company, 1976.

- [179] Y. Saad, *Iterative methods for sparse linear systems*, PWS, 1996, 2nd edition available at <http://www.cs.umn.edu/~saad/books.html>.
- [180] M. Save and W. Prager (eds.), *Structural optimization, vol. 1, optimality criteria*, Plenum Press, New York, 1985.
- [181] M. Save and W. Prager (eds.), *Structural optimization, vol. 2, mathematical optimization*, Plenum Press, New York, 1990.
- [182] O. Scherzer, *An iterative multi level algorithm for solving nonlinear ill-posed problems*, Numer. Math. **80** (1998), 579–600.
- [183] M. Schinnerl, U. Langer, and R. Lerch, *Multigrid simulation of electromagnetic actuators*, ZAMM, Z. Angew. Math. Mech. **81** (2001), no. 3, 729–730.
- [184] M. Schinnerl, J. Schöberl, M. Kaltenbacher, U. Langer, and R. Lerch, *Multigrid methods for the fast numerical simulation of coupled magnetomechanical systems*, ZAMM Z. Angew. Math. Mech. **80** (2000), no. 1, 117–120.
- [185] A. Schleupen, K. Maute, and E. Ramm, *Adaptive FE-procedures in shape optimization*, Struct. Multidisc. Optim. **19** (2000), 282–302.
- [186] J. Schöberl, *NETGEN - An advancing front 2D/3D-mesh generator based on abstract rules*, Comput. Vis. Sci. (1997), 41–52.
- [187] C. Schwab, *p- and hp- finite element methods: Theory and applications to solid and fluid mechanics*, Oxford University Press, 1998.
- [188] J. Sethian and A. Wiegmann, *Structural boundary design via level set and immersed interface methods*, J. Comp. Phys. **163** (2000), 489–528.
- [189] R. E. Showalter, *Hilbert space methods for partial differential equations*, Pitman, London, San Francisco, 1977.
- [190] O. Sigmund and J. Petersson, *Numerical instabilities in topology optimization: A survey on procedures dealing with checkerboards, mesh-dependencies and local minima*, Struct. Multidisc. Optim. **16** (1998), no. 1, 68–75.
- [191] C. A. C. Silva and M. L. Bittencourt, *An object-oriented structural optimization program*, Struct. Multidisc. Optim. **20** (2000), 154–166.
- [192] P. P. Silvester and R. L. Ferrari, *Finite elements for electrical engineers*, Cambridge University Press, 1983.
- [193] J. Simon, *Differentiation with respect to the domain in boundary value problems*, Numer. Funct. Anal. Optimization **2** (1980), 649–687.
- [194] S. L. Sobolev, *Applications of functional analysis in mathematical physics*, American Mathematical Society, Providence, 1963.
- [195] J. Sokolowski and A. Zochowski, *On the topological derivative in shape optimization*, SIAM J. Control Optimization **37** (1999), 1251–1272.

- [196] J. Sokolowski and J.-P. Zolesio, *Introduction to shape optimization*, Springer Series in Computational Mathematics, no. 16, Springer, Berlin, 1992.
- [197] L. Solymar, *Lectures on electromagnetic theory*, Oxford University Press, 1987.
- [198] W. Stadler, *Non-existence of solutions in optimal structural design*, *Optim. Control Appl. Methods* **7** (1986), 243–258.
- [199] C. W. Steele, *Numerical computation of electric and magnetic fields*, Van Nostrand Reynold Co., New York, 1987.
- [200] G. Strang and G. J. Fix, *An analysis of the finite element method*, Prentice-Hall, 1973.
- [201] J. A. Stratton, *Electromagnetic theory*, Mc Graw-Hill Book Co., Inc., New York, 1941.
- [202] K. Suzuki and N. Kikuchi, *A homogenization method for shape and topology optimization*, *Comput. Methods Appl. Mech. Eng.* **93** (1991), no. 3, 291–318.
- [203] K. Svanberg, *The method of moving asymptotes – a new method for structural optimization*, *Int. J. Numer. Methods Eng.* **24** (1987), 359–373.
- [204] ———, *A class of globally convergent optimization methods based on conservative convex separable approximations*, *SIAM J. Optim.* **12** (2002), no. 2, 555–573.
- [205] B. Szabó and I. Babuška, *Finite element analysis*, John Wiley & Sons, 1991.
- [206] N. Takahashi, *Optimization of die press model*, Proceedings of the TEAM Workshop in the Sixth Round (Okayama, Japan), March 1996.
- [207] P.-S. Tang and K.-H. Chang, *Integration of topology and shape optimization for design of structural components*, *Struct. Multidisc. Optim.* **22** (2001), 65–82.
- [208] The MathWorks, Inc., *Matlab user manual*, 1993.
- [209] H. Thomas, L. Zhou, and U. Schramm, *Issues of commercial optimization software development*, *Struct. Multidisc. Optim.* **23** (2002), no. 2, 97–110.
- [210] D. Tscherniak and O. Sigmund, *A web-based topology optimization program*, *Struct. Multidisc. Optim.* **22** (2001), no. 3, 179–187.
- [211] U. van Rienen, *Numerical methods in computational electrodynamics. linear systems in practical applications*, Lecture Notes in Computational Science and Engineering, vol. 12, Springer, Berlin, 2001.
- [212] K. Washizu, *Variational methods in elasticity and plasticity*, Pergamon, Oxford, Great Britain, 1983.
- [213] Y. M. Xie and G. P. Steven, *Evolutionary structural optimization*, Springer, Berlin, 1997.
- [214] J. Yoo and N. Kikuchi, *Topology optimization in magnetic fields using the homogenization design method*, *Int. J. Numer. Meth. Eng.* **48** (2000), 1463–1479.
- [215] K. Yuge, N. Iwai, and N. Kikuchi, *Optimization of 2-d structures subjected to nonlinear deformations using the homogenization method*, *Struct. Multidisc. Optim.* **17** (1999), no. 4, 286–298.

- [216] K. Yuge and N. Kikuchi, *Optimization of a frame structure subjected to a plastic deformation*, Struct. Multidisc. Optim. **10** (1995), no. 3–4, 197–208.
- [217] O. C. Zienkiewicz, *The finite element method in engineering science*, London etc.: McGraw-Hill XIV, 1971.
- [218] O. C. Zienkiewicz and R. L. Taylor, *The finite element method (parts 1–3)*, Butterworth Heinemann, Oxford, 2000.
- [219] M. Zlámal, *On the finite element method*, Numer. Math. **12** (1968), 394–409.
- [220] J.-P. Zolesio, *The material derivative (or speed) method for shape optimization*, Optimization of Distributed Parameter Structures, Part II (E. J. Haug and J. Cea, eds.), NATO Adv. Study Inst., 1981, pp. 1089–1151.
- [221] G. Zoutendijk, *Methods of feasible directions*, Elsevier, Amsterdam, 1960.
- [222] J. Zowe, M. Kočvara, and M. P. Bendsøe, *Free material optimization via mathematical programming*, Math. Program. **79** (1997), no. 1–3, Ser. B, 445–466.
- [223] A. K. Zvedin and V. A. Kotov, *Modern magneto-optics and magneto-optical materials*, Institute of Physics Publishing Bristol and Philadelphia, 1997.

Curriculum vitae

Personal data

Name: Ing. Dalibor Lukáš
Date and place of birth: June 10, 1976 in Ostrava, Czech Republic

Education

1990 - 1994: High Electrotechnical School (SPŠE) in Havířov, Czech Republic
1994 - 1999: studies in Informatics and Applied Mathematics at the VŠB–Technical University of Ostrava, Czech Republic
since November 1999: Ph.D. student in Informatics and Applied Mathematics at the VŠB–Technical University of Ostrava, Czech Republic

Career history

August 1996: programmer at the company Q-gir in Ostrava, Czech Republic
April–July 1999: programmer at the company Voest-Alpine Industrienlangebau GmbH in Leoben, Austria
April 2000–March 2001: research fellow at the Spezialforschungsbereich SFB013 “Numerical and Symbolic Scientific Computing”, Project F1309 “Hierarchical Optimal Design Methods”, University Linz, Austria
December 2002–May 2003: research assistant at the Department of Applied Mathematics, VŠB–Technical University of Ostrava, Czech Republic
since June 2003: research fellow at the SFB013 “Numerical and Symbolic Scientific Computing”, Project F1309 “Multilevel Solvers for Large Scale Discretized Optimization Problems”, University Linz, Austria

Publications

M.Sc. thesis [127], 1 article to appear in a refereed journal [120], 2 articles in refereed conference proceedings [123, 124], 4 refereed technical reports [121, 122, 125, 126], 5 contributed talks and 3 posters at international conferences, and 1 tutorial of a student conference project [119]

Research interests

Shape optimization, finite element method, mathematical modelling, scientific computing, magnetostatics

For more information see: <http://lukas.am.vsb.cz>