# Surrogate Optimization with <br> Algebraic Notes and Applications within the Electromagnetics Context 

vorgelegt von<br>M.Sc.<br>Mirsad Hadžiefendić

# an der Fakultät IV - Elektrotechnik und Informatik der Technischen Universität Berlin zur Erlangung des akademischen Grades 

Doktor der Ingenieurwissenschaften

- Dr.-Ing. -
genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Jürgen Bruns (TU Berlin, Fakultät IV)
Gutachter: Prof. Dr. Rolf Schuhmann (TU Berlin, Fakultät IV)
Gutachter: Prof. Dr. Fredi Tröltzsch (TU Berlin, Fakultät II)
Gutachter: Jun.-Prof. Dr. Ulrich Römer (TU Braunschweig)

Tag der wissenschaftlichen Aussprache: 14. Juli 2021

## Declaration of Authorship

I, Mirsad Hadžiefendić, M.Sc., declare that this thesis titled, "Surrogate Optimization with Algebraic Notes and Applications within the Electromagnetics Context" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:
Date:
"Ima jedna modra rijeka - valja nama preko rijeke."
Mehmedalija Mak Dizdar

# TECHNISCHE UNIVERSITÄT BERLIN 

Abstract<br>Fakultät Elektrotechnik und Informatik<br>Institut für Hochfrequenz- und Halbleiter-Systemtechnologien Fachgebiet Theoretische Elektrotechnik<br>Doktor der Ingenieurwissenschaften (Dr.-Ing.)<br>Surrogate Optimization with<br>Algebraic Notes and Applications within<br>the Electromagnetics Context<br>by Mirsad Hadžiefendić, M.Sc.

This thesis deals with surrogate optimization for applications within the electromagnetics context. Regarding the electromagnetics context, in particular, the magnetoquasistatic model of Maxwell's theory is discussed. Moreover, relevant points regarding the magnetoquasistatic model's numerical simulation and numerical optimization are examined.

The key notion surrogate optimization is thoroughly elaborated which is partitioned into three sub-notions: (1) surrogate modeling \& simulation, (2) surrogatebased optimizaton, and (3) surrogate-guided optimization. The various notions of surrogate optimization are tagged with algebraic notes in order to anticipate the toolset of the formal language of category theory. Moreover, the capability of the category theory toolset as an algebraic modeling framework for applications in surrogate optimization is investigated.

Finally, representatives of the class of inductive components are invoked and the surrogate optimization tools of the present work are applied to four high-fidelity optimization problems that are embedded within the setting of a two-dimensional linear boundary value problem and a three-dimensional linear boundary value problem, respectively. Concerning these optimization problems, some promising spots for a useful application of the category theory toolset are illuminated.

From the bird's-eye view, this thesis achieves some progress in the scientific thicket of full automation of the virtual prototyping of power electronic systems.

From the frog's-eye view, i.e., at a more technical level, some of the present work's achievements deal with hybrid model management strategies of surrogateguided optimization methods, the repercussions of the choice of a sampling plan on these methods, and formalization issues regarding surrogate optimization with multiple low-fidelity models.

# TECHNISCHE UNIVERSITÄT BERLIN 

# Kurzzusammenfassung 

Fakultät Elektrotechnik und Informatik Institut für Hochfrequenz- und Halbleiter-Systemtechnologien Fachgebiet Theoretische Elektrotechnik<br>Doktor der Ingenieurwissenschaften (Dr.-Ing.)

Surrogate Optimization with
Algebraic Notes and Applications within
the Electromagnetics Context
von Mirsad Hadžiefendić, M.Sc.

Diese Arbeit beschäftigt sich mit der Optimierung mit Ersatzmodellen für Anwendungen innerhalb der elektromagnetischen Feldtheorie. Im Kontext der elektromagnetischen Feldtheorie geht diese Arbeit insbesondere auf das magnetoquasistatische Modell der Maxwell'schen Gleichungen ein. Zudem werden relevante Punkte hinsichtlich der numerischen Simulation und Optimierung des magnetoquasistatischen Modells besprochen.

Es wird der Schlüsselbegriff „Surrogate optimization" (in Dt.: Optimierung mit Ersatzmodellen) ausführlich ausgearbeitet, der im Rahmen dieser Arbeit in drei Unterbegriffe aufgeteilt wird: (1) „Surrogate modeling \& optimization" (in Dt.: Modellierung und Simulation mit Ersatzmodellen), (2) „Surrogate-based optimization" (in Dt.: auf Ersatzmodellen basierende Optimierung) und (3) „Surrogate-guided optimization" (in Dt.: durch Ersatzmodelle geführte Optimierung). Man beachte, dass die verschiedenen Begriffe bzgl. der Optimierung mit Ersatzmodellen mit algebraischen Anmerkungen versehen werden, um den mathematischen Werkzeugkasten der formalen Sprache der Kategorientheorie vorwegzunehmen. Es werden im Speziellen die Möglichkeiten untersucht, den kategorientheoretischen Werkzeugkasten als algebraische Modellierungsumgebung für Anwendungen im Rahmen der Optimierung mit Ersatzmodellen zu verwenden.

Abschließend werden Vertreter aus der Klasse der induktiven Komponenten präsentiert und die im Rahmen dieser Arbeit vorgestellten Werkzeuge bzgl. der Optimierung mit Ersatzmodellen werden auf vier hochgenaue Optimierungsprobleme angewendet. Diese Optimierungsprobleme sind in das Umfeld von zweidimensionalen und dreidimensionalen linearen Randwertproblemen eingebettet. Hinsichtlich dieser Optimierungsprobleme werden einige vielversprechende Stellen für eine nützliche Anwendung des kategorientheoretischen Werkzeugkastens beleuchtet.

Aus der Vogelperspektive betrachtend erreicht diese Dissertation einen gewissen Fortschritt im wissenschaftlichen Dickicht der vollständigen Automatisierung des virtuellen Prototypings von leistungselektronischen Systemen.

Aus der Froschperspektive betrachtend, d. h., auf einer eher technischen Ebene, beschäftigen sich einige der Errungenschaften dieser Arbeit mit hybriden Modell-management-Strategien von durch Ersatzmodelle geführte Optimierungsmethoden, mit den Auswirkungen der Wahl des Stichprobenplans auf diese Methoden und mit Formalisierungsproblemen bzgl. der Optimierung mit Ersatzmodellen im Falle von mehreren ungenauen Modellen.

## Acknowledgements

I sincerely thank Prof. Dr. Rolf Schuhmann for his mentorship, advice, and openmindedness. I thank also all my colleagues at the group Theoretische Elektrotechnik for a pleasant and inspirational working environment. A special thanks has to be given to Marcus Christian Lehmann, Albert Piwonski, and Rodrigo Silva Rezende with whom I have shared many pursuits of knowledge.

I thank Prof. Dr. Fredi Tröltzsch and Jun.-Prof. Dr. Ulrich Römer for their time and energy to provide the second review and third review regarding the present work, respectively. And I would like to thank Prof. Dr. Jürgen Bruns for taking over the chairmanship of the doctoral committee.

Finally, I thank all my friends and my whole family for their support. My greatest and warmest thanks are due to my parents, Mensur and Abasa, and my brothers, Admir and Emir.

## Contents

Declaration of Authorship ..... iii
Abstract ..... vii
Kurzzusammenfassung ..... ix
Acknowledgements ..... xi
1 Introduction ..... 1
1.1 A bigger picture: The ideal long-term goal ..... 1
1.2 Glimpse at the details: Surrogate optimization ..... 3
1.3 Setting a horizon: The research scope \& goals ..... 6
2 Magnetoquasistatic Maxwell's theory - Modeling, simulation, and opti- mization ..... 11
2.1 Magnetoquasistatic Model of Maxwell's theory ..... 12
2.1.1 The fundamental problem statement of electromagnetism ..... 12
2.1.2 The system of Maxwell's equations ..... 12
2.1.3 The magnetoquasistatic subsystem \& the magnetostatic sub- system ..... 16
2.2 Numerical simulation of the magnetoquasistatic model ..... 18
2.2.1 The weak formulation ..... 18
2.2.2 Numerical approximation ..... 21
2.2.3 Parametric mathematical model ..... 22
2.3 Numerical optimization with the magnetoquasistatic model ..... 23
2.3.1 Optimization with a partial differential equation ..... 24
2.3.2 Nonlinear optimization problem ..... 26
2.3.3 Optimization algorithms ..... 28
2.4 In closing ..... 39
3 Surrogate optimization ..... 41
3.1 Surrogate modeling \& simulation ..... 42
3.1.1 An abstract setting ..... 42
3.1.2 Deterministic and probabilistic data-fit low-fidelity models ..... 59
3.1.3 Simplified-physics low-fidelity models ..... 75
3.2 Surrogate-based optimization ..... 81
3.2.1 Optimization with test functions by data-fit low-fidelity models ..... 81
3.2.2 Optimization with test functions by emulated simplified-physics low-fidelity models ..... 96
3.3 Surrogate-guided optimization ..... 100
3.3.1 Sequential kriging optimization ..... 102
3.3.2 Optimization within the space-mapping paradigm ..... 104
3.3.3 Co-kriging optimization ..... 117
3.4 In closing ..... 126
4 An algebraic modeling framework using the category theoretical language for applications in surrogate optimization ..... 129
4.1 Recapitulating and enlarging the contextual landscape ..... 130
4.1.1 Recapitulating \& the context of full automation of SGO ..... 130
4.1.2 Relevant related work ..... 132
4.2 Category theory toolset ..... 133
4.2.1 Fostering some intuition ..... 133
4.2.2 Applying some rigor ..... 134
4.2.3 Computational facets ..... 140
4.3 Using the CT toolset for SGO methods ..... 142
4.3.1 Specifying a general optimization problem ..... 143
4.3.2 Specifying surrogate-guided optimization methods ..... 145
4.4 Use cases of the CT toolset within the electromagnetics context ..... 153
4.4.1 Use case \#1: Simplified-physics low-fidelity models ..... 153
4.4.2 Use case \#2: Coordinate transformations ..... 157
4.5 Future use cases for the CT toolset ..... 161
4.6 In closing ..... 163
5 Surrogate optimization with the magnetoquasistatic model ..... 165
5.1 Solenoid with a core ..... 166
5.1.1 Preliminary consideration ..... 166
5.1.2 Optimization problem I ..... 172
5.1.3 Optimization problem II ..... 178
5.2 Common-Mode Choke ..... 182
5.2.1 Preliminary consideration ..... 182
5.2.2 Optimization problem I ..... 183
5.2.3 Optimization problem II ..... 186
5.3 In closing ..... 190
6 Conclusion and outlook ..... 191
6.1 Conclusion ..... 191
6.2 Outlook ..... 193
A Multivariate polynomials (§3.1.2) ..... 195
A. 1 Reparametrization using mean-centered arguments ..... 195
A. 2 Bernstein polynomials ..... 196
A. 3 Chebyshev polynomials ..... 199
B Solenoid with a core (§ 5.1) ..... 203
B. 1 An electrical network viewpoint ..... 203
B. 2 A visualization of evaluated data-fit low-fidelity models regarding (5.12), (5.14c), and (5.15) ..... 206
Bibliography ..... 211

## List of Figures

1.1 An inductive component in various representations ..... 2
1.2 A schematic depiction of a low-fidelity model. ..... 5
1.3 Illustrating two generic devices under test. ..... 7
1.4 A schematic orientation aid for the present work ..... 10
2.1 Illustrations of a magneotquasistatic subsystem's domain. ..... 17
2.2 Representations of the six test functions in Table 2.1. ..... 35
2.3 Representations of the six test functions in Table 2.1 (highlighting the neighborhood of the global minimum). ..... 36
2.4 Depicting $\operatorname{grad}(f)\left(x_{1}, x_{2}\right)$ with $\left(x_{1}, x_{2}\right) \in U$ as a projection on the con- tour representation of the six test functions in Table 2.1. ..... 37
2.5 Depicting $s_{i}\left(f,\left(x_{1}, x_{2}\right)\right)$ from (2.49) with $i \in\{1,2\}$ and $\left(x_{1}, x_{2}\right) \in U$ w.r.t. the test functions in Table 2.1. ..... 38
3.1 Representations of different sampling plans (part one). ..... 49
3.2 Representations of different sampling plans (part two). ..... 50
3.3 The Audze-Eglais LHC sampling plan in Figure 3.2 adapted for the contour representation in Figure 2.2b. ..... 51
3.4 Sampling plan $X_{\mathbf{s}}$ (condition number $\kappa\left(B^{T} B\right)$ ) ..... 66
3.5 Basis under consideration for the space $P_{\leq 2}^{1}$. (i) Monomial basis, (ii) Bern- stein basis, (iii) Chebyshev basis. ..... 67
3.6 The six radial basis functions in Table 3.2. ..... 69
3.7 A schematic depiction of a user-prescribed hierarchy of generic bound- ary value problems. ..... 76
3.8 A schematic depiction of a user-prescribed hierarchy of magnetoqua- sistatic and magnetostatic problems. ..... 77
3.9 Using the Sobol quasi-random sequence sampling plan in Figure 3.2, a 2-variate monomial polynomial model $p(x) \in \mathrm{P}_{\leq 2}^{2}$ via regression of the test functions in Figure 2.2. ..... 83
3.10 Using the Sobol quasi-random sequence sampling plan in Figure 3.2, a radial basis function $\phi=r \mapsto \phi(r)$ with thin plate spline assignment via interpolation of the test functions in Figure 2.2. ..... 84
3.11 Using the Sobol quasi-random sequence sampling plan in Figure 3.2, a kriging low-fidelity model via interpolation of the test functions in Figure 2.2. ..... 85
3.12 The value of $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}$ and $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{NR}}$ evaluated at the number of folds $k$ and the number of testing points $m_{g}$ w.r.t. the test functions in Figure 2.2 with normalized values. ..... 90
3.13 The value of $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}$ evaluated at the number of folds $k$ and the number of testing points $m_{g}$ w.r.t. the test functions in Figure 2.2 with normalized values. ..... 91
3.14 The value of $\bar{r}_{\hat{y} \tilde{y}, \mathrm{cv}}^{2}$ evaluated at the number of folds $k$ and the number of testing points $m_{g}$ w.r.t. the test functions in Figure 2.2. ..... 92
3.15 The value of $S_{\tilde{\tilde{y}}, i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the test functions in Figure 2.2 and the number of training points $m_{t}$ ..... 93
3.16 Emulated simplified-physics low-fidelity models for the modified Branin test function in Figure 2.2 via the assignment rule in (3.105). ..... 98
3.17 Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ as a projection on the contour representa- tion of the emulated simplified-physics low-fidelity models for the modified Branin test function in Figure 3.16. ..... 99
4.1 A schematic depiction of a three-dimensional helical coil of 5 turns in xyz-coordinates and in uvw-coordinates. ..... 158
5.1 A representation of a solenoid with core within FEMM4.2. ..... 167
5.2 Contour representation of $\tilde{\hat{j}}_{\bar{P}_{L}, \omega_{i}}$ and $\tilde{\hat{Q}}_{L, \omega_{i}}$ in (5.31) for various fre- quencies. ..... 180
5.3 A representation of a common-mode choke within FEMM4.2. ..... 182
5.4 A prototypical version of a simplistic EMC filter. ..... 184
5.5 A representation of two common-mode choke within FEMM4.2. ..... 187
5.6 Contour representation of $\tilde{\hat{Q}}_{L, \omega_{0}}$ in (5.47). ..... 189
A. 1 Instances of a Chebyshev grid as a sampling plan $X_{\mathbf{s}} \subset\left(X^{2}\right)^{m_{1} \times m_{2}}$. ..... 200
B. 1 Circuit diagram representation of the three fundamental passive elec- trical components. ..... 203
B. 2 Two representatives from the equivalence class of circuit diagrams for real inductive components. ..... 204
B. 3 The magnitude and the phase of the impedances associated with the representatives in Figure B.2. ..... 205
B. 4 By using the Sobol quasi-random sequence sampling plan with $m:=21$ and the data-fit low-fidelity models in $\S 3.2$.1; representations of $\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$and $\tilde{V}_{u t}\left(x_{1}, x_{2}\right)$ where $\omega_{0}:=2 \pi 100 \mathrm{kHz}$.207
B. 5 By using the Sobol quasi-random sequence sampling plan with $m:=21$ and the data-fit low-fidelity models in §3.2.1; representations of $\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega_{0}}\left(x_{1}, x_{2}\right)$and $\tilde{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)$ where $\omega_{0}:=2 \pi 100 \mathrm{kHz} . . . . . . . . . . . . . . . . . .208$B. 6 Depicting $\operatorname{grad}(\tilde{K})\left(x_{1}, x_{2}\right)$ as a projection on the contour representa-tion of the data-fit low-fidelity models for the functions in Figure B.4.209
B. 7 Depicting $\operatorname{grad}(\tilde{K})\left(x_{1}, x_{2}\right)$ as a projection on the contour representa-tion of the data-fit low-fidelity models for the functions in Figure B.5.210

## List of Tables

2.1 Test functions of form $f=\left(x_{1}, x_{2}\right) \mapsto f\left(x_{1}, x_{2}\right): U \rightarrow \mathbb{R}$. ..... 29
2.2 The normalized global first-order sensitivity measure $S_{i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the Figure 2.2b. ..... 31
2.3 The normalized global first-order sensitivity measure $S_{i}^{N}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the Figure 2.3b. ..... 31
2.4 The normalized global first-order sensitivity measure $S_{i}^{\mathrm{N}}$ evaluated at $f_{\mathrm{R}_{N_{\xi}}}$ w.r.t. the domain $[-2.0,2.0]^{N_{\tilde{\zeta}}}$ with $N_{\xi} \in\{2,3,4,5,6,7\}$ ..... 32
2.5 Check exemplarily the Ackley function's global optimum. ..... 33
3.1 Given some pairs $(k, d) \in \mathbb{N}^{2}$, the dimension of $P_{\leq k}^{d}$ and $P_{k}^{d}$. ..... 61
3.2 Given a generic radial basis function $\phi=r \mapsto \phi(r)$ with function sig- nature $\mathbb{R}^{+} \rightarrow \mathbb{R}$, six different definitions for the assignment $\phi(r)$. ..... 69
3.3 The normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. a 2 -variate monomial poly- nomial. ..... 86
3.4 The normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. a radial basis function with thin plate spline assignment. ..... 86
3.5 The normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. a kriging low-fidelity model. ..... 87
3.6 The normalized global first-order sensitivity measure $S_{\tilde{y}, i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the 2 -variate monomial polynomial in Figure 3.9b. ..... 87
3.7 The normalized global first-order sensitivity measure $S_{\hat{y}, i}^{N}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the radial basis function with thin plate spline as- signment in Figure 3.10b. ..... 87
3.8 The normalized global first-order sensitivity measure $S_{\tilde{\tilde{y}}, i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the kriging low-fidelity model in Figure 3.11b. ..... 87
3.9 The normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. a kriging low-fidelity model of a generalized version of the Rosenbrock test function in Table 2.1 (without normalized values). ..... 89
3.10 The normalized global first-order sensitivity measure $S_{i}^{N}$ evaluated at $f_{\mathbb{R}_{N_{\xi}}}$ from (2.52) w.r.t. the domain $[-2.0,2.0]^{N_{\tilde{\xi}}}$ with $N_{\xi} \in\{2,3,4,5,6,7\}$. ..... 94
3.11 The low-fidelity models' normalized global first-order sensitivity mea- sures (LFSM) error in (3.37) evaluated at $f_{\mathrm{R}_{N_{\bar{\xi}}}}$ from (2.52) w.r.t. the domain $[-2.0,2.0]^{N_{\tilde{\xi}}}$ with $N_{\tilde{\xi}} \in\{2,3,4,5,6,7\}$ ..... 94
3.12 Surrogate-based optimization w.r.t. the modified Branin function us- ing data-fit low-fidelity models. ..... 96
3.13 The choice of the 4 -tuple of parameters $(\alpha, \beta, \gamma, \delta)$ in Figure 3.16. ..... 96
3.14 The normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. emulated simplified-physics low-fidelity models in Figure 3.16a. ..... 97
3.15 The normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. emulated simplified-physics low-fidelity models in Figure 3.16b. ..... 97
5.1 The rough heuristic estimate of the scale of the winding losses in (5.13) for various operating frequencies $f_{0}$. ..... 171
5.2 (Ia) $S_{i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at (a) $f \equiv \tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}$ and (b) $f \equiv \tilde{V}_{\mathrm{ut}}$ w.r.t. the Figure B.6; (Ib) Given $m_{j}:=50, \operatorname{LFSM}$ error $\mathfrak{e}_{m_{j}}\left(S_{\tilde{\tilde{y}}, i}^{\mathrm{N}}\right)$ w.r.t. (Ia); (IIa) $S_{i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at (a) $f \equiv \tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega_{0}}$ and (b) $f \equiv \tilde{\hat{Q}}_{L, \omega_{0}}$ w.r.t. the Figure B.7; (IIb) Given $m_{j}:=50$, LFSM error $\mathfrak{e}_{m_{j}}\left(S_{\tilde{\tilde{y}}, i}^{\mathrm{N}}\right)$ w.r.t. (IIa). ..... 174
5.3 The mean SSPCC $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k:=5}$ within the $k$-fold cross validation method w.r.t. the simplified-physics low-fidelity model regarding the entities in (5.17) ..... 176
5.4 The $\log$ data in an abridged version w.r.t. the setting of the $\log$ data in (5.19) for different operating frequencies. ..... 179
5.5 The $\log$ data in an abridged version w.r.t. the setting of the $\log$ data in (5.19) for the operating frequency $1 \times 10^{8} \mathrm{~Hz}$ and different initial points. ..... 179
A. 1 The condition number w.r.t. a sampling plan from Figure 3.4 without and with reparametrization in (A.2). ..... 195
A. 2 The trace of $\frac{1}{m}\left(B^{\mathrm{T}} B\right)$ and the trace of $\frac{1}{m}\left(B_{\zeta}^{\mathrm{T}} B_{\zeta}\right)$. ..... 197
A. 3 The condition number $\kappa\left(B^{\mathrm{T}} B+\lambda I\right)$ and $\kappa\left(B_{\zeta}^{\mathrm{T}} B_{\zeta}+\lambda I\right)$. ..... 198

## List of Abbreviations

| 1D | One-dimensional |
| :---: | :---: |
| 2D | Two-dimensional |
| 3D | Three-dimensional |
| ADE | Adaptive Differential Evolution |
| AMMO | Approximation and Model Management Optimization |
| BVP | Boundary Value Problem |
| CCC | Cartesian Closed Category |
| CM | Common-Mode |
| CMC | Common-Mode Choke |
| COBYLA | Constrained Optimization by Linear Approximation |
| CT | Category Theoretical/Category Theory |
| DIRECT | Dividing Rectangles |
| DM | Differential-Mode |
| DoFF | Degree of Forgetfulness |
| EMC | Electromagnetic Compatibility |
| FE | Finite Element |
| GA | Genetic Algorithm |
| GPU | Graphics Processing Unit |
| KKT | Karush-Kuhn-Tucker |
| L-BFGS | Limited-memory Broyden-Fletcher-Goldfarb-Shanno |
| LBVP | Linear Boundary Value Problem |
| LFSM | Low-Fidelity Models' Normalized Global First-Order Sensitivity Measures |
| LHC | Latin Hypercube |
| LU | Lower-Upper |
| MEA | Modified Evolutionary Algorithm |
| MLE | Maximum Likelihood Estimate |
| MM | Manifold Mapping |
| MMA | Method of Moving Asymptotes |
| MS | Moore-Skelboe |
| NEGE | Normalized Empirical Generalization Error |
| NLBVP | Non-Linear Boundary Value Problem |
| NMS | Nelder-Mead Simplex |
| NREGE | Normalized Root Empirical Generalization Error |
| PDE | Partial Differential Equation |
| PL | Programming Language |
| PS | Particle Swarm |
| RPM | Response and Parameter Mapping |
| S- | Scattering |
| SGO | Surrogate-guided Optimization |
| SM | Space Mapping |
| SQP | Sequential Quadratic Programming |
| SSPCC | Squared Sample Pearson Correlation Coefficient |
| SVD | Singular Value Decomposition |

TPS RBF Thin Plate Spline Radial Basis Function<br>TRASM Trust Region Aggressive Space Mapping<br>UMP Universal Mapping Property<br>WPBVP Well-Posed Boundary Value Problem

# Physical \& Mathematical Constants 

```
speed of light in vacuum
c
vacuum magnetic permeability }\mp@subsup{\mu}{0}{}=4\pi\times1\mp@subsup{0}{}{-7}\mp@subsup{\textrm{Hm}}{}{-1
vacuum electric permittivity
\epsilon}\mp@subsup{\epsilon}{0}{=1/( ( 
pi
    \pi=3.1415926535897...
```


## List of Symbols (Selection)

| $\mathbf{J}$ | electric current flux density |
| :--- | :--- |
| $\mu$ | magnetic permeability |
| $\sigma$ | electric conductivity |
| $\omega$ | angular frequency |
| $\mathbf{A}$ | magnetic vector potential |
| $\phi$ | electric scalar potential |
| $I_{0}$ | fixed current intensity |
| $\bar{P}_{\mathrm{L}}$ | time-averaged ohmic loss |
| $\bar{W}_{\mathrm{m}}$ | time-averaged magnetic energy |
| $R$ | resistance |
| $L$ | inductance |
| $f_{0}$ | operating frequency |
| $\omega_{0}$ | operating angular frequency |
| $V_{\mathrm{ut}}$ | volume under test |
| $\mathbb{R}$ | set of real numbers |
| $\mathbf{x}$ | space variable |
| $\Omega$ | space region |
| $\partial \Omega$ | boundary of $\Omega$ |
| $f: A \rightarrow B$ | domain $A$ and codomain $B$ of function $f$ with the signature $A \rightarrow B$ |
| $f=x \mapsto f(x)$ | function $f$ with the assignment rule $x \mapsto f(x)$ |
| div, grad, curl | differential operators: divergence, gradient, and curl |
| $\forall$ | universal quantifier |
| $\exists$ | existential quantifier |
| $V$ | Hilbert space |
| $\\|\cdot\\|_{V}$ | appropriate norm on $V$ |
| $Q$ | quantity of interest |
| $h$ | mesh size parameter |
| $\mathcal{T}_{h}$ | simplicial triangulation |
| $\mathbf{u}_{h}$ | discrete solution |
| $N_{\tilde{F}}$ | number of parameters |
| $\xi$ | parameter point |
| $\mathbf{f}$ | parametric solution function |
| $\hat{Q}_{\tilde{F}}$ | reduced parametric quantity of interest |
| $\hat{j}$ | reduced objective function |
| $\circ$ | composition operator |
| $s_{i}$ | local first-order sensitivity measure w.r.t. component $i$ |
| $S_{i}^{N}$ | normalized global first-order sensitivity measure w.r.t. component $i$ |
| $d$ | dimensionality |
| $\mathfrak{e}_{\mathrm{H}}($ K $)$ | high-fidelity function approximation error |
| $\mathbf{s}$ | sample |
| $X_{\mathrm{s}}$ | sampling plan |
| $m$ | sample size |


| $\mathfrak{e}_{\mathrm{H}, \mathrm{s}}\left(\hat{Q}_{\xi}\right)$ | empirical surrogate modeling error |
| :---: | :---: |
| $\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\underline{\xi}}\right)$ | normalized empirical generalization error |
| $r_{\hat{y} \tilde{y}}^{2}$ | squared sample Pearson correlation coefficient |
| $\mathrm{P}_{\leq k}^{d}$ | space of $d$-variate polynomials of total degree at most $k$ |
| $\mathcal{N}_{m}\left(y \mid \underline{y}_{\underline{\prime}}, \Sigma\right)$ | probability density of an m-dimensional Gaussian distribution at $y$ |
| $\Psi$ | correlation matrix |
| C | covariance matrix |
| $L_{l n}$ | ln-likelihood function |
| $\Delta_{A x=b}$ | threshold for a termination criterion of an iterative solver |
| P | domain-oriented correction map |
| $\Delta^{(k+1)}$ | $k+1$-th iteration trust-region radius |
| $x^{*}$ | optimal solution of the high-fidelity optimization problem |
| X | object |
| $g$ | morphism |
| $\mathcal{A}$ | category |
| F | functor |
| $\alpha$ | natural transformation |
| $S_{i j}$ | ( $i, j$ )-th scattering parameter |
| $N$ | number of turns of a winding |
| $m_{\text {SGO,sm }}$ | number of high-fidelity function evaluations (space-mapping) |
| $m_{\text {SGO,ck }}$ | number of high-fidelity function evaluations (co-kriging) |
| $m_{w}$ | number of operating frequencies |

Dedicated to my parents, Mensur \& Abasa

## Chapter 1

## Introduction

In this chapter, we encounter the background, the scope, and the research goals of the present dissertation:

- First, I discuss the bigger picture, more precisely, the ideal long-term goal which originates in the engineering domain of power electronics and inspires the starting point and the direction of the research project.
- Second, I sketch the general path to which the dissertation contributes; that is, the development and application of surrogate modeling, simulation, and optimization methods in, primarily, the electromagnetic field theory's realms of magnetostatics and magnetoquasistatics.
- Third, I conclude the chapter by providing the path-dependent research goals that guide the remainder of the work. ${ }^{1}$


### 1.1 A bigger picture: The ideal long-term goal

Implicitly or explicitly, every research project has an ideal long-term goal which helps to establish the investigation's concrete context and the actual research goals. The thesis' ideal long-term goal is the full automation of the virtual prototyping of power electronic systems. Let me elucidate briefly this goal.

The domain of power electronics is concerned with the control and the conversion of electrical energy by means of fast switching semiconductor components (see, e.g., [152], [89], [60]); two representatives of the broad class of power electronic systems are three-phase rectifiers and electromagnetic compatibility (EMC) filters (see Fig. 1.1). ${ }^{2}$ Given this domain, my notion of "full automation" is that an ideal software system processes a user's input specifications and it outputs an appropriate power electronic system - without any additional user's intervention. Finally, I understand the term "virtual prototyping" as a proxy for "mathematical modeling, numerical simulation \& optimization".

Note that full automation is still far away; but there has already been prolific research regarding virtual prototyping of power electronic systems (see, e.g., [222], [220], [41]). The corresponding real-world engineering optimization problems consist of different levels of design complexity: from the materials' design over the

[^0]

Figure 1.1: An inductive component in various representations: (a) in a circuit diagram (figure from [154, p. 7]), (b) in a realworld EMC filter (source: Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration IZM), (c) in the 3D simulation tool CST Studio Suite ${ }^{\circledR 3}$, and (d) in the 2D simulation tool FEMM4.2 (see [149]).
components' design to the systems' design. Mind that these problems involve intricate interactions between various physical domains such as electromagnetics, fluid dynamics or structural mechanics; additionally, they involve several conflicting objectives such as performance, cost or efficiency. Hence, formalizing properly and solving efficiently these problems are challenging tasks.

To date, the reported optimization procedures utilize predominantly concepts from the area of multidisciplinary design optimization (see, e.g., [2], [147]) and from the area of multiobjective optimization (see, e.g., [150], [146]):

Multiobjective optimization. If multiple objective functions are taken into account, then multiobjective optimization - or vector optimization - expresses an optimal design by the notion of Pareto optimality, i.e., an optimal design is Pareto-optimal if an improvement concerning one objective leads inevitably to a degradation concerning another objective. Common multiobjective optimization techniques include a transformation of the multiple objectives into a single objective, for instance, by the weighted sum method: First, the objectives are multiplied by weights (non-negative numbers that add up to one); next, the weighted objectives are summed up. An immediate complication is the need for selecting a specific combination of weights which is reflecting a specific preference of objectives.

[^1]Multidisciplinary design optimization. Optimal input variables attained by an optimization using a single physical discipline rarely equal the optimal input variables attained by an optimization using multiple physical disciplines - especially, if there are interdependencies between the different disciplines. For the preceding observation, multidisciplinary design optimization offers a framework to keep track of the input variables and all the involved output variables. However, one important issue is how to establish compatibility regarding the variables; another important issue is how to choose an adequate architecture, i.e., how to coordinate the analysis of the multiple interdependent physical disciplines. These issues influence the selection of a solution method and the reasoning about the optimal design.

Regarding virtual prototyping of power electronic systems, a noteworthiness of the deployed procedures is that they utilize different computational and noncomputational models of variable degrees of fidelity, e.g., finite element simulations, closed-form expressions, physical experiments, etc. The areas of surrogate optimization (see, e.g., [70]), and multifidelity optimization (see, e.g., [166]), respectively, are dedicated to exploit such different models for optimization purposes.

Mind that, to my best knowledge, concepts from surrogate optimization have not yet been exhaustively discussed in the context of virtual prototyping of power electronic systems. However, as I have elaborated above, the complexity of realworld power electronic systems' design is tremendous. Therefore, the focus of the present work's applications is on particular optimization problems concerning inductive components (see Fig. 1.1). Inductive components represent significant devices under test since they contribute heavily to the losses of a power electronic system, and they demand a lot of space within a power electronic system (cf. [154, p. 2]).

All in all, the surrogate optimization of inductive components constitutes the starting point and the direction of the research project.

### 1.2 Glimpse at the details: Surrogate optimization

Surrogate optimization for engineering design problems is a vast research area that spans several decades of intensive investigations (see, e.g., surveys in [125], [127], and [126]). At a conceptional level, the notion surrogate optimization encompasses three sub-notions:

- surrogate modeling \& simulation,
- surrogate-based optimization, and
- surrogate-guided optimization. ${ }^{4}$

Surrogate modeling \& simulation. The basic assumption is that the evaluation of a given function - aka high-fidelity function or model - is too expensive; hence, there is a need to approximate this function in a meaningful manner by another function - aka low-fidelity function or model - whose evaluation costs are, by design, much lower than those of the high-fidelity model. An example of a high-fidelity

[^2]model is the joule loss functional computed by a high-order finite element simulation (see, e.g., [227]). An example of a low-fidelity model is a fit to data collected by sampling the high-fidelity model.

Some immediate issues are concerned with error bounds or error estimates in order to assess the quality of the low-fidelity model.

Concerning this sub-notion, the surrogate model is identical to the low-fidelity model; and surrogate simulation means evaluating the low-fidelity model.

Surrogate-based optimization. Assuming that the surrogate model is sufficiently accurate, the basic idea of surrogate-based optimization is to replace the optimization problem regarding the high-fidelity function by an optimization problem regarding the low-fidelity function - without any additional interaction with the highfidelity function.

Next, the optimal solution corresponding to the low-fidelity optimization problem is computed, for instance, by deterministic algorithms such as the sequential quadratic programming (SQP) (see, e.g., [158, ch. 18]) or stochastic algorithms such as the genetic algorithm (GA) (see, e.g., [49, p. 39-43]). Finally, the computed optimal solution is checked within the high-fidelity optimization problem (cf. [31, p. 2]).

An issue concerning this sub-notion is connected to the assessment of the computed optimal solution - since the optimal solution of the high-fidelity optimization problem is unknown apriori.

Thus, the low-fidelity optimization problem's optimal solution is either accepted as a proxy - to some extent - of a high-fidelity optimization problem's optimal solution; or it is utilized as a starting point within the high-fidelity optimization problem.

Surrogate-guided optimization. Compared to the previous optimization approach, the key difference is that there is an interaction between the high-fidelity optimization problem and the low-fidelity optimization problem. During the search for the optimal solution of the high-fidelity optimization problem, the role of the low-fidelity function is to speedup the search; whereas the role of the high-fidelity function is to ensure convergence of the search.

A common issue is concerned with the general theoretic characterization of optimal solutions by the first-order necessary conditions - i.e., the Karush-Kuhn-Tucker (KKT) conditions (see, e.g., [210, p. 17f]).

Concerning this sub-notion, the surrogate model is not necessarily identical to the low-fidelity model (cf. [194, p. 28]) since it depends on the type of interaction - or model management strategy (cf. [166, p. 554f]) - between the high-fidelity model and the low-fidelity model.

In the remaining text, I use the terms surrogate-guided optimization and multifidelity optimization (recall § 1.1) interchangeably.

Motivated by the field of application in the present work (recall § 1.1), the semantics of the models is mainly determined by the electromagnetic field theory's realms of magnetostatics and magnetoquasistatics (see, e.g., [139] or [103]).

The preceding elaborations already hint at the pivotal role played by the lowfidelity model in the area of surrogate optimization. In Fig. 1.2, there is a schematic depiction of a low-fidelity model depending on the available information about the high-fidelity model. Considering the high-fidelity model as a black-box, a graybox or a white-box model influences one classification of low-fidelity models into


Figure 1.2: A schematic depiction of a low-fidelity model (encoded by $\tilde{K}$ ) depending on the available information about the high-fidelity model (encoded by K). The vertical arrows merely emphasize a connection between a high-fidelity model and a low-fidelity model; the horizontal arrows indicate input and output entities. The boxes associated with the low-fidelity model $\tilde{K}$ solely indicate schematically different potential representations of $\tilde{\mathrm{K}}$. The high-fidelity model K is considered as (1) a black-box model, (2) a gray-box model or (3) a white-box model. The vertical black line separating (1) and (2) from (3) indicates that, in the present work, the focus is on (1) and (2).
(1) data-fit, (2) simplified-physics, and (3) projection-based models (cf. [166, p. 556]). ${ }^{5}$ Another possible way to classify low-fidelity models (indicated by the vertical black line in Fig. 1.2) is to ask whether the models are intrusive or non-intrusive - where I understand "intrusive" as a need to modify the numerical software underlying the high-fidelity model (cf. [77, p. 3f]).

In my investigation, I reduce the area of focus on low-fidelity models of data-fit type - for instance, kriging models (see, e.g., [137]) - and of simplified-physics type for instance, coarse-grid discretization models (see, e.g., [125, p. 159]). In order to provide a complete picture and to comprehend the reduced focus, I address briefly low-fidelity models of projection-based type.

Brief digression: projection-based low-fidelity models. The upcoming exposition is very condensed. Thus, for a more elaborate exposition, I refer to [189] and [20].

The basic mechanism behind this type of low-fidelity models is: The high-fidelity model is given as a system of equations in a high-dimensional space and a corresponding low-dimensional subspace is constructed such that some desired characteristics of the system are preserved. The low-fidelity model constitutes the projection of the high-fidelity model onto the low-dimensional subspace.

In the context of electrical engineering, this type of low-fidelity models is intensively discussed for circuit simulations and electromagnetic field simulations. Regarding applications in circuit simulations, see [21] for a collection of detailed investigations. Regarding applications in electromagnetic field simulations, there are various investigations depending on the meaning of the parameter under consideration (encoded by $x$ in Fig. 1.2). Common meanings of the parameter are: frequency

[^3](see, e.g., [223], [118]), material (see, e.g., [115], [44]), and geometry (see, e.g., [40], [30]).

Undoubtedly, the projection-based type of low-fidelity models is a very important type because it does not depend on domain-specific knowledge of an expert. On the one hand, this independence is valuable for the automated construction of a lowfidelity model; especially, if theoretically sound error bounds and error estimators are available. On the other hand, it is questionable why the domain specific knowledge of an expert - for example, in the form of a large number of different models should not be exploited.

With regard to the complexity of a real-world engineering design problem (recall $\S 1.1$ ), there are, inevitably, a lot of open challenges concerning the theory and the implementation of low-fidelity models of projection-based type. However, it is arguably reasonable to state that a harmoniously balanced interaction between all three types of low-fidelity models has the potential to be a fruitful approach in the long run - as recent promising results (see, e.g., [165, p. A3163]) indicate.

Finally, I have sketched the general path to which the present dissertation contributes. Next, I identify critical points on this general path and specify the research scope and goals.

### 1.3 Setting a horizon: The research scope \& goals

In order to have a chance to reconcile the ideal-long term goal (see § 1.1) and surrogate optimization (see § 1.2), there are at least two critical points that one encounters:
(1) In real-world design optimization problems, various high-fidelity models and low-fidelity models from various sources are used non-formally - that lack rigorously proven error bounds and error estimators (see, e.g., [179] or [32]); despite the non-formal usage, these models and their relationships have proven to be useful in practice.
(2) In general, the task of comparing optimization algorithms is non-trivial (see, e.g., [224]). With regard to surrogate optimization, there is a variety of methods discussed in the literature but the task of choosing an appropriate method for a given problem is non-trivial. An obstacle is to find a proper way to classify the numerous methods. Especially, there is a lack of well-defined benchmarks that could enable a standardized benchmark-focused comparison.

For illustration purpose of the two critical points, I exhibit briefly two examples:
(E1) An example concerning (1) is connected to the computation of the ohmic loss (as the quantity of interest) of a three-dimensional helical coil of $N$ turns (as the device under test, see (i) in Fig. 1.3). ${ }^{6}$ If we replace this helical coil by a collection of $N$ toroids (see (ii) in Fig. 1.3), then the ohmic loss computation associated with the coil represents the high-fidelity model, and the ohmic loss computation associated with the toroids represents the low-fidelity model. The comparison of the two models is usually based on the comparison of the respective computed ohmic loss encoded as a non-negative real number.

[^4]Mind that, for instance from a topological viewpoint (see, e.g., [92]) or a boundary value problem viewpoint (see, e.g., [174]), the two devices under test are not necessarily the same in general. However, they are commonly assumed as approximately the same regarding the ohmic loss computation - i.e., the sameness of the high-fidelity model and the low-fidelity model implies sameness of the respective devices under test.


Figure 1.3: Two generic devices under test: (i) a three-dimensional helical coil of 5 turns, and (ii) a collection of 5 toroids. The devices are created within CST Studio Suite ${ }^{\circledR}$.
(E2) An example concerning (2) is to choose a surrogate-guided optimization method from the class of methods following the space mapping paradigm (see, e.g., [125, p. 50]) for the optimization of a three-dimensional helical coil of $N$ turns as a device under test.
Note that, in the space mapping paradigm, the low-fidelity model and the surrogate model are not identical (cf. [194, p. 28]). Various approaches have been proposed to construct the surrogate model (see, e.g., [49, ch. 3]). An attempt to classify some methods within this class is by assessing the quality of the lowfidelity models and the surrogate models with regard to convergence properties of the corresponding algorithms (see, e.g., [120], [121]).

I argue that the two critical points (1) and (2) are natural bounds to the full recognition of surrogate optimization methods by practitioners in the industrial sector. Additionally, the two points bound naturally the research scope in the present work at a problem- or application-oriented level and at a theory-oriented level.

Bound at a problem- or application-oriented level. Since there is no realistic possibility to test all conceivable classes of use cases by all surrogate optimization methods, there is a need to restrict the investigation to a subclass of use cases and a subclass of methods. Therefore, the use cases are restricted to applications associated with inductive components; and the methods are restricted to those methods that are using simplified-physics and data-fit as low-fidelity models and that are using the space mapping paradigm (see, e.g., [125, p. 50]) and co-kriging approach (see, e.g., [70, p. 167]) as model management strategies. According to the terminology in [166, p. 555], the space mapping paradigm is a subtype of the model management strategy adaptation, and the co-kriging approach is a subtype of the model management strategy fusion.

At the transition between the two levels, there are inevitable software issues regarding, for instance, finite element (FE) simulation tools or programming languages (PLs). Commercial FE software (e.g., in CST Studio Suite ${ }^{\circledR}$ ), open-source FE
software (e.g., FEMM4.2), and in-house programs for the algorithms (written, e.g., in MATLAB ${ }^{\circledR 7}$ and Julia ${ }^{8}$ ) are all employed in the present work.

Bound at a theory-oriented level. If one applies the high-fidelity model and the low-fidelity model of the example (E1) in context of the example (E2), then one can observe that the current formal languages in surrogate-guided optimization (see, e.g., [127], [166]) enable only insufficiently to encode the semantics (or interpretation) that one model is derived from the other. Considering point (1), such an encoding is beneficial in order to preserve and organize formally the practical prior knowledge about the models and their relationships - which is also beneficial as a stage of model preparation in context of point (2).

Commonly, questions concerning semantics (and syntax) are rather investigated by tools from logical analysis than by tools from numerical analysis. Mostly, in numerical analysis, questions regarding logical sound footing for a reliable reasoning about numerical models are associated with the notions of validation and verification (see, e.g., [159]).

However, a promising mediator between these apparently different tool sets is the formal language of category theory which is a holistic-structural approach to mathematics (see, e.g., [11], [177], [180]). Its usefulness in physics (see, e.g., [73], [46]) and in computer science (see, e.g., [167], [16]) has already been recognized. Moreover, its usefulness is gradually getting recognition in electrical engineering and computational electromagnetics (see, e.g., [13], [133]). Thus, the category theoretical language opens up a new opportunity to complement the primarily numerical analytic perspective in the context of surrogate optimization.

To draw the research scope completely, it is also necessary to mention directions that are closely related to the present work but which will not be pursued.

Disclaimer: What is not considered in the dissertation. I provide a list of three trends in the context of surrogate optimization (see $\S 1.2$ ). Note that the list is certainly not exhaustive, though:

1. In real-world applications, there are many sources of uncertainties such as manufacturing imperfections that result in, for instance, uncertain material, shape or excitation information of a problem under consideration. Hence, the first trend is to investigate mathematical methods of uncertainty quantification (see, e.g., [201], [178], [50], [30], [122]).
2. The need for finding quickly an optimal solution associated with a high-fidelity model is a reason for using surrogate optimization. However, an accelerated search is also conceivable if the overall computational costs of a highfidelity model are reduced by utilizing concepts from parallel computing (see, e.g., [190], [212]). Thus, a second trend is to explore the applicability of parallel computing (see a survey, e.g., in [87]).
3. In surrogate optimization, as mentioned before, selecting a proper method for a given problem is a non-trivial task since the selection depends heavily on

[^5]the given problem. Therefore, there is a lack of generally valid guiding principles for the selection process. However, considering machine learning techniques (see, e.g., [192]), a third trend is concerned with the automation of the selection process (see, e.g., [185]).

After providing the research scope, I can state the superordinate research goals:

- investigate the applicability of a surrogate optimization's subsegment to applications associated with inductive components;
- investigate the benefits and drawbacks of the category theoretical language as an algebraic modeling toolbox in the context of surrogate optimization.

In order to assess the achievements of the present work, it aids to consider the superordinate research goals from a methodological point of view (cf. [199, p. 3] ${ }^{9}$ ): The first goal is largely concerned with utilizing long-researched techniques for new applications; whereas the second goal is largely concerned with introducing a new area of knowledge to a long-researched area of knowledge.

Finally, I present the outline of the work:

- In chapter 2, I discuss particularly the magnetoquasistatic model of Maxwell's theory. Moreover, some relevant aspects regarding the numerical simulation of the magnetoquasistatic model are presented. Finally, a few key points concerning the numerical optimization with the magnetoquasistatic model are illuminated. Mind that, in the exposition, I take also a few small detours in order to show by familiar examples some facets of the formal language of category theory in advance. At the end, I address a zoo of optimization algorithms regarding nonlinear optimization problems. Furthermore, six test functions are introduced and a gradient-based interpretation of sensitivity measures are deployed that are primarily applied to models such as, e.g., data-fit low-fidelity models, that permit the determination of derivative information by forward mode automatic differentiation.
- In chapter 3, I elaborate thoroughly on the key notion surrogate optimization and on the proposed partitioning of this notion in $\S 1.2$ into the three sub-notions: (1) surrogate modeling \& simulation, (2) surrogate-based optimization, and (3) surrogate-guided optimization. The various notions of surrogate optimization are tagged with algebraic notes in order to anticipate the toolset of the category theoretical language. Concerning the sub-notion (1), the notion of a high-fidelity model, a low-fidelity model, and a surrogate model are pinned down, for instance. Concerning the sub-notion (2), a numerical scaffolding of a benchmark-focused classification of test functions is carved out, for example. Concerning the sub-notion (3), given an optimization procedure within the space-mapping paradigm and a co-kriging low-fidelity model, we encounter, e.g., the elucidation of potential hybrid model management strategies.

[^6]- In chapter 4, a formalization-oriented viewpoint is deepened by introducing the category theory toolset. I focus solely on core tools and attempt to balance intuition and rigor regarding this toolset. Moreover, the toolset is used to specify a general optimization problem and to specify surrogate-guided optimization methods where the focus rests on optimization procedures within the space-mapping paradigm. In addition, we face also other use cases for the toolset related to high- and low-fidelity models associated with examples of applications in electrical engineering.
- In chapter 5 , we look at a solenoid with a core and a common-mode choke as representatives of the class of inductive components where I elaborate on four optimization problems within the setting of a two-dimensional linear boundary value problem and a three-dimensional linear boundary value problem, respectively. Supposing the context of an electrical engineering design workflow, a strategy of using the tools from chapter 3 in practical applications is presented and some relevant spots are carved out where the tools from chapter 4 can have a favorable influence, too.
- In chapter 6, I distill a conclusion from the presented research and present an outlook.

In order to furnish one with some kind of visual orientation aid for maneuvering within the present work, the following figure depicts schematically four generic levels (the level of programs, the level of algorithms, the level of (generalized) functions, and the level of applications) to which the essence of the respective discussion in chapters $2,3,4$ and 5 can be roughly assigned to. In addition, some abovementioned terms are associated with these four levels as well.


Figure 1.4: A schematic orientation aid for the present work (inspired by [225, p.3]). The index PL refers to "programming language". The assignment of ch. 6 is omitted. The dotted lines merely indicate a connection between the large ellipses. Each large ellipse represents one of the generic levels: programs, algorithms, (generalized) functions, and applications. The respective small ellipse within a large ellipse symbolizes a sub-area of interest. A colored asterisk within a small ellipse encodes a use case that is stated as colored text.

## Chapter 2

## Magnetoquasistatic Maxwell's theory - Modeling, simulation, and optimization

In the present work, the physical framework is primarily restricted to macroscopic scale electromagnetic phenomena described by Maxwell's theory - in which the magnetic energy and the power loss (weighted with a time of oscillation) are much bigger than the electric energy such that physical effects concerning electromagnetic wave propagation can be disregarded. The majority of the thesis' central applications under investigation is embedded in this particular physical framework. Therefore, I choose to expand on this particular physical framework in the subsequent sections and to leave the common details of the general physical framework to the standard literature (see, e.g., [139], [103]).

If the operating frequency is greater than zero, then, as customary, the mathematical representation of the physical framework is given by the magnetoquasistatic model of Maxwell's theory; otherwise the mathematical representation is given by the magnetostatic model of Maxwell's theory.

Respecting the standard approach in electrical engineering, let us discuss the corresponding mathematical models in the language of vector analysis. Hence, in order to express Maxwell's theory (see, e.g., in [139], [103]), one has to assume a familiarity with notions such as vector fields, scalar fields, the differential operators div, grad, curl, etc.

We are not concerned with a thorough numerical analysis of the models - since we abstract over most of their inner workings in the remaining chapters. However, the modern treatment regarding the numerical simulation and optimization of these models makes it necessary to involve some basic concepts from the languages of functional analysis (see, e.g., [226], [140], [218], [9], [153]) and differential geometry (see, e.g., [93], [42], [83], [66], [134]) that provide methodological and terminological guidance. Thus, one has to suppose a working knowledge of elementary definitions and results concerning notions such as Hilbert spaces, bounded linear operators, manifolds and similar; but the explanations do not follow strictly the socalled "definition-theorem-proof model of mathematics" (cf. [204, p. 3]).

Furthermore, these two languages assist in tracing some intuitions concerning the structural perspective that is emphasized by the language of category theory that I employ in ch. 4.

### 2.1 Magnetoquasistatic Model of Maxwell's theory

The section provides a detailed description of the physical realm of the present work: Starting from a brief exposition of the fundamental problem statement of electromagnetism, we discuss the statement's mathematical representation by the system of Maxwell's equations. From the general system, we derive the magnetoquasistatic subsystem and the magnetostatic subsystem; and for the former subsystem, we arrive at a strong formulation that serves as an orientation point for the discussion about the numerical simulation in the subsequent section.

### 2.1.1 The fundamental problem statement of electromagnetism

In the present work, we focus exclusively on Maxwell's theory of electromagnetism and corresponding mathematical models. For more details on the distinction between theory, model, and formulation, I refer to, e.g., [132, p. 5-9].

In [205, p. 273], the author states the fundamental problem of electromagnetism:

- Given a space region and a time interval,
- Given the nature of the materials that fill the region,
- Given the boundary conditions,
- Given the initial values of the configuration variables,
- Given the space and time distribution of charges and currents,
- Find the configuration of the field at every point and at every later instant. ${ }^{1}$


### 2.1.2 The system of Maxwell's equations

Associated with the problem statement in the previous section is the system of Maxwell's equations that represents its mathematical model. Let us formalize the model by the language of vector analysis.

The system of Maxwell's equations contains the following field functions: the electric field intensity $\mathbf{E}$, the electric field flux density $\mathbf{D}$, the magnetic field flux density $\mathbf{B}$, the magnetic field intensity $\mathbf{H}$, the electric charge density $\rho$, and the electric current flux density $\mathbf{J}$.

All field functions are defined as functions of space and time. It is assumed that a three-dimensional Euclidean space as a model for space and a one-dimensional Euclidean space as a model for time are given (see, e.g., [33, p. 109]). Additionally, it is assumed that there are no mechanically moving parts involved. The space variable $\mathbf{x}$ is a member of the space region $\Omega \subset \mathbb{R}^{3}$ and the time variable $t$ is a member of the time interval $I_{T}:=[0, T] \subset \mathbb{R}$.

Moreover, the field functions are categorized into two types: vector fields and scalar fields, i.e., given an instant in time, vector fields map a point in space to a vector and scalar fields map a point in space to a scalar. Thus, the field functions E, $\mathbf{D}, \mathbf{B}, \mathbf{H}$, and $\mathbf{J}$ are vector fields with the function signature $\Omega \times I_{T} \rightarrow \mathbb{R}^{3}$; the field

[^7]function $\rho$ is a scalar field with the function signature $\Omega \times I_{T} \rightarrow \mathbb{R}$. ${ }^{2}$ Regarding the notation, however, it should be noted that the symbols for the field functions can also mean the evaluated field functions - for instance, $\mathbf{E} \equiv \mathbf{E}(\mathbf{x}, t), \mathbf{D} \equiv \mathbf{D}(\mathbf{x}, t)$ etc.

Notice that, e.g., the field functions E and $\mathbf{B}$ constitute configuration variables (see § 2.1.1). The electric current density can be decomposed in a conduction part $\mathbf{J}_{\text {cond }}$ due to an electrically conductive medium, and a source part $\mathbf{J}_{\text {src }}$ that is imposed externally; hence, $\mathbf{J}:=\mathbf{J}_{\text {cond }}+\mathbf{J}_{\text {src }}$.

Customarily, the system of Maxwell's equations is displayed in the integral version or in the differential version. Let $A \subset \Omega$ be an oriented surface with boundary $\partial A$, and let $V \subset \Omega$ denote a volume with boundary $\partial V$. Mind that the symbol $\partial$ behaves polymorphically, i.e., it is utilized to declare a boundary operator and a partial time derivative operator $\partial_{t}$. The integral version reads as

$$
\begin{align*}
& \text { (i) } \forall A \cdot \int_{\partial A} \mathbf{H} \cdot \mathrm{~d} \mathbf{s}=\int_{A} \mathbf{J} \cdot \mathrm{~d} \mathbf{A}+\mathrm{d}_{t} \int_{A} \mathbf{D} \cdot \mathrm{~d} \mathbf{A}, \\
& \text { (ii) } \forall A \cdot \int_{\partial A} \mathbf{E} \cdot \mathrm{~d} \mathbf{s}=-\mathrm{d}_{t} \int_{A} \mathbf{B} \cdot \mathrm{~d} \mathbf{A},  \tag{2.1}\\
& \text { (iii) } \forall V \cdot \int_{\partial V} \mathbf{D} \cdot \mathrm{~d} \mathbf{A}=\int_{V} \rho \mathrm{~d} V \\
& \text { (iv) } \forall V \cdot \int_{\partial V} \mathbf{B} \cdot \mathrm{~d} \mathbf{A}=0 .
\end{align*}
$$

The system of Maxwell's equations is completed by the three constitutive equations that relate the corresponding field functions and express their interaction with matter. Assuming time-invariant, linear, homogeneous, and isotropic material, the equations are given by

$$
\begin{align*}
& \text { (i) } \mathbf{D} \stackrel{\text { mat }}{=} \epsilon \mathbf{E}, \\
& \text { (ii) } \mathbf{B} \stackrel{\text { mat }}{=} \mu \mathbf{H},  \tag{2.2}\\
& \text { (iii) } \mathbf{J}_{\text {cond }} \stackrel{\text { mat }}{=} \sigma \mathbf{E},
\end{align*}
$$

where the notation $\stackrel{\text { mat }}{=}$ follows the style of [205, p. 33]. The electric permittivity $\epsilon$, the magnetic permeability $\mu$, and the electric conductivity $\sigma$ are considered as functions of space. The absolute electric permittivity $\epsilon_{0}$ is incorporated in $\epsilon$ and the absolute magnetic permeability $\mu_{0}$ is incorporated in $\mu$. Notice that it depends on the context whether $\epsilon \equiv \epsilon(\mathbf{x}), \mu \equiv \mu(\mathbf{x})$, and $\sigma \equiv \sigma(\mathbf{x})$. In the case of non-linear and inhomogeneous magnetic material, it is customary to introduce the magnetization $\mathbf{M}$ as an additional field function (see, e.g., [178, p. 3]). In the presence of permanent magnets, it is customary to introduce an additional magnetic field strength $\mathbf{H}_{\mathrm{pm}}$ (see, e.g., [30, p. 10]). However, driven by the domain of applications in the present work (recall $\S 1.1$ ), we are mainly concerned with constitutive equations given by (2.2).

[^8]If we prescribe a unit normal vector to $\partial \Omega$, then one can extract the tangential and normal components of the field functions in (2.1) at the boundary of the space region. If subregions $\Omega_{1}$ and $\Omega_{2}$ of the space region $\Omega$ exhibit different material properties, then additional conditions have to be taken into account at the material interfaces. For more details on the handling of all these conditions - especially by trace operators in a functional analytic setting -, I refer to [218], [32], and [179].

Providing initial values of the corresponding field functions, and the space and time information of the sources, all requirements according to the problem statement in $\S$ 2.1.1 are fulfilled. Applying the theorem of Stokes and the theorem of Gauss on (2.1), we derive the system of Maxwell's equation in the differential version

$$
\begin{cases}\text { (i) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} . \operatorname{curl} \mathbf{H}=\mathbf{J}+\partial_{t} \mathbf{D}, \\ \text { (ii) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} \cdot \operatorname{curl} \mathbf{E}=-\partial_{t} \mathbf{B},  \tag{2.3}\\ \text { (iii) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} \cdot \operatorname{div} \mathbf{D}=\rho, \\ \text { (iv) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} \cdot \operatorname{div} \mathbf{B}=0,\end{cases}
$$

where suitable boundary conditions and reasonable properties regarding the space region $\Omega$ are assumed - which will be discussed later.

From (2.3), we can recover the continuity equation that encodes local charge conservation

$$
\begin{equation*}
\forall(\mathbf{x}, t) \in \Omega \times I_{T} . \operatorname{div} \mathbf{J}=-\partial_{t} \rho . \tag{2.4}
\end{equation*}
$$

Commonly, potential field functions such as the electric scalar potential $\phi$ and the magnetic vector potential A are also employed in the context of Maxwell's equations. Starting from (2.3), these two potential field functions are introduced, e.g., in the representation
(i) $\mathbf{E}=:-\boldsymbol{\operatorname { g r a d }} \phi-\partial_{t} \mathbf{A}$,
(ii) $\mathbf{B}=: \operatorname{curl} \mathbf{A}$.

The potential field functions are particularly relevant for the formulation of magnetostatics problem and magnetoquasistatics problem where, for instance, the socalled A- $\phi$ formulation plays an important role in the numerical approximation of these problems (see, e.g., [179, ch. 6]).

Before we move on to the approximation of the system of Maxwell's equations in magnetostatics and in magnetoquasistatics, let us seize the opportunity for a detour to discuss shortly an important structural property of the differential operators regarding the physical space. By discussing this property, I want to carve out some intuitions concerning the category theory's structural approach in ch. 4.

Detour 1: a structural perspective on a structural property. Recall that the three differential operators grad, curl, and div exhibit an important structural property for contractible domains: Firstly, given a field function expressed via grad, applying curl to this function results in the zero vector field; secondly, given a field function expressed via curl, applying div to this function results in the zero scalar field. Simplisticly, the structural property is expressed as curl grad $\equiv \mathbf{0}$ and div curl $\equiv 0$ which is encoded in the Poincaré lemma (cf. [32, p. 298]).

If we emphasize the function signature of the three differential operators, i.e., grad has the type "scalar field $\rightarrow$ vector field", curl has the type "vector field $\rightarrow$ vector field", and div has the type "vector field $\rightarrow$ scalar field", then one can systematize the previous structural property as

[^9]The inverted commas indicate that the systematization of the structural property cannot be formalized properly in the standard language of vector analysis (see, e.g., [83, p. 31]). For this purpose, there is a need to state the types and maps more precisely. The language of differential geometry and the language of functional analysis are capable to encode properly the structural property which, in these languages, is the algebraic expression called exact sequence (cf. [32, p. 132f]). Let us expose briefly this expression in these two languages. Leaving the majority of details to the numerous textbooks that have been mentioned at the chapter's beginning, we focus only on the bare minimum of technicalities since the exposition's purpose is to abstract the structural essence of the common algebraic expression.

Within the manifold-based differential geometric approach, the full system of Maxwell's equations is formulated based on the machinery of differential forms and exterior calculus. Using this approach, the vector field functions in (2.1) are called "vector proxies" (cf. [33, p. 132]). For instance, the electric field strength is merely a representative of an observable entity, more precisely, the assignment of a voltage (i.e., the electromotive force) to an oriented line. Hence, the map

$$
e=l \mapsto \int_{l} \mathbf{E} \cdot \mathrm{~d} \mathbf{s}
$$

is called a differential form of degree 1 - abbreviated as 1 -form. Assuming a smooth manifold $M$, we denote the space of 1 -forms as $\Lambda^{1}(M)$, thus, $e$ is an element of $\Lambda^{1}(M)$. If we associate other field functions with other geometric objects such as points, surfaces, and volumes, then one can designate the corresponding spaces: the space of 0forms as $\Lambda^{0}(M)$, the space of 2 -forms as $\Lambda^{2}(M)$, and the space of 3 -forms as $\Lambda^{3}(M)$, respectively. Additionally, one can instantiate a notion of a differential operator via the exterior derivative d which maps a differential form of degree $k$ to a differential form of degree $k+1$ such that one can formalize the abovementioned systematization of the structural property in vector analysis as the algebraic expression

$$
\begin{equation*}
0 \rightarrow \Lambda^{0}(M) \xrightarrow{d_{1}} \Lambda^{1}(M) \xrightarrow{d_{2}} \Lambda^{2}(M) \xrightarrow{d_{3}} \Lambda^{3}(M) \rightarrow 0 \tag{2.6}
\end{equation*}
$$

Observe that the structural property itself is encoded in a defining property of the exterior derivative: $\forall a \in \Lambda^{k}(M) .\left(\mathrm{d}_{k+1} \circ \mathrm{~d}_{k}\right)(a) \equiv 0$; or concisely: $\mathrm{d}_{k+1} \circ \mathrm{~d}_{k} \equiv 0$.

Picking the functional analytic approach, technically, we are deploying the machinery of Sobolev spaces and weak differential operators such that, given a regular bounded, contractible domain $D$ of the Euclidean space, we have to choose appropriate Sobolev spaces for the field functions in (2.3) and in (2.5), i.e.: $L_{\text {grad }}^{2}(D)$, $\mathbb{L}_{\text {curl }}^{2}(D), \mathbb{L}_{\text {div }}^{2}(D)$, and $L^{2}(D)$ where the notational convention by [32, p. 128] is employed in which $L^{2}(D)$ denotes the space of square-integrable functions over $D$ and $\mathbb{L}^{2}(D)$ denotes the space of square-integrable vector fields over $D$ (see, e.g., [32, p. 69]). Moreover, we have to set the domains and codomains of the weak differential operators such that, by construction, we arrive at the algebraic expression which is conceptually similar to (2.6):

$$
\begin{equation*}
0 \rightarrow L_{\text {grad }}^{2}(D) \xrightarrow{\text { grad }} \mathbb{L}_{\text {curl }}^{2}(D) \xrightarrow{\text { curl }} \mathbb{L}_{\text {div }}^{2}(D) \xrightarrow{\text { div }} L^{2}(D) \rightarrow 0 . \tag{2.7}
\end{equation*}
$$

One significance of diagrams such as (2.6) and (2.7) is that they provide a guidance for the construction of a discrete representation of the full system of Maxwell's equations in the sense that, in numerical approximations, such type of diagrams should be preserved (cf. [33, p. 145]) in order to mimic the continuous properties
at the discrete level. Therefore, such diagrams enable some kind of consistency check (cf. [111]). Following the spirit of the diagrammatic notation in (2.6) and in (2.7), one can systematize the full system of Maxwell's equations in the so-called Maxwell's house (cf. [32, p. 134]). See, in addition, the so-called Tonti's classification diagrams of electromagnetism, in short: Tonti's diagrams (cf. [205, p. 307-323]).

Regarding an in-depth elaboration on the relationship between (2.6) and (2.7), I refer to the discussion in [6] about the de Rham complex and Hilbert complexes, and I refer to the discussion in [131] about the Sobolev space setting.

From a purely structural perspective, though, the essence of the algebraic expression in (2.6) and in (2.7) is that there are, in general, four spaces $U, V, W, X$ and three maps $f_{1}, f_{2}, f_{3}$ such that the algebraic expression reads as

$$
\begin{equation*}
0 \rightarrow U \xrightarrow{f_{1}} V \xrightarrow{f_{2}} W \xrightarrow{f_{3}} X \rightarrow 0 . \tag{2.8}
\end{equation*}
$$

At an intuitive level, one can regard the expression in (2.8) as the syntax, whereas one can consider the expression in (2.6) as one possible semantics and the expression in (2.7) as another possible semantics. More interestingly, one can identify another semantics if, in (2.6) and in (2.7), one regards the spaces as vector spaces and the maps as linear maps. This interplay of syntax and semantics shows us a flavor of a structural perspective that foreshadows the category theoretical language which we encounter in ch. 4.

### 2.1.3 The magnetoquasistatic subsystem \& the magnetostatic subsystem

Due to the applications addressed in the present work, we are chiefly interested in subsystems of Maxwell's equations in (2.1) and in (2.3), respectively, where wave propagating effects are neglected, and, therefore, the term $\partial_{t} \mathbf{D}$ is neglected. Additionally, the electric charge density $\rho$ is assumed to be the zero scalar field function. These restrictions lead to the magnetoquasistatic subsytem of Maxwell's equation which, in the literature (see, e.g., [179, p. 7]), is also called eddy current approximation or magnetoquasistatic approximation of the Maxwell's equation. Furthermore, if one neglects all time-dependencies, then one arrives at the magnetostatic subsystem of Maxwell's equation.

The two subsystems represent approximations, hence, there is a need for justification. Let us assume that the magnetic energy and power loss (weighted with a time of oscillation) are much bigger than the electric energy. There are additional quantifiable tools (cf. [178, p. 6]) to check our assumption: (1) Given an operating angular frequency $\omega$ in a domain, the product $\omega \epsilon$ has to be much smaller compared to $\sigma$; and (2) the diameter of a bounded domain has to be much smaller than the corresponding minimal wavelength within the bounded domain. For more details on the mathematical justification, see, e.g., [179, ch. 2] or [198].

Reducing the system in (2.3) according to the corresponding restrictions, the magnetoquasistatic subsystem of Maxwell's equations reads as

$$
\begin{cases}\text { (i) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} . \operatorname{curl} \mathbf{H}=\sigma \mathbf{E}+\mathbf{J}_{\text {src }}, \\ \text { (ii) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} . \operatorname{curl} \mathbf{E}=-\partial_{t}(\mu \mathbf{H}),  \tag{2.9}\\ \text { (iii) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} . \operatorname{div}(\epsilon \mathbf{E})=0, \\ \text { (iv) } & \forall(\mathbf{x}, t) \in \Omega \times I_{T} . \operatorname{div}(\mu \mathbf{H})=0 .\end{cases}
$$

If we focus on the time-harmonic case where the field functions exhibit a sinusoidal time-dependency, one can formulate the magnetoquasistatic subsystem of

Maxwell's equations in the frequency domain as


Remark 2.1.1. In the frequency domain, the notation of field functions indicates complexvalued field functions. As mentioned before, the symbols for the complex-valued field functions can also mean the evaluated complex-valued field functions-for instance, $\boldsymbol{E} \equiv \boldsymbol{E}(\boldsymbol{x}, j \omega)$. Hence, the notation $\overline{\boldsymbol{E}}$ indicates the componentwise conjugation of $\boldsymbol{E}$ and $\boldsymbol{E}(\boldsymbol{x}, j \omega)$, respectively.

Remark 2.1.2. In Figure 2.1, I illustrate schematically two common representatives of a magnetoquasistatic subystem's domain in application.

Remark 2.1.3. Moving from (2.9) to (2.10) means that we have moved from an initialboundary value problem (IBVP) to a boundary value problem (BVP). In the time-harmonic case, the equation (iv) in (2.9) can been dropped since it can be recovered from the equation (ii) in (2.10). Moreover, the equation (iii) in (2.10) holds for all non-conducting subregions ( $\Omega_{n c}$ ), whereas the equations (i) and (ii) refer to the whole physical (or computational) domain under consideration ( $\Omega$ ). However, mind that the electric conductivity $\sigma$ is supposed to be greater than zero in a conducting subregion $\left(\Omega_{c}\right)$, and to be equal to zero in a non-conducting subregion $\left(\Omega_{n c}\right)$. Finally, one has to assume that div $J_{s r c}=0$ in $\Omega_{n c}$ that follows immediately from the continuity equation in (2.4).


Figure 2.1: A schematic illustration of two common representatives of a magneotquasistatic subsystem's domain in application.

If we apply the frequency-domain representation of the potential field functions from (2.5) to the subsystem in (2.10), one can state this subsystem in the A- $\phi$ formulation:

$$
\left\{\begin{array}{l}
\text { (i) } \quad \operatorname{curl}\left(\mu^{-1} \operatorname{curl} \mathbf{A}\right)=-\sigma \operatorname{grad} \phi-j \omega \sigma \mathbf{A}+\mathbf{J}_{\mathrm{src}} \text { in } \Omega,  \tag{2.11}\\
\text { (ii) } \\
\operatorname{div}(-\epsilon \operatorname{grad} \phi-j \omega \epsilon \mathbf{A})=0 \text { in } \Omega_{\mathrm{nc}} .
\end{array}\right.
$$

By setting $\omega \equiv 0$ in (2.11), one can immediately derive the magnetostatic subsystem of Maxwell's equations in the $\mathbf{A}-\phi$ formulation:

$$
\begin{cases}\text { (i) } & \operatorname{curl}\left(\mu^{-1} \operatorname{curl} \mathbf{A}\right)=-\sigma \operatorname{grad} \phi+\mathbf{J}_{\mathrm{src}} \text { in } \Omega,  \tag{2.12}\\ \text { (ii) } & \operatorname{div}(-\epsilon \operatorname{grad} \phi)=0 \text { in } \Omega_{\mathrm{nc}} .\end{cases}
$$

A necessary remark is concerned with the gradient field $\operatorname{grad} \phi$ in (2.12) and in (2.11). Some authors (see, e.g., [178, p. 9]) neglect this term, other authors (see, e.g., [30, p. 11]) introduce the field function $\mathbf{J}_{\text {src }}$ via the term $-\sigma \operatorname{grad} \phi$. In either cases, the investigation focuses only on finding the vector potential A. For a treatment of the term grad $\phi$ in a more general setting, see, e.g., the discussion in [97].

In order to obtain uniqueness of the vector potential $\mathbf{A}$, it is necessary to introduce appropriate gauge conditions and boundary conditions. A common gauge condition is the Coulomb gauge

$$
\begin{equation*}
\operatorname{div} \mathbf{A}=0 \text { in } \Omega . \tag{2.13}
\end{equation*}
$$

Let $\mathbf{n}$ denote the exterior unit normal at the computational domain's boundary $\partial \Omega$, then common boundary conditions are Dirichlet boundary conditions

$$
\begin{equation*}
\mathbf{A} \cdot \mathbf{n}=0 \text { on } \Gamma_{\mathrm{D}} \subset \partial \Omega, \tag{2.14}
\end{equation*}
$$

and Neumann boundary conditions

$$
\begin{equation*}
\mu^{-1}(\operatorname{curl} \mathbf{A}) \times \mathbf{n}=0 \text { on } \Gamma_{N} \subset \partial \Omega, \tag{2.15}
\end{equation*}
$$

where it is assumed that $\Gamma_{D} \cup \Gamma_{\mathrm{N}} \equiv \partial \Omega$. If we assume a simply-connected computational domain $\Omega$, then, based on (i) in (2.11) - and, analogously, based on (i) in (2.12) -, one can wrap-up the previous pieces of information in a so-called strong formulation

$$
\begin{cases}\operatorname{curl}\left(\mu^{-1} \operatorname{curl} \mathbf{A}\right)+\sigma \operatorname{grad} \phi+j \omega \sigma \mathbf{A}=\mathbf{J}_{\operatorname{src}} & \text { in } \Omega,  \tag{2.1.}\\ \operatorname{div} \mathbf{A}=0 & \text { in } \Omega, \\ \mathbf{A} \cdot \mathbf{n}=0 & \text { on } \Gamma_{\mathrm{D}}, \\ \mu^{-1}(\operatorname{curl} \mathbf{A}) \times \mathbf{n}=0 & \text { on } \Gamma_{\mathrm{N}} .\end{cases}
$$

It is customary that due to, for instance, rotational or translational symmetry (see, e.g., (ii) in Fig. 1.3), a two-dimensional setting is applied. Hence, the formulation in (2.16) has to be adapted accordingly. For further discussion on this adaptation, I refer to [178, p. 17f], [30, p. 11f], or, more generally, [83], [32].

### 2.2 Numerical simulation of the magnetoquasistatic model

As mentioned at the very beginning of the chapter, instead of a thorough numerical analysis of the model, the primary concern is rather to utilize the methodological and terminological guidance by basic concepts from functional analysis in the discussion of the numerical simulation of the model. Hence, let us discuss abstractly the weak formulation and its numerical approximation; and let us conclude the section by an exposition of a parametric mathematical model.

### 2.2.1 The weak formulation

A strong formulation such as in (2.16) is the orientation point of the numerical simulation of the magnetoquasistatic model. However, due to continuity issues, the solvability of the given problem in (2.16) is not generally guaranteed. Starting from the strong formulation, a weak formulation has to be derived by formally multiplying (2.16) with a test function $\mathbf{v}$ as a member of a Hilbert space $V(\Omega)$ and integrating over the space region $\Omega$. Under certain conditions, a solution to the weak formulation is a solution to the strong formulation.

The weak formulation is a means to show that the problem in (2.16) is well-posed in the sense of Hadamard, i.e., it exists a solution that is unique and that depends continuously on the given data (e.g., boundary conditions or source). Furthermore, the weak formulation is used for a finite element numerical approximation.

For the model-specific technicalities regarding the weak formulation of (2.16), I refer to, e.g., [179, ch. 6] because, hereafter, I illuminate the weak formulation merely abstractly in a Hilbert space setting.

First, let us set $V:=V(\Omega)$ and encode the magnetic vector potential A by the solution function $\mathbf{u}$ which is a member of the Hilbert space $W:=W(\Omega)$. Second, looking ahead to the numerical approximation by the finite element method considered as a special case of the Ritz-Galerkin method (see, e.g., [32, p. 73]), [218, p. 45]), we set the solution function space $W$ equal to the test function space $V$, i.e., $W:=V$. Third, assuming that the Hilbert space's underlying field is $\mathbb{R}$, let the map $a: V \times V \rightarrow \mathbb{R}$ be the bilinear form, and let the map $l: V \rightarrow \mathbb{R}$ be the linear form. Finally, one can state the weak formulation abstractly as

$$
\begin{equation*}
\text { find } \mathbf{u} \in V \text { such that } \forall \mathbf{v} \in V \cdot a(\mathbf{u}, \mathbf{v})=l(\mathbf{v}) \tag{2.17}
\end{equation*}
$$

Remark 2.2.1. In (2.17), the Hilbert space $V$ has to provide a notion of weak derivatives, thus, $V$ has to be a Sobolev space (see, e.g., [218, p. 419f]). For instance, the corresponding spaces in (2.7) are considered as Sobolev spaces.

The boundary conditions are incorporated in the weak formulation, and, conventionally, the excitation is incorporated in the linear form. If the bilinear form satisfies certain requirements such as boundedness and coerciveness and if the linear form is bounded as well, then the weak formulation is well-posed.

Let us consider two restatements of (2.17). By observing that the linear form $l$ is a member of $V^{\prime}$ that is the dual of $V$, one can introduce the so-called natural pairing that is a non-degenerate bilinear map $\left\langle\cdot, \cdot>: V^{\prime} \times V \rightarrow \mathbb{R}\right.$ such that $\left.l(\mathbf{v})=<l, \mathbf{v}\right\rangle$. Hence, the first restatement of (2.17) is

$$
\begin{equation*}
\text { find } \mathbf{u} \in V \text { such that } \forall \mathbf{v} \in V . a(\mathbf{u}, \mathbf{v})=\langle l, \mathbf{v}\rangle \tag{2.18}
\end{equation*}
$$

A benefit of this presentation is that it is an aid in the conceptual distinction of the various quantities involved since moving from the infinite-dimensional to the finitedimensional case, this distinction could be overlooked.

A second restatement of (2.17) is achieved if we only partially evaluate the bilinear form $a$ regarding the first argument such that $a(\mathbf{u}, \cdot): V \rightarrow \mathbb{R}$. One can observe that the map $a(\mathbf{u}, \cdot)$ and the linear form $l$ are members of $V^{\prime}$. By introducing a map $L$ such that $L=\mathbf{u} \mapsto a(\mathbf{u}, \cdot): V \rightarrow V^{\prime}$, one can restate (2.17) as

$$
\begin{equation*}
\text { find } \mathbf{u} \in V \text { such that } \forall \mathbf{v} \in V .(L \mathbf{u})(\mathbf{v})=l(\mathbf{v}), \tag{2.19}
\end{equation*}
$$

where, by omitting additional brackets such as $(\mathrm{L}(\mathbf{u}))(\mathbf{v})$, the conventional order of evaluation is assumed.

Conceiving the map $L$ from (2.19) as a member of the collection hom $\left(V, V^{\prime}\right)$, that is, the collection of all structure-preserving maps from $V$ to $V^{\prime}$, one can represent, e.g., a homogeneous partial differential equation (PDE) by the equation $L(\mathbf{u})=0$ (cf. [96, p. 2]). Hence, compared to (2.17), a benefit of the presentation in (2.18) is the more explicit representation of the mathematical model involved. Examining
parametric mathematical models in the last subsection, another benefit becomes apparent by the logical connection between the parametric mathematical model and its corresponding weak formulation.

In application, the solution function $\mathbf{u}$ is especially utilized to determine an observable physical quantity. Such a quantity is formally encoded in the so-called quantity of interest which is denoted by a non-linear functional $Q: V \rightarrow \mathbb{R}$. For instance, one can express a quantity of interest with an appropriate norm on a space region under investigation $\Omega_{i} \subset \Omega$ (cf. [211, p. 6]) such that one can write

$$
\begin{equation*}
Q=\mathbf{u} \mapsto\|\mathbf{u}\|_{\Omega_{i}}: V \rightarrow \mathbb{R} . \tag{2.20}
\end{equation*}
$$

If one denotes another functional by $q: V \rightarrow \mathbb{R}$ (cf. [178, p. 35]), one can also represent $Q$ in the form

$$
\begin{equation*}
Q=\mathbf{u} \mapsto \int_{\Omega_{i}} q(\mathbf{u}) \mathrm{d} \mathbf{x}: V \rightarrow \mathbb{R} . \tag{2.21}
\end{equation*}
$$

In the magnetoquasistatic model, two common interpretations of the evaluated quantity of interest $Q(\mathbf{u})$ are the magnetic energy and the power loss.

The notion of a quantity of interest is also relevant in the context of error quantification regarding high-fidelity models and corresponding low-fidelity models. We take a closer look at these models in the next chapters.

The authors in [160] discuss the estimation of errors in quantities of interest of two related solution functions $\mathbf{u} \in V$ and $\mathbf{u}_{0} \in V$ - whereas they embed their discussion within the topic of model validation.

The solution function $\mathbf{u}$ is determined by (2.17) which represents a high-fidelity model; the solution function $\mathbf{u}_{0}$ is determined by using a different bilinear form $a_{0}$ in (2.17) which represents a low-fidelity model. A conceivable distinction between the models lies, e.g., in the different modeling of the material properties. Hence, the error regarding the solution function $\mathfrak{E}(\mathbf{u}) \in \mathbb{R}^{+}$and the modeling error $\mathfrak{E}(Q) \in \mathbb{R}^{+}$ can be defined as

$$
\begin{align*}
& \mathfrak{E}(\mathbf{u}):=\left\|\mathbf{u}-\mathbf{u}_{0}\right\|_{V},  \tag{2.22a}\\
& \mathfrak{E}(Q):=\left\|Q(\mathbf{u})-Q\left(\mathbf{u}_{0}\right)\right\|_{l_{2}}, \tag{2.22b}
\end{align*}
$$

where $\|\cdot\|_{V}$ denotes an appropriate norm on $V$ and $\|\cdot\|_{l_{2}}$ denotes the standard $l_{2}$-norm. In (2.22b), choosing the absolute-value norm $|\cdot|$ instead of the standard $l_{2}$-norm is possible as well.

Note that, in (2.22), it is assumed that both solutions are members of the same space $V$. However, in the more generic setting of surrogate optimization, error estimates or error bounds for (2.22) might not exist. There are various situations in which the quantity of interest to be compared has to be represented by two different linear functionals. Some examples are: If $\mathbf{u}$ is determined in a three-dimensional space region and $\mathbf{u}_{0}$ is determined in a two-dimensional space region; or if the space region under investigation (cf. (2.20)) may exhibit different topological properties (see, e.g., Fig. 1.3); or if different numerical methods are employed. In such a generic setting, a comparison relying merely on the real number $\mathfrak{E}(Q)$ conceals the characters of the models under investigation and their relationships. The category theoretical language in ch. 4 provides tools to express formally at least parts of these characters and relationships.

### 2.2.2 Numerical approximation

Recalling (2.16), the main focus of the exposition is on the time-harmonic case, thus, let us solely pay attention to the spatial discretization in the context of the finite element method.

The initial step of this method is the simplicial triangulation $\mathcal{T}_{h}$ of the space region $\Omega$, i.e., the space region is spatially subdivided into a collection of tetrahedra. Notice that if $\Omega \subset \mathbb{R}^{2}$, then the triangulation $\mathcal{T}_{h}$ is a subdivision of $\Omega$ into a collection of triangles. Let us refer to $h$ as mesh size parameter.

The next step is to choose a family of finite dimensional subspaces $V_{h}$ of $V$ such that one can seek the discrete solution $\mathbf{u}_{h} \in V_{h}$ by solving the discrete problem of the weak formulation in (2.17), more precisely,

$$
\begin{equation*}
\text { find } \mathbf{u}_{h} \in V_{h} \text { such that } \forall \mathbf{v} \in V_{h} \cdot a_{h}\left(\mathbf{u}_{h}, \mathbf{v}\right)=l_{h}(\mathbf{v}), \tag{2.23}
\end{equation*}
$$

where the map $a_{h}: V_{h} \times V_{h} \rightarrow \mathbb{R}$ denotes a bilinear form and the map $l_{h}: V_{h} \rightarrow \mathbb{R}$ denotes a linear form. Notice well that, recalling $\S 2.2 .1$, it is tacitly assumed that the spaces' underlying field is $\mathbb{R}$. However, technically speaking, the formulation in (2.16) requires to consider the field of complex numbers $\mathbb{C}$ which, in turn, demands to invoke the notion of a sesquilinear form and an anti-linear form. For the sake of exposition, let us not dwell on these specific technicalities and their implications, though.

By choosing an appropriate basis of $V_{h}$, the corresponding matrix representation of (2.23) expresses the computation of $\mathbf{u}_{h}$ by solving a system of linear equations. For the construction of the finite element subspaces by associating each element of $\mathcal{T}_{h}$ with shape functions and degrees of freedom, I refer to, e.g., [6, p. 82f].

To construct a convergent numerical method that approximates properly the solution $\mathbf{u}$, the family of finite dimensional subspaces $V_{h}$ has to fulfill certain properties (cf. [6, p. 55ff]). For a more elaborated discussion on the consistency, stability, and convergence of numerical methods, see, e.g., [7].

To close this paragraph, let us look closer at a structural property that is related to the detour in § 2.1.2.

Detour 2: a structural perspective on another structural property. Recall the diagrammatic presentation of the algebraic expression in (2.7). I have argued that such an diagram is significant since it provides guidance as one moves from the continuous to the discrete representation. The mimicry of the continuous level's structural property at the discrete level can be encoded by the commuting diagram of the form

where $L_{\text {grad }, h}^{2}(D), \mathbb{L}_{\text {curl }, h}^{2}(D), \mathbb{L}_{\text {div }, h}^{2}(D)$, and $L_{h}^{2}(D)$ denote the finite element subspaces (see, e.g., [178, p. 21]); the differential operators behave polymorphically; and the maps $\pi_{h}^{\text {grad }}, \pi_{h}^{\text {curl }}, \pi_{h}^{\text {div }}, \pi_{h}$ indicate projections (see, e.g., [218, p. 401-405]).

From a purely structural perspective, the essence of the algebraic expression
in (2.24) is that there are different spaces and maps equipped with a notion of composition in a certain context in which one can draw diagrams of the form

which are commutative such that it reflects the equality of various paths. Mind that, in a more generic context, diagrams in (2.25) are not necessarily commutative.

Utilizing the intuition from the detour in $\S 2.1 .2$, one can regard the expression in (2.25) as the syntax, whereas one can consider the expression in (2.24) as a possible semantics. Thus, this example hints more accurately at the style of reasoning employed in the category theoretical language.

### 2.2.3 Parametric mathematical model

Up to this point, the solution function $\mathbf{u}$ is only considered within a space region $\Omega$. In application, though, one is additionally interested in a solution function that is dependent on $N_{\xi}$ parameters where $N_{\xi} \in \mathbb{N}$. These parameters are encoded in the parameter point $\boldsymbol{\xi} \in X \subset \mathbb{R}^{N_{\bar{\xi}}}$. The map $\mathbf{f}=\boldsymbol{\xi} \mapsto \mathbf{u}: X \rightarrow V$ encodes the parametric solution function. The expression $\boldsymbol{f}(\boldsymbol{\xi})$ denotes the solution function for the parameter point $\boldsymbol{\xi}$.

The corresponding partial differential equation depends on $\mathbf{u}$ and on $\xi$ such that the map L has to be extended to $\mathrm{L}: X \times V \rightarrow V^{\prime}$ which is leading to $\mathrm{L}(\boldsymbol{\xi}, \mathbf{u})=0$ (cf. § 2.2.1). Commonly (see, e.g., [96, p. 2f]), it is assumed that the corresponding partial differential equation is well-posed, and $\forall \boldsymbol{\zeta} \in X . \exists!\mathbf{f}(\boldsymbol{\xi}) \in V$.

In the next step, let us adapt the weak formulation in (2.17) in order to state a parametric weak formulation (cf. [94, p. 16]). Therefore, we have to extend the bilinear form and the linear form to $a: V \times V \times X \rightarrow \mathbb{R}$ and $l: V \times X \rightarrow \mathbb{R}$, respectively. The bilinearity and linearity are with respect to the $V$-related arguments. The parametric weak formulation (aka strong-weak formulation) reads as

$$
\begin{equation*}
\text { given } \boldsymbol{\xi} \in X, \text { find } \mathbf{f}(\boldsymbol{\xi}) \in V \text { such that } \forall \mathbf{v} \in V \cdot a(\mathbf{f}(\boldsymbol{\xi}), \mathbf{v}, \boldsymbol{\xi})=l(\mathbf{v}, \boldsymbol{\xi}) . \tag{2.26}
\end{equation*}
$$

Regarding the well-posedness of the formulation (2.26), one has to suppose the requirements of the non-parametric case (2.19). For further details, I refer to, e.g., [94]. Recalling the strong formulation in (2.16), one can observe that the physical meaning of the parameters originates from either the material, the geometry or the source. In general, the individual components of the parameter point $\boldsymbol{\xi}$ can have different physical meanings.

Similarly to (2.21), let us introduce two functionals $Q_{\xi}$ and $\hat{Q}_{\xi}$ : the parametric quantity of interest $Q_{\xi}: V \times X \rightarrow \mathbb{R}$ and the reduced parametric quantity of interest $\hat{Q}_{\xi}: X \rightarrow \mathbb{R}$. Given the matching of the two functionals' codomains such that $\operatorname{cod}\left(\hat{Q}_{\tilde{\xi}}\right) \equiv \operatorname{cod}\left(Q_{\tilde{\xi}}\right)$, it is assumed that the evaluations of the functionals yield the same numerical result, i.e.,

$$
\begin{equation*}
\forall \boldsymbol{\xi} \in X \cdot \hat{Q}_{\tilde{\xi}}(\boldsymbol{\xi})=Q_{\tilde{\xi}}(\mathbf{f}(\boldsymbol{\xi}), \boldsymbol{\xi}) . \tag{2.27}
\end{equation*}
$$

The evaluated functional $\hat{Q}_{\mathcal{\xi}}(\mathcal{\xi})$ can be interpreted as, for instance, the numerical value of the magnetic energy - or the numerical value of the power loss - for certain
geometry parameters. Analogously, the evaluated functional $Q_{\boldsymbol{\xi}}(\mathbf{f}(\boldsymbol{\xi}), \boldsymbol{\xi})$ can be interpreted as the numerical value of the magnetic energy or the power loss; however, it emphasizes the role of the parametric solution function $f$ as well.

For further discussion on parametric mathematical models in the context of magnetoquasistatic Maxwell's theory, I refer to the functional analytic setting, e.g., in [178], and I refer to the differential geometric setting, e.g., in [174].

Remark 2.2.2. Since the functionals' domains do not match, i.e., $\operatorname{dom}\left(\hat{Q}_{\tilde{\xi}}\right) \not \equiv \operatorname{dom}\left(Q_{\tilde{\xi}}\right)$, one cannot conclude from (2.27) that the maps $\hat{Q}_{\xi}$ and $Q_{\xi}$ are equal by function extensionality.

In principle, one should be cautious with the equality of the evaluated quantities of interests such as in (2.27). For instance, consider the so-called magnetic energy functional and the so-called magnetic coenergy functional. Their assignment rules are different and the numerical results of their evaluations are only equal in the linear case (cf. [32, p. 194]).

Additionally, recall the example (E1) in § 1.3 regarding the loss computation of a threedimensional helical coil and the loss computation of a corresponding representation by toroids. Let $\boldsymbol{\xi}$ comprise geometry parameters that are the same in some sense for both the coil and the toroids, let $\hat{Q}_{\xi, 1}(\xi)$ denote the loss of the coil, and let $\hat{Q}_{\xi, 2}(\xi)$ denote the loss of the toroids, respectively. Then, an elemental tool of comparison is to check

$$
\begin{equation*}
\forall \boldsymbol{\xi} \in X . \hat{Q}_{\boldsymbol{\xi}, 1}(\boldsymbol{\xi})=\mathbb{R} \hat{Q}_{\xi, 2}(\boldsymbol{\xi}), \tag{2.28}
\end{equation*}
$$

where the notation $=_{\mathbb{R}}$ indicates a test of equality of real numbers. If the statement in (2.28) holds true, then one can conclude that the maps $\hat{Q}_{\xi, 1}$ and $\hat{Q}_{\bar{\xi}, 2}$ are equal by function extensionality, thus, $\hat{Q}_{\xi, 1}=X \rightarrow \mathbb{R} \hat{Q}_{\xi, 2}$ - such that one can substitute one map for the other. However, the general statement in (2.28) might be undecidable. For the sake of completeness, it is unlikely that the maps are equal by function intensionality as well; because it is unlikely that the internal definitions of the maps are equal. In (3.2) in ch. 3, we encounter a situation similar to the statement in (2.28) from the perspective of approximation theory.

Mind that the previous considerations are relevant from a rather logical analysis viewpoint. Especially, if one imagines other loss computations by in some sense corresponding representations of the helical coil, then it becomes appealing to look out for further tools of comparison at the map level. Putting an emphasis on the map level is a peculiarity of the category theoretical language.

### 2.3 Numerical optimization with the magnetoquasistatic model

Establishing the well-posedness property of a mathematical model is a demanding major task in its own right. This property is a prerequisite for any optimization procedure that is build on top of it. Hence, in the present work, it is assumed that all mathematical models under investigation are well-posed.

Let us begin the section by outlining some theoretical considerations and limitations regarding the optimization theory with partial differential equations and its finite dimensional formulation as a nonlinear optimization problem. A particular feature of the optimization problems is that the evaluation of the objective function or of the constraints or of both requires the solving of a PDE. In the present work, as opposed to its treatment as an explicit equality constraint, the discrete version of the PDE is only considered implicitly within a given optimization problem.

We end the section by an illustration of a subset of optimization test functions and various types of optimization algorithms.

### 2.3.1 Optimization with a partial differential equation

The modern solution theory regarding optimization problems with partial differential equation is in tandem with the modern solution theory regarding partial differential equations; more precisely, it is rooted in the infinite-dimensional Banach space setting and Hilbert space setting, respectively. However, since a thorough discussion in such settings is out of the scope of the present work, let us only consider briefly some aspects in order to be consistent with the abstract discussion in the previous section, and to shine a light on some questions regarding optimization problems. For an in-depth look at the infinite-dimensional case, I refer to, e.g., [96], [210].

Let us use the parametric mathematical model $\mathrm{L}(\boldsymbol{\xi}, \mathbf{u})=0$ from $\S 2.2 .3$ as a basis for the investigation of the objective functional (or cost functional) $J: X \times V \rightarrow \mathbb{R}$. Moreover, let us introduce a Hilbert space $R$ and a closed convex cone $K \subset R$ such that one can define a map $C: X \times V \rightarrow R$ in order to encode an abstract inequality constraint as $C(\boldsymbol{\xi}, \mathbf{u}) \in K$. Hence, one can define abstractly the following optimization problem (cf. [96, p. 2])

$$
\begin{align*}
& \text { minimize } J(\boldsymbol{\xi}, \mathbf{u}) \text { over }(\boldsymbol{\xi}, \mathbf{u}) \in X \times V  \tag{2.29a}\\
& \text { subject to } L(\boldsymbol{\xi}, \mathbf{u})=0, C(\boldsymbol{\xi}, \mathbf{u}) \in K \tag{2.29b}
\end{align*}
$$

where it is assumed that the objective functional $J$ is sufficiently smooth. The evaluation of the objective functional $J$ relies on solving accurately the discrete version of the partial differential equation that is encoded in $\mathrm{L}(\boldsymbol{\xi}, \mathbf{u})=0$. Supposing the wellposedness of the partial differential equation and using the map $\mathbf{f}$ from § 2.2.3, one can redefine the optimization problem as

$$
\begin{gather*}
\min . J(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \text { over }(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \in X \times V,  \tag{2.30a}\\
\text { s.t. } L(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))=0, C(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \in K . \tag{2.30b}
\end{gather*}
$$

Analogously to (2.27), one can define the reduced objective functional $\hat{J}: X \rightarrow \mathbb{R}$ such that $\hat{J}(\boldsymbol{\xi})=J(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))$. Let us consider shortly an instance of the reduced objective functional. For the sake of presentation, I replace the finite-dimensional space $X$ by the infinite-dimensional space $\tilde{X}$ and introduce the variable $\tilde{\zeta}$ such that $\tilde{\xi} \in \tilde{X}$. Hence, the instantiated reduced objective functional $\hat{\jmath}$ reads as

$$
\begin{equation*}
\hat{J}=\tilde{\boldsymbol{\xi}} \mapsto \frac{\alpha}{2}\left\|\mathbf{f}(\tilde{\boldsymbol{\xi}})-\mathbf{y}_{d}\right\|_{A}^{2}+\frac{\beta}{2}\|\tilde{\boldsymbol{\xi}}\|_{B}^{2}: \tilde{X} \rightarrow \mathbb{R}, \tag{2.31}
\end{equation*}
$$

where $\alpha, \beta>0$ (with $0<\alpha+\beta$ ) indicate some fixed scalars for, e.g., weighting or regularization purposes, $\|\cdot\|_{A}$ and $\|\cdot\|_{B}$ indicate some appropriate norms, and $\mathbf{y}_{d}$ denotes a fixed desired solution function. Objective functionals similar to (2.31) are under investigation, for instance, in the context of $\tilde{\xi}$ being the source current density. The corresponding optimization problems deal with the optimal control of electromagnetic fields (see, e.g., [211]).

From an application viewpoint, it is desirable to consider the role of (reduced) parametric quantities of interest regarding the optimization problem in (2.30). There are various possible combinations. For instance, the instantiated reduced objective functional $\hat{J}$ in (2.31) does not necessarily encode a reduced parametric quantity of interest $\hat{Q}_{\tilde{\xi}}$ (see, e.g., [80]). Thus, the value $\hat{Q}_{\tilde{\tilde{\xi}}}(\tilde{\tilde{\xi}})$ would only be determined in a post-optimization step. Nevertheless, it is conceivable to instantiate the reduced
objective functional $\hat{\jmath}$ as

$$
\begin{equation*}
\hat{J}=\tilde{\xi} \mapsto \hat{Q}_{\tilde{\xi}}(\tilde{\xi}): \tilde{X} \rightarrow \mathbb{R}, \tag{2.32}
\end{equation*}
$$

or if a desired value $Q_{d} \in \mathbb{R}$ is provided, then one can instantiate $\hat{\jmath}$ as

$$
\begin{equation*}
\hat{J}=\tilde{\xi} \mapsto\left\|\hat{Q}_{\tilde{\tilde{\xi}}}(\tilde{\xi})-Q_{d}\right\|_{2_{2}}^{2}: \tilde{X} \rightarrow \mathbb{R} . \tag{2.33}
\end{equation*}
$$

If $N_{\hat{Q}}$ quantities of interest with $N_{\hat{Q}} \in \mathbb{N}$ are considered, then a possible instantiation of $\hat{J}$ can be written as

$$
\begin{equation*}
\hat{J}=\tilde{\xi} \mapsto \sum_{i=1}^{N_{\hat{Q}}} \eta_{i} \cdot \mathbb{R} \hat{Q}_{\tilde{\tilde{\xi}}, i}(\tilde{\xi}): \tilde{X} \rightarrow \mathbb{R}, \tag{2.34}
\end{equation*}
$$

or, analogous to (2.33), if $N_{\hat{Q}}$ desired values $Q_{d, 1}, \ldots, Q_{d, N_{\hat{Q}}} \in \mathbb{R}$ are provided, then one can extend (2.34) to the expression

$$
\begin{equation*}
\hat{J}=\tilde{\boldsymbol{\xi}} \mapsto \sum_{i=1}^{N_{\hat{Q}}} \eta_{i} \cdot \mathbb{R}\left\|\hat{Q}_{\tilde{\zeta}, i}(\tilde{\xi})-Q_{d, i}\right\|_{l_{2}}^{2}: \tilde{X} \rightarrow \mathbb{R}, \tag{2.35}
\end{equation*}
$$

where $\eta_{i}>0$ (with, e.g., $\eta_{1}+\cdots+\eta_{N_{\hat{Q}}}=1$ ) denote $N_{\hat{Q}}$ fixed weighting constants and $\cdot \mathbb{R}^{R}$ indicates the standard multiplication on the real numbers. Additionally, one could incorporate the evaluated quantity of interests in the constraints (2.30b) as well.

From a solution theory viewpoint, there are a number of important questions (cf. [96, p. 3f]) regarding the optimization problem in (2.30):
(a) One question is whether there exits an optimal argument for an optimal objective functional value.
(b) An immediate second question is whether this optimal argument is unique.
(c) A third question is concerned whether the optimal argument respects the constraints, hence, whether the optimality conditions, the so-called Karush-KuhnTucker (KKT) conditions, are satisfied. Mostly, first-order necessary optimality conditions are elaborated since the investigation of second-order necessary and sufficient optimality conditions is harder.
(d) The fourth question is concerned with corresponding optimization algorithms. Ideally, the algorithms respect the KKT conditions; thus, in the search of the optimal solution, they rely on information about the first derivative (gradient) or the second derivative (hessian) of, e.g., the objective functional. Generally, such algorithms are guaranteed to find local minimal objective function values of the problem in (2.30). Under certain conditions such as, e.g., convexity, they even find the global minimal objective function value.
For instance, in the case of linear elliptic partial differential equations such as in (2.16), there is much understanding concerning the questions (a) - (d) in the context of the optimization problem in (2.30). However, to the extent of my present understanding, there is still not yet a complete general solution theory regarding the consideration of quantities of interests in different combinations in the optimization problem and for different physical meanings of the parameters.

The above-mentioned algorithms' tendency of finding local minima inspires the introduction of some kind of randomness in the search of a potential global minimum. If the algorithms exhibit some kind of randomness in the search, let us label
them as stochastic; otherwise, let us label them as deterministic. For stochastic algorithms, there is, to my best knowledge, no established theory dealing with guaranteed optimal solutions of an optimization problem such as (2.30).

Finally, in order to solve numerically the optimization problem in (2.30), one has to transform it into a nonlinear optimization problem, i.e., one has to move from the infinite-dimensional case to the finite-dimensional case.

### 2.3.2 Nonlinear optimization problem

Transforming the infinite-dimensional optimization problem in (2.30) into a finitedimensional optimization problem leads to the identification of the spaces involved with subspaces of the standard Euclidean space and the restriction to finitely many equality constraints and inequality constraints. Then, one can define abstractly the so-called nonlinear optimization problem (see, e.g., [158], [18], [35], [142]) as

$$
\begin{align*}
& \min . j(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \text { over }(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \in \mathbb{R}^{N_{\tilde{\xi}}} \times \mathbb{R}^{n},  \tag{2.36a}\\
& \text { s.t. } \forall i \in \mathrm{D} . l_{i}(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))=0,  \tag{2.36b}\\
& \quad \forall i \in \mathrm{E} . c_{i}(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \leq 0, \tag{2.36c}
\end{align*}
$$

where $j: X_{a d} \times V_{a d} \rightarrow \mathbb{R}$ is the smooth objective function, $X_{a d} \subseteq \mathbb{R}^{N_{\bar{\xi}}}$ is the set of admissible parameter points, $V_{a d} \subseteq \mathbb{R}^{n}$ is the set of admissible parametric solution functions for a given parameter point, D is the set of indices for the equality constraints $l_{i}: \mathbb{R}^{N_{\xi}} \times \mathbb{R}^{n} \rightarrow \mathbb{R}$, and E is the set of indices for the inequality constraints $c_{i}: \mathbb{R}^{N_{\tilde{\xi}}} \times \mathbb{R}^{n} \rightarrow \mathbb{R}$. Let us leave the arguments $\boldsymbol{\xi}$ and $\mathbf{f}(\boldsymbol{\xi})$ unchanged since the altered context is clear, thus, there is no risk of confusion. If one incorporates the constraints into a set of admissible arguments $W_{a d}:=X_{a d} \times V_{a d}$, then one can express the optimization problem in (2.36) compactly as

$$
\begin{equation*}
\min _{(\xi, f(\xi)) \in W_{a d}} j(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) . \tag{2.37}
\end{equation*}
$$

Introducing a reduced objective function $\hat{j}: X_{a d} \rightarrow \mathbb{R}$ such that $\hat{j}(\boldsymbol{\xi}):=j(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))$ and invoking set-builder notation to define the set of admissible parameter points $X_{a d}$ as

$$
\begin{equation*}
X_{a d}:=\left\{\xi \in \mathbb{R}^{N_{\xi}}: \forall i \in \mathrm{D} \cdot l_{i}(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))=0 \wedge \forall i \in \mathrm{E} \cdot c_{i}(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \leq 0\right\}, \tag{2.38}
\end{equation*}
$$

one can state the reduced optimization problem compactly as

$$
\begin{equation*}
\min _{\boldsymbol{\xi} \in X_{a d}} \hat{j}(\boldsymbol{\xi}) \tag{2.39}
\end{equation*}
$$

where, technically, one can assume that there are reduced constraint functions $l_{i}(\boldsymbol{\xi})$ and $c_{i}(\boldsymbol{\xi})$ such that $l_{i}(\boldsymbol{\xi}):=l_{i}(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))$ and $c_{i}(\boldsymbol{\xi}):=c_{i}(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))$. A decisive property of the class of nonlinear optimization problems such as in (2.36), in (2.37), and (2.39), respectively, is that the evaluation of the objective function or of the constraints or of both requires the solving of the discrete version of a partial differential equation. In the present work, however, the discrete version of $L(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))=0$ is only implicitly considered (see, e.g., [3]) as opposed to its treatment as an explicit equality constraint (see, e.g., [96] or [210]).

Regarding the above-mentioned class of nonlinear optimization problems, the fundamental assumption is that the numerical simulation of the corresponding mathematical model - in our case, the magnetoquasistatic model - dominates the overall computational costs of the optimization procedure. This assumption inspires the
use of so-called low-fidelity mathematical models in order to reliably accelerate the optimization procedure. Notice that the low-fidelity models are implicitly associated with low computational costs. The chapter 3 is devoted to the discussion about optimization schemes using low-fidelity models.

In ch. 3, we discuss a particular class of optimization schemes that are following the so-called space-mapping paradigm (see, e.g., [125, p.47]). Within this class, there is an emphasis on a reduced objective function $\hat{\hat{j}}$ 。f $: X_{a d} \rightarrow \mathbb{R}$ where $\hat{\hat{j}}: V_{a d} \rightarrow \mathbb{R}$ and the condition $\hat{\hat{j}}(\mathbf{f}(\boldsymbol{\xi}))=j(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi}))$ is applied. Considering (2.39), the condition $\hat{\hat{j}}(\mathbf{f}(\boldsymbol{\xi}))=$ $\hat{j}(\boldsymbol{\xi})$ is supposed, too. Finally, one can state the reduced optimization problem as

$$
\begin{equation*}
\min _{\xi \in X_{a d}}(\hat{j} \circ \mathbf{f})(\boldsymbol{\xi}), \tag{2.40}
\end{equation*}
$$

where the composition operator $\circ: \operatorname{hom}\left(V_{a d}, \mathbb{R}\right) \times \operatorname{hom}\left(X_{a d}, V_{a d}\right) \rightarrow \operatorname{hom}\left(X_{a d}, \mathbb{R}\right)$ is assumed.

Confronted with the optimization problems in (2.37), in (2.39), and in (2.40), let us apply briefly the structural perspective from the detours in $\S$ 2.1.2 and in § 2.2.2. We discuss this perspective more thoroughly in ch. 4.

Detour 3: a structural perspective on the objective functions. Recall the diagrammatic presentation of the abstract algebraic expressions in (2.8) and in (2.25). Using this style of presentation in the context of the optimization problems in (2.37), in (2.39), and in (2.40) results in, among other aspects, stressing the function signatures of the involved objective functions. Therefore, the objective function in (2.37) can be expressed by its assignment rule together with its signature, i.e.,

$$
\begin{equation*}
j=(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})) \mapsto j(\boldsymbol{\xi}, \mathbf{f}(\boldsymbol{\xi})): X_{a d} \times V_{a d} \rightarrow \mathbb{R}, \tag{2.41}
\end{equation*}
$$

the objective function in (2.39) can be expressed by its assignment rule together with its signature, i.e.,

$$
\begin{equation*}
\hat{j}=\boldsymbol{\xi} \mapsto \hat{j}(\boldsymbol{\xi}): X_{a d} \rightarrow \mathbb{R}, \tag{2.42}
\end{equation*}
$$

and the objective function in (2.40) can be expressed by its assignment rule together with its signature, i.e.,

$$
\begin{equation*}
\hat{\hat{j}} \circ \mathbf{f}=\boldsymbol{\xi} \mapsto(\hat{\hat{j}} \circ \mathbf{f})(\boldsymbol{\xi}): X_{a d} \rightarrow \mathbb{R} . \tag{2.43}
\end{equation*}
$$

Recalling (2.28), an elemental tool of comparison for (2.42) and (2.43) is to check

$$
\begin{equation*}
\forall \boldsymbol{\xi} \in X_{a d} \cdot \hat{j}(\boldsymbol{\xi})==_{\mathbb{R}}(\hat{j} \circ \mathbf{f})(\boldsymbol{\xi}) . \tag{2.44}
\end{equation*}
$$

If the statement in (2.44) holds true, then one can conclude that the maps $\hat{j}$ and $\hat{\dot{j}}$ 。f are equal by function extensionality, thus, $\hat{j}=x_{a d} \rightarrow \mathbb{R} \hat{j} \circ \mathbf{f}$ - such that one can substitute one map for the other. However, from a purely structural perspective, the essence of the previous designated objective functions can be captured by four distinct objects $A, B, C$, and $A \times B$ and by five distinct maps $j_{0}, j_{1}, j_{2}, f_{2}$, and $j_{2} \circ f_{2}$ such that one
can draw three diagrams


The benefits of such diagrams are diverse:
(1) Such diagrams disclose pictorially the different decisions for formulating the objective function at the signature level. Hence, they reflect in some sense the available information of the problem at hand.
(2) From the viewpoint of syntax and semantics (recall the detours in $\S$ 2.1.2 and in $\S$ 2.2.2), such diagrams' level of abstraction is particular useful as a unifying guiding tool. For instance, the map $f_{2}$ can be a parametric quantity of interest and the map $j_{2}$ can be chosen such that the map $j_{2} \circ f_{2}$ encodes an objective function similar to (2.33).
(3) Especially in the wider context of validation and verification (see, e.g., [178, p. 11ff] and some references therein such as, e.g., [12], [182], and [160]), the diagrammatic presentation is helpful as a formal organizing tool when we consider various models of various fidelity at the signature level.

### 2.3.3 Optimization algorithms

In order to solve the optimization problem in (2.37), in (2.39) or in (2.40) by means of a computer, we have to apply an appropriate optimization algorithm which seeks iteratively the solution. It is intricate to define a generally accepted taxonomy of optimization algorithms (see, e.g., [158, p. 422]) as it is intricate to define a generally accepted taxonomy of optimization problems (see, e.g., the internet website of the NEOS Server [84]). Therefore, it depends on the user to select an appropriate algorithm for a given problem (cf. [158, p. 2]).

Another user-dependent decision is a software-related issue, more precisely, the choice of the programming language (PL) in which to implement the algorithm. Two dynamically-checked programming languages - and corresponding libraries - are employed: the well-known MATLAB ${ }^{\circledR}$ PL and the relatively young Julia PL. A thorough discussion of this software-related issue is out of the scope of the present work. Though, two of my reasons to utilize the Julia PL are:
(1) its promising outlook to reconcile performance issues and productivity issues;
(2) and its expressive type system and support of functional programming idioms.

The rationale behind point (1) is the observation that performance issues are predominantly discussed at the algorithm level; thus, performance issues at the program level are rarely tackled explicitly in the literature. However, in order to comprehensively assess the Julia PL's capabilities and limitations in comparison to other programming languages, there are much more benchmarks of test problems from various sources needed (see, e.g., [172]).

The rationale behind point (2) is the observation that the category theoretical language (which we encounter in ch. 4) is closely related to type theory and functional programming languages. However, the expressiveness of the Julia PL's type system and the range of its functional programming features are not sufficient to fully match the category theoretical language.

Let us utilize the Julia PL to discuss some widely used optimization algorithms. In our discussion, let us take a pragmatic viewpoint in the sense that we leave the details about the algorithms to the corresponding references and we describe concisely their behavior regarding a subset of test functions of form $f=\left(x_{1}, x_{2}\right) \mapsto$ $f\left(x_{1}, x_{2}\right): U \rightarrow \mathbb{R}$ which are at least members of the differentiability class $C^{1}$ on the open set $U \subset \mathbb{R}^{2}$. The test functions are: Ackley, the Unit sphere, Booth, Rosenbrock, Michalewicz, and the modified Branin (see, e.g., [116], [70], [202]). These test functions (see the Table 2.1) cover various shapes that pose challenges for the algorithms. The Figure 2.2a provides the test functions' surface representation and the Figure 2.2b provides the test functions' contour representation together with a mark of a global minimum $\left(x_{1}^{*}, x_{2}^{*}\right)$. Additionally, the Figure 2.3 provides a close-up of the neighborhood of the test functions' global minimum.

TABLE 2.1: Test functions of form $f=\left(x_{1}, x_{2}\right) \mapsto f\left(x_{1}, x_{2}\right): U \rightarrow \mathbb{R}$.

| Test function $f$ | Definition $f\left(x_{1}, x_{2}\right)$ | Global minimum ( $x_{1}^{*}, x_{2}^{*}$ ) |
| :---: | :---: | :---: |
| Ackley | $\begin{aligned} & \quad-20 \exp \left(-0.2 \sqrt{\frac{1}{2} \sum_{i=1}^{2} x_{i}^{2}}\right)- \\ & \exp \left(\frac{1}{2} \sum_{i=1}^{2} \cos \left(2 \pi x_{i}\right)\right)+20+ \\ & \exp (1) \end{aligned}$ | $(0,0)$ s.t. $f(0,0)=0$ |
| Unit sphere | $\sum_{i=1}^{d=2} x_{i}^{2}$ | $(0,0)$ s.t. $f(0,0)=0$ |
| Booth | $\left(x_{1}+2 x_{2}-7\right)^{2}+\left(2 x_{1}+x_{2}-5\right)^{2}$ | $(1,3)$ s.t. $f(1,3)=0$ |
| Rosenbrock ${ }^{3}$ | $\sum_{i=1}^{2-1} 5\left(x_{i+1}-x_{i}^{2}\right)^{2}+\left(x_{i}-1\right)^{2}$ | $(1,1)$ s.t. $f(1,1)=0$ |
| Michalewicz ${ }^{4}$ | $-\sum_{i=1}^{2} \sin \left(x_{i}\right) \sin ^{20}\left(\frac{i x_{i}^{2}}{\pi}\right)$ | $\begin{aligned} & (2.20,1.57) \text { s.t. } \\ & f(2.20,1.57)=-1.801 \end{aligned}$ |
| Modified Branin ${ }^{5}$ | $\begin{array}{r} 1 \cdot\left(x_{2}-\frac{5.1}{4 \pi^{2}} x_{1}^{2}+\frac{5}{\pi} x_{1}-6\right)^{2}+ \\ 10 \cdot\left(\left(1-\frac{1}{8 \pi}\right) \cos \left(x_{1}\right)+1\right)+5 x_{1} \end{array}$ | $\begin{aligned} & (-3.689,13.630) \text { s.t. } \\ & f(-3.689,13.630)=-16.644 \end{aligned}$ |

Let us set up the nonlinear optimization problem under test by choosing the respective test function as the objective function and by defining the admissible

[^10]sets $\mathfrak{F}_{1}$ (also known as box constraints), $\mathfrak{F}_{2}$, and $\mathfrak{F}:=\mathfrak{F}_{1} \cap \mathfrak{F}_{2}$ such that
\[

$$
\begin{align*}
& \mathfrak{F}_{1}:=\left\{\left(x_{1}, x_{2}\right) \in \mathbb{R} \times \mathbb{R}: x_{1, l}-x_{1} \leq 0 \wedge x_{1}-x_{1, u} \leq 0 \wedge x_{2, l}-x_{2} \leq 0 \wedge x_{2}-x_{2, u} \leq 0\right\},  \tag{2.46a}\\
& \mathfrak{F}_{2}:=\left\{\left(x_{1}, x_{2}\right) \in \mathbb{R} \times \mathbb{R}:\left(x_{1}-x_{1}^{*}\right)^{2}+\left(x_{2}-x_{2}^{*}\right)^{2}-r_{0}^{2} \leq 0\right\}, \tag{2.46b}
\end{align*}
$$
\]

where the point $\left(x_{1, l}, x_{1, u}\right)$ and the point $\left(x_{2, l}, x_{2, u}\right)$ represent the lower and upper bound of the component $x_{1}$ and $x_{2}$, respectively; the point ( $x_{1}^{*}, x_{2}^{*}$ ) denotes an optimal argument of the respective test function that is known apriori; and the scalar $r_{0}$ encodes the radius of a disk $D\left(M, r_{0}\right)$ with the midpoint $M:=\left(x_{1}^{*}, x_{2}^{*}\right)$. The scalar $r_{0}$ is set to 10 , i.e., $r_{0}:=10$, and the lower and upper bounds are set such that the optimal argument's quadrant is captured, e.g., for the Rosenbrock function in Fig. 2.2b, it holds that $\left(x_{1, l}, x_{1, u}\right)=(0,10)$ and $\left(x_{2, l}, x_{2, u}\right)=(0,10)$. In the cases, in which the optimal argument is the zero point (see (i) and (ii) in Fig. 2.2b), the first quadrant is selected.

The test functions can be composed of generic Julia functions, hence, one can employ the package ForwardDiff. jl (see [175]) in order to determine derivative information by forward mode automatic differentiation (see, e.g., [81], [156]). Since it is assumed that the six test functions in Table 2.1 are at least members of $C^{1}, \mathrm{I}$ depict in Figure 2.4 the image of $U$ under the vector field $\operatorname{grad} f$, i.e., $\operatorname{grad}(f)\left(x_{1}, x_{2}\right)$ where ( $x_{1}, x_{2}$ ) $\mathcal{U}$, as a projection on the test functions' contour representation (see Figure 2.2b and Figure 2.3b). ${ }^{6}$

Mind that if one assumes that the domain $U$ is convex, that is, the condition

$$
\begin{equation*}
\forall p_{1}, p_{2} \in U . \forall t \in[0,1] .(1-t) p_{1}+_{\mathbb{R}^{2}} t p_{2} \in U \tag{2.47}
\end{equation*}
$$

holds to be true; and if one assumes that, e.g., the condition

$$
\begin{equation*}
\forall p_{1}, p_{2} \in U . f\left(p_{1}\right) \geq f\left(p_{2}\right)+\operatorname{grad}(f)\left(p_{2}\right) \cdot \mathbb{R}^{2}\left(p_{1}-\mathbb{R}^{2} p_{2}\right) \tag{2.48}
\end{equation*}
$$

holds to be true as well, then one can conclude that a test function $f$ is convex. However, for arbitrary domains and arbitrary functions in practical applications, in most cases, one cannot exploit analytical examinations of the convexity property. Furthermore, by means of numerical examinations, one can only test the condition in (2.47) and the condition in (2.48) for some $p_{1}, p_{2} \in U$; hence, one can gain at most a clue for convexity. In the present work, therefore, I refrain from elaborating on the convexity property for all the functions under consideration.

Assuming a map $s_{1}$ and a map $s_{2}$ who share the same signature that reads as $C^{1}(U, \mathbb{R}) \times U \rightarrow \mathbb{R}^{+}$, one can conceive these maps as local first-order sensitivity measures if one defines their assignment rules as

$$
\begin{equation*}
s_{1}(f, p):=\left(\partial_{\mathbf{e}_{x_{1}}}(f)(p)\right)^{2}, \quad s_{2}(f, p):=\left(\partial_{\mathbf{e}_{x_{2}}}(f)(p)\right)^{2} \tag{2.49}
\end{equation*}
$$

Hence, let us deploy a gradient-based interpretation of sensitivity measures (see, e.g., [129]). ${ }^{7}$ In Figure 2.5, $s_{1}(f, p)$ and $s_{2}(f, p)$ are depicted w.r.t. Figure 2.3b.

[^11]Exploiting the assignment rules in (2.49), one can define the maps $S_{1}$ and $S_{2}$ that possess the same signature, that is, $C^{1}(U, \mathbb{R}) \rightarrow \mathbb{R}^{+}$, whose assignment rules read as

$$
\begin{equation*}
S_{1}(f):=\int_{U} s_{1}(f, \mathbf{x}) \mathrm{d}^{2} \mathbf{x}, \quad S_{2}(f):=\int_{U} s_{2}(f, \mathbf{x}) \mathrm{d}^{2} \mathbf{x} \tag{2.50}
\end{equation*}
$$

Thus, one can conceive the maps $S_{1}$ and $S_{2}$ as global first-order sensitivity measures. In addition, one can define normalized global first-order sensitivity measures $S_{1}^{N}$ and $S_{2}^{N}$ whose assignment rules read as

$$
\begin{equation*}
S_{1}^{\mathrm{N}}(f):=\frac{S_{1}(f)}{\sum_{i=1}^{2} S_{i}(f)}, \quad S_{2}^{\mathrm{N}}(f):=\frac{S_{2}(f)}{\sum_{i=1}^{2} S_{i}(f)} \tag{2.51}
\end{equation*}
$$

where $\forall f . S_{1}^{\mathbb{N}}(f)+S_{2}^{\mathbb{N}}(f)=1.0$ holds to be true in exact arithmetic.
By using the package HCubature.jl (see [105]), $S_{1}^{N}(f)$ and $S_{2}^{N}(f)$ in (2.51) can be computed by means of numerical integration with regard to the Figure 2.2b and with regard to the Figure 2.3b. In Table 2.2 and in Table 2.3, respectively, the corresponding results are presented. ${ }^{8}$

Table 2.2: The normalized global first-order sensitivity measure $S_{i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the Figure 2.2b.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $S_{1}^{\mathrm{N}}(f)$ | 0.5000 | 0.5000 | 0.4894 | 0.9965 | 0.3702 | 0.7234 |
| $S_{2}^{\mathrm{N}}(f)$ | 0.5000 | 0.5000 | 0.5106 | 0.0035 | 0.6298 | 0.2766 |
| $\Sigma_{i=1}^{2} S_{i}^{\mathrm{N}}(f)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

TAbLE 2.3: The normalized global first-order sensitivity measure $S_{i}^{N}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the Figure 2.3b.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $S_{1}^{\mathrm{N}}(f)$ | 0.5000 | 0.5000 | 0.5000 | 0.9109 | 0.2595 | 0.9277 |
| $S_{2}^{\mathrm{N}}(f)$ | 0.5000 | 0.5000 | 0.5000 | 0.0891 | 0.7405 | 0.0723 |
| $\Sigma_{i=1}^{2} S_{i}^{\mathrm{N}}(f)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

If we consult the Table 2.1, then one can assert that the results in Table 2.2 and in Table 2.3 seem to pass some plausibility checks. More precisely: The choice of the domain $U$ can have an impact on the sensitivity measures if the corresponding evaluated test function does not show some kind of symmetry. In the cases from (i) to (iii), for all points $p \in U$ the evaluated squared instantaneous rate of change is, roughly speaking, equal w.r.t. both variables $x_{1}$ and $x_{2}$; whereas, in the cases from (iv) to (vi), for all points $p \in U$ the evaluated squared instantaneous rate of change is, roughly speaking, either greater w.r.t. $x_{1}$ or greater w.r.t. $x_{2}$. Hence, from a practical applications viewpoint, the Table 2.2 and the Table 2.3 furnish us with a valuable

[^12]quantitative screening of the importance of the variables regarding the respective test function.

Mind that the elaborations are without loss of generality regarding the number of parameters $N_{\tilde{\zeta}}$ (recall $\S 2.2 .3$ ). To exemplify this kind of generality, let us explore the Rosenbrock test function in Table 2.1 since, without much ado, it is amenable to the number of parameters $N_{\xi}$ with $N_{\xi} \in\{2,3,4,5,6,7\}$. Hence, the $N_{\xi}$-dimensional Rosenbrock test function $f_{\mathrm{R}_{N_{\xi}}}$ can be written as

$$
\begin{equation*}
f_{\mathrm{R}_{N_{\bar{\xi}}}}=x \mapsto f_{\mathrm{R}_{N_{\bar{\xi}}}}(x):=\sum_{i=1}^{N_{\tilde{\xi}}-1} 5\left(x_{i+1}-x_{i}^{2}\right)^{2}+\left(x_{i}-1\right)^{2}: U_{N_{\tilde{\xi}}} \rightarrow \mathbb{R} \tag{2.52}
\end{equation*}
$$

where $N_{\tilde{\xi}} \in\{2,3,4,5,6,7\}$, and $U_{N_{\tilde{\xi}}}$ is an open set $U_{N_{\tilde{\xi}}} \subset \mathbb{R}^{N_{\tilde{\xi}}}$. Notice that, in each case of $N_{\xi}$, the global minimum w.r.t. (2.52) is at the point $(1,1, \ldots, 1) \in \mathbb{R}^{N_{\tilde{\xi}}}$. Adapting the normalized global first-order sensitivity measures in (2.51) to the use case of the $N_{\xi^{-}}$ dimensional Rosenbrock test function $f_{\mathrm{R}_{N_{\bar{\xi}}}}$ I report in Table 2.4 the corresponding results.The observable pattern is reasonable if one unrolls the term $f_{\mathrm{R}_{N_{\xi}}}(x)$ in (2.52).

Table 2.4: The normalized global first-order sensitivity measure $S_{i}^{\mathrm{N}}$ evaluated at $f_{\mathrm{R}_{N_{\xi}}}$ w.r.t. the domain $[-2.0,2.0]^{N_{\tilde{\xi}}}$ with $N_{\tilde{\xi}} \in\{2,3,4,5,6,7\}$.

| $N_{\xi}$ | $S_{i}^{\mathrm{N}}(f)$ | $i:=1$ | $i:=2$ | $i:=3$ | $i:=4$ | $i:=5$ | $i:=6$ | $i:=7$ | $\sum_{i=1}^{N_{\bar{\xi}}} S_{i}^{\mathrm{N}}(f)$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2 | 0.9109 | 0.0891 | - | - | - | - | - | 1.0000 |  |
| 3 | 0.4008 | 0.5600 | 0.0392 | - | - | - | - | 1.0000 |  |
| 4 | 0.2569 | 0.3590 | 0.3590 | 0.0251 | - | - | - | 1.0000 |  |
| 5 | 0.1892 | 0.2641 | 0.2641 | 0.2641 | 0.0185 | - | - | 1.0000 |  |
| 6 | 0.1494 | 0.2090 | 0.2090 | 0.2090 | 0.2090 | 0.0146 | - | 1.0000 |  |
| 7 | 0.1239 | 0.1728 | 0.1728 | 0.1728 | 0.1728 | 0.1728 | 0.0121 | 1.0000 |  |

Finally, let us invoke four packages that contain various types of optimization algorithms:
(Opkg1) the package NLopt.jl (see [107]) provides an interface to the open-source NLopt library for nonlinear optimization,
(Opkg2) the package BlackBoxOptim.jl(see [64]) which provides some meta-heuristic ${ }^{9}$ stochastic algorithms for global optimization,
(Opkg3) the package optim.jl (see [151]) provides some deterministic and stochastic algorithms for box-constrained local and global optimization, and
(Opkg4) the package Intervaloptimisation.jl (see [186]) which provides guaranteed deterministic global optimization algorithms using interval arithmetic.

From (Opkg1), I employ two gradient-based local optimization algorithms more precisely, a sequential quadratic programming (SQP) algorithm based on [128] and a method of moving asymptotes (MMA) algorithm based on [203] - on the admissible set $\mathfrak{F}$; I apply two derivative-free local optimization algorithms - i.e., a

[^13]Nelder-Mead simplex (NMS) algorithm based on [176] and a constrained optimization by linear approximations (COBYLA) algorithm based on [170] - on the admissible set $\mathfrak{F}_{1}$ and on the admissible set $\mathfrak{F}$, respectively; and I apply two derivative-free global optimization algorithms - that is, the dividing rectangles (DIRECT) algorithm based on [109] and a modified evolutionary algorithm (MEA) based on [193] - on the admissible set $\mathfrak{F}_{1} .{ }^{10}$

From (Opkg2), I pick an adaptive differential evolution (ADE) algorithm from a collection of stochastic algorithms in order to perform a stochastic global optimization on the admissible set $\mathfrak{F}_{1}$.

From (Opkg3), I utilize a primal interior-point algorithm (see, e.g., [158, ch. 19] or [116, ch. 10.9]) on the admissible set $\mathfrak{F}_{1}$ in which I employ for the inner optimization algorithms a gradient-based, i.e., a limited-memory Broyden-Fletcher-Gold-farb-Shanno (L-BFGS) algorithm (see, e.g., [158, ch. 7.2]); a derivative-free, i.e., a Nelder-Mead simplex algorithm (see above); and a stochastic global, i.e., a particle swarm (PS) algorithm based on [228].

By providing the initial point $\left(\hat{x}_{1}, \hat{x}_{2}\right)=(1.1,1.1)$, let us check exemplarily the Ackley function's global optimal argument $\left(x_{1}^{*}, x_{2}^{*}\right)=(0.0,0.0)$ with the corresponding optimal function value $f\left(x_{1}^{*}, x_{2}^{*}\right)=0.0$ (see (i) in Figure 2.2b). As expected, all algorithms find the optimal solution within a certain numerical tolerance (see Table 2.5). But also as expected, choosing an initial point closer to the admissible set's borders, the gradient-based local optimization algorithms tend to be trapped in one of the Ackley function's many local minima. Similarly, the behavior of the algorithms with respect to the other test functions (see (ii)-(vi) in Figure 2.2b) - that has been well investigated in the literature - can be recapitulated.

Table 2.5: Check exemplarily the Ackley function's global optimum.

| Opkg | SQP | MMA | NMS | COBYLA | DIRECT | MEA | ADE | L-BFGS | PS |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 1 | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ | $\checkmark$ |  |  |  |
| 2 |  |  |  |  |  |  | $\checkmark$ |  |  |
| 3 |  |  | $\checkmark$ |  |  |  |  | $\checkmark$ | $\checkmark$ |

In order to assess the ambit of the solution found, a common practice in many applications is: Apply a global optimization algorithm; and use its solution as a starting point for a local optimization algorithm. However, another possibility to assess the area of validity is to use interval arithmetic (see, e.g., [213], [104]) in the context of deterministic global optimization. ${ }^{11}$ In (Opkg4), such a possibility is pursued by a Moore-Skelboe (MS) algorithm (see, e.g, [59]). Mind that the result is not comprised of the optimal component values $\left(x_{1}^{*}, x_{2}^{*}\right)$ and the optimal function value $f\left(x_{1}^{*}, x_{2}^{*}\right)$ as with the aforementioned algorithms; instead the result is comprised of intervals that contain guaranteed the optimal component values $\left[x_{1, l}^{*}, x_{1, u}^{*}\right] \times\left[x_{2, l}^{*}, x_{2, u}^{*}\right]$ and the optimal function value $\left[f\left(x_{1, l}^{*}, x_{2, l}^{*}\right), f\left(x_{1, u}^{*}, x_{2, u}^{*}\right)\right]$.

Finally, when one moves from the test functions such as in Figure 2.2a to functions from applications, one has to recall two common issues:

[^14]- The test functions exhibit a fairly complete picture that facilitates the choice of an appropriate algorithm. However, in many applications, the choice of an appropriate optimization algorithm for a given problem is difficult due to incomplete preliminary information. Hence, there is also no preference for a particular type of optimization algorithms.
- A test function evaluation is computationally cheap. However, in many applications, the evaluation of the objective function or the constraint functions or both (see § 2.3.1) depends on a computationally expensive numerical simulation (see § 2.2) such that, presumably, an exhaustive coverage of the parameter space is prohibitive. Hence, an exhaustive reconstruction of the shape (or landscape) that can be associated with the function under investigation is unlikely.

We encounter these two issues again in the upcoming chapter 3 that is concerned with the discussion about optimization schemes using low-fidelity models.

Moreover, in chapter 5, we consider high-fidelity optimization problems as a concrete instances of the abstract optimization problem in (2.36) where the semantics of the magnetoquasistatic model is applied. More precisely, we encounter functions that encode, for instance, the time-averaged ohmic loss, the time-averaged ohmic loss density or the inductance at different operating frequencies. Hence, the investigation presented in this section is valuable as a preliminary study and anchor point to develop and assess the studies of chapter 5 .

(A) Surface representation of $z:=f\left(x_{1}, x_{2}\right)$.

(в) Contour representation of $z:=f\left(x_{1}, x_{2}\right)$.

The red cross indicates a global minimum.

Figure 2.2: Representations of the test functions in Table 2.1.
(i) Ackley, (ii) Unit sphere, (iii) Booth,
(iv) Rosenbrock, (v) Michalewicz, (vi) Modified Branin.

(в) Contour representation of $z:=f\left(x_{1}, x_{2}\right)$.

The red cross indicates a global minimum.

Figure 2.3: Representations of the test functions in Table 2.1 (highlighting the neighborhood of the global minimum).
(i) Ackley, (ii) Unit sphere, (iii) Booth,
(iv) Rosenbrock, (v) Michalewicz, (vi) Modified Branin.

(A) Depicting $\operatorname{grad}(f)\left(x_{1}, x_{2}\right)$ within Figure 2.2b.

(B) Depicting $\operatorname{grad}(f)\left(x_{1}, x_{2}\right)$ within Figure 2.3b.

Figure 2.4: Depicting $\operatorname{grad}(f)\left(x_{1}, x_{2}\right)$ with $\left(x_{1}, x_{2}\right) \in U$ as a projection on the contour representation of the test functions in Table 2.1.
(i) Ackley, (ii) Unit sphere, (iii) Booth,
(iv) Rosenbrock, (v) Michalewicz, (vi) Modified Branin.

(A) Depicting $s_{1}\left(f,\left(x_{1}, x_{2}\right)\right)$ w.r.t. Figure 2.3b.

Dark colors indicate low values; bright colors indicate high values.

(в) Depicting $s_{2}\left(f,\left(x_{1}, x_{2}\right)\right)$ w.r.t. Figure 2.3b. Dark colors indicate low values; bright colors indicate high values.

Figure 2.5: Depicting $s_{i}\left(f,\left(x_{1}, x_{2}\right)\right)$ from (2.49) with $i \in\{1,2\}$ and $\left(x_{1}, x_{2}\right) \in U$ w.r.t. the test functions in Table 2.1.
(i) Ackley, (ii) Unit sphere, (iii) Booth,
(iv) Rosenbrock, (v) Michalewicz, (vi) Modified Branin.

### 2.4 In closing

The chapter's primary purpose has been to lay out the technical landscape in which the remaining chapters are placed. The languages of vector analysis, differential geometry, and functional analysis served as a methodological and terminological guidance for formulating the relevant notions.

More precisely, we have elaborated the magnetoquasistatic model of Maxwell's theory by presenting the fundamental problem statement of electromagnetism and the corresponding system of Maxwell's equations. From this system, we have derived the magnetoquasistatic subsystem and the magnetostatic subsystem.

Using the magnetoquasistatic model as a directing representative, we have examined its numerical simulation in the common procedure, i.e., we have recapitulated concisely the concepts of the weak formulation, the numerical approximation, and the parametric mathematical model.

Finally, we have discussed notions regarding the optimization with a partial differential equation and its relation to nonlinear optimization problems. We have sketched various types of optimization algorithms and we have outlined a subset of optimization test functions. By deploying a gradient-based interpretation of sensitivity measures to the test functions - which permit the determination of derivative information by forward mode automatic differentiation -, we have completed the discussion.

## Chapter 3

## Surrogate optimization

In this chapter, I provide an in-depth elaboration of the key notion surrogate optimization. Furthermore, I provide an in-depth elaboration of the proposed partitioning of this notion in § 1.2 into the three sub-notions: (1) surrogate modeling \& simulation, (2) surrogate-based optimization, and (3) surrogate-guided optimization.

Within the limited scope of the explanations, let us anticipate consistently algebraic tools from the category theoretical language in ch. 4 in order to tag the various notions of surrogate optimization with algebraic notes. Additionally, these algebraic tools facilitate the smooth transition between the various layers in Figure 1.4.

Concerning the sub-notion (1) surrogate modeling \& simulation, let us forge an abstract setting in which we state common classes of mathematical problems. Within the context of these classes, we embed the notion of a high-fidelity model and a low-fidelity model. Subsequently, we define the high-fidelity approximation error, the notion of a sampling plan, and the empirical surrogate modeling error as one among other indicators within surrogate optimization. Afterwards, let us attempt to sketch an holistic understanding of some deterministic data-fit low-fidelity models, i.e., multivariate polynomials and radial-basis functions, and some probabilistic data-fit low-fidelity models, i.e., kriging low-fidelity models. We close this subpart by applying a formalization-driven perspective on simplified-physics low-fidelity models.

Concerning the sub-notion (2) surrogate-based optimization, let us examine the optimization with the test functions in $\S 2.3 .3$ by data-fit low-fidelity models and by emulated simplified-physics low-fidelity models. We carve out a numerical scaffolding of a benchmark-focused classification of test functions (more generally, highfidelity models) and we elucidate different procedures to find a solution of the highfidelity optimization problem.

Concerning the sub-notion (3) surrogate-guided optimization, let us dwell briefly on the sequential kriging optimization and its construction principles as a subkind of the model management strategy adaptation. Afterwards, we dwell on optimization procedures within the space-mapping paradigm which are a subkind of the model management strategy adaptation; and we dwell on the basic building blocks of the co-kriging optimization which can be seen as a subkind of the model management strategy fusion. By applying a formalization-oriented viewpoint, it is attempted to illuminate potential hybrid model management strategies and to pin down properly, e.g., the conceptional distinction between a low-fidelity model and a surrogate model within the space-mapping paradigm. Furthermore, driven by heuristics, we construct formally some statements to provide a novel access to the delicate aspect of convergence-related issues regarding the optimization within the space-mapping paradigm and the co-kriging optimization.

### 3.1 Surrogate modeling \& simulation

Notice well that, due to the work of so many diverse research communities in the vast field of surrogate optimization, it seems impossible to provide a unifying methodological and terminological guidance concerning surrogate optimization that suits every research community.

Regarding probability low-fidelity models, for instance, there is the delicate aspect of the interpretation of probability (see, e.g., [201, p. 29ff]). One main interpretation leads to the school of thought called Bayesian statistics (see, e.g., [155, ch. 5]), another main interpretation leads to the school of thought called Frequentist statistics (see, e.g., [155, ch. 6]).

Therefore, let a general guiding principle of ours be that we aim at being as indifferent as possible to potential issues of interpretation or semantics.

Thus, similarly to a first principles approach, we focus on stripping surrogate optimization down to a bare syntactical minimum - and, then, to argue from this bare syntactical minimum, adding layers of syntax and semantics when they are needed.

### 3.1.1 An abstract setting

After discussing different classes of mathematical problems, we discuss the concepts high-fidelity function approximation error, sampling plan, and empirical surrogate modeling error. Among others, we introduce the empirical generalization error and the squared sample Pearson correlation coefficient (SSPCC). We close this subsection by illuminating a link between the SSPCC and the normalized global first-order sensitivity measures (see § 2.3.3).

## Classes of mathematical problems

In the previous chapter, we have encountered the concepts of modeling, simulation and optimization in the context of the magnetoquasistatic Maxwell's theory. If we apply a map-based viewpoint, one can assign abstractly each of these concepts to one of the following classes of mathematical problems:

$$
\begin{align*}
& \text { given } x \in X \text { and } y \in Y \text {, find } \mathrm{K} \in \operatorname{hom}(X, Y) \text { such that } \mathrm{K}(x)=y,  \tag{3.1a}\\
& \text { given } \mathrm{K} \in \operatorname{hom}(X, Y) \text { and } x \in X \text {, find } y \in Y \text { such that } \mathrm{K}(x)=y,  \tag{3.1b}\\
& \text { given } \mathrm{K} \in \operatorname{hom}(X, Y) \text { and } y \in Y \text {, find } x \in X \text { such that } \mathrm{K}(x)=y, \tag{3.1c}
\end{align*}
$$

where, for instance, K denotes a linear map, $X$ and $Y$ denote linear spaces over an underlying field $\mathbb{F}$, and $\operatorname{hom}(X, Y)\left({\left.\text { or } \operatorname{hom}_{\mathbb{F}}(X, Y)\right) \text { connotes a vector space as well. }}_{\text {. }}\right.$

Following the terminology in [140, p. 23], let us call the problems of the form in (3.1a) as identification problems, the form in (3.1b) as direct problems, and the form in (3.1c) as inverse problems. ${ }^{1}$ Thus, in the context of the previous chapter, one can assign modeling to (3.1a), simulation and optimization to (3.1c).

Observe that, for example, the evaluation of the reduced parametric quantity of interest $\hat{Q}_{\xi}(\boldsymbol{\xi})$ (see $\S 2.2 .3$ ) can be assigned to (3.1b). Seizing this example, let us pin down a few notions regarding a surrogate model. Some chunks of approximation theory and statistical learning theory are utilized which aid us to frame coherently and to state economically the necessary notions.

[^15]Let us return to the statement in (2.28) and reformulate it slightly using the terms in (3.1); hence, we deal with the statement

$$
\begin{equation*}
\forall x \in X . \mathrm{K}(x)=\gamma \tilde{\mathrm{K}}(x), \tag{3.2}
\end{equation*}
$$

where $\mathrm{K}, \tilde{\mathrm{K}}: \mathrm{X} \rightrightarrows \mathrm{Y}$. In calculations that are of practical interest, it is assumed that the map K possesses certain undesired properties, e.g., its evaluation is exceeding reasonable finite computing time budgets or it is not straightforwardly available for operations such as differentiation or integration. A consequence of these properties is that application-oriented optimizations relying on the map K are prohibitive. Therefore, the aim is to surrogate the map K with the map $\tilde{\mathrm{K}}$ that possesses userprescribed desired properties. Commonly, the map K is called a high-fidelity model to emphasize the user's prescribed assessment of the model's predictive power or a high-cost model - to emphasize the user's prescribed assessment of the model's computational costs. Then, the map $\tilde{\mathrm{K}}$ is called a low-fidelity model, a low-cost model, a meta-model or a surrogate model.

Technically, one can substitute the map $K$ for the map $\tilde{K}$ if the statement in (3.2) holds true such that the maps are equal by function extensionality, thus, $\mathrm{K}={ }_{X \rightarrow Y} \tilde{\mathrm{~K}}$. However, this line of thought is left to the next chapter since the usual starting point regarding surrogate models focuses on, e.g., interpolating polynomials or splines and linear or nonlinear regression models. Inspired by the origins of these concrete examples in approximation theory and statistical learning theory, one can define the class of data-fit low-fidelity models which can be subdivided into the subclass of deterministic low-fidelity models, and the subclass of probabilistic low-fidelity models (see, e.g., [76, p. 132], [116, p. 275]). Facing numerical simulations, one can additionally define the class of projection-based low-fidelity models and the class of simplifiedphysics low-fidelity models.

Recalling $\S 1.2$, the class of projection-based low-fidelity models is not pursued. Furthermore, mind that the term meta-model is primarily a paraphrase for the term data-fit low-fidelity model, and vice versa; and that, in the context of the space mapping paradigm or the defect correction paradigm, the term low-fidelity model and the term surrogate model are distinguished (see, e.g., [194, p. 28f] or [49, p. 56f]).

It is assumed that a high-fidelity model is deterministic in the sense that repeated use of the same input results in the same output each time (see, e.g., [184, p. 409]). If the choice of a high-fidelity model is, e.g., the reduced parametric quantity of interest $\hat{Q}_{\xi}$ that is attained by a FE simulation, then a deterministic FE simulation is considered as opposed to a stochastic FE simulation (see, e.g., [74]). However, some kind of randomness in the form of noise $\varepsilon$ in the image $\hat{Q}_{\tilde{\xi}}(\tilde{\xi})$ is taken into account but this noise does not stem from randomness in the argument $\boldsymbol{\xi}$. The noise $\varepsilon$ encodes, for instance,

- "random errors from unobserved variables" [61, p. 39],
- errors in the presence of a "parameter controlling missing physics" [201, p. 93],
- or a "systematic error ... caused by insufficient mesh resolution" [70, p. 5].

Hence, let us summarize in a single observational noise $\varepsilon$ all kinds of noise such as the computational noise with regard to the mesh size parameter $h$ (recall § 2.2.2) and represent this observational noise as a random variable, even though we utilize a deterministic high-fidelity model. Mind that this representation is a usual trick in order to make the machinery of probabilistic low-fidelity models amenable to deterministic high-fidelity models (see, e.g., [70, p. 5]).

## High-fidelity function approximation error

Let us suppose that $X, Y$, and $Y^{X} \equiv \operatorname{hom}(X, Y)$ comprise a more finely layered structure, more precisely, let us suppose that they are normed linear spaces and they are equipped with corresponding norm-induced metrics $d_{X}, d_{Y}$, and $d_{Y^{X}}$ such that

$$
\begin{align*}
& d_{X}=\left(x_{1}, x_{2}\right) \mapsto\left\|x_{1}-x_{2}\right\|_{X}: X \times X \rightarrow \mathbb{R}^{+},  \tag{3.3a}\\
& d_{Y}=\left(y_{1}, y_{2}\right) \mapsto\left\|y_{1}-y_{2}\right\|_{Y}: Y \times Y \rightarrow \mathbb{R}^{+},  \tag{3.3b}\\
& d_{Y^{X}}=\left(\mathrm{K}_{1}, \mathrm{~K}_{2}\right) \mapsto\left\|\mathrm{K}_{1}-\mathrm{K}_{2}\right\|_{Y^{X}}: Y^{X} \times Y^{X} \rightarrow \mathbb{R}^{+} . \tag{3.3c}
\end{align*}
$$

Notice well that, for the sake of simplicity, it is merely tacitly assumed that, for all elaborations, additional properties such as compactness, Lipschitz continuity, and the like hold to be true if the respective notions require these properties.

If we choose a surrogate model $\tilde{\mathrm{K}}$ from a prescribed class of functions called hypothesis space H (cf. [51, p. 9ff]), that is, $\tilde{\mathrm{K}} \in \mathrm{H}$ and $\mathrm{H} \subseteq Y^{X}$, then we introduce an error which one can capture by the metrics in (3.3). The theoretical capability of a surrogate model to approximate accurately a high-fidelity model can be investigated by an analysis of the convergence of a sequence of surrogate models $\left(\tilde{\mathrm{K}}_{n}\right)_{n \in \mathbb{N}}$ to the high-fidelity model K . It is assumed that $\tilde{\mathrm{K}}_{n}: X \rightarrow Y$ and $\left(\tilde{\mathrm{K}}_{n}\right)_{n \in \mathbb{N}}=n \mapsto \tilde{\mathrm{~K}}_{n}: \mathbb{N} \rightarrow Y^{X}$.

Often, the notion of uniform convergence (see, e.g., [183, p. 147ff] or [169, p. 203ff]) is deployed where it is set that

$$
\begin{equation*}
d_{Y^{X}}\left(\mathrm{~K}_{1}, \mathrm{~K}_{2}\right):=\sup _{x \in X}\left\|\mathrm{~K}_{1}(x)-\mathrm{K}_{2}(x)\right\|_{Y}, \tag{3.4}
\end{equation*}
$$

such that, in abbreviated form, $\left(\tilde{\mathrm{K}}_{n}\right)_{n \in \mathbb{N}} \rightarrow \mathrm{~K}$ denotes the convergence of the sequence to the limit function which can be expressed figuratively as

$$
\begin{equation*}
\left(\tilde{\mathrm{K}}_{n}\right)_{n \in \mathbb{N}} \rightarrow \mathrm{~K}: \Leftrightarrow \lim _{n \rightarrow \infty} d_{Y_{X}^{X}}\left(\tilde{\mathrm{~K}}_{n}, \mathrm{~K}\right)=0 . \tag{3.5}
\end{equation*}
$$

There are various theorems like the Stone-Weierstrass theorem for multivariate polynomials (see, e.g., in [140], [169], [183]) which establish the convergence of different surrogate models in a sense that is similar to (3.5). The expression in (3.4) and the expression in (3.5) guide the definition of the high-fidelity function approximation error $\mathfrak{e}_{\mathrm{H}}(\mathrm{K})$ with respect to a fixed surrogate model $\tilde{\mathrm{K}}_{n}$.

Definition 3.1.1 (High-fidelity function approximation error). Given a high-fidelity model $\mathrm{K}: X \rightarrow Y$ and a fixed surrogate model $\tilde{\mathrm{K}}_{n}: X \rightarrow Y$ from a hypothesis space H , i.e., $\tilde{\mathrm{K}}_{n} \in \mathrm{H}$ and $\mathrm{H} \subseteq Y^{X}$, the high-fidelity function approximation error $\mathfrak{e}_{\mathrm{H}}(\mathrm{K}) \in \mathbb{R}^{+}$ with respect to a fixed surrogate model $\tilde{\mathrm{K}}_{n}$ is constituted by

$$
\begin{equation*}
\mathfrak{e}_{\mathrm{H}}(\mathrm{~K}):=\sup _{x \in \mathrm{X}}\left\|\mathrm{~K}(x)-\tilde{\mathrm{K}}_{n}(x)\right\|_{\mathrm{Y}} . \tag{3.6}
\end{equation*}
$$

Remark 3.1.1. If it is unambiguous, then the adjunct high-fidelity is dropped.
Remark 3.1.2. An important special case of (3.6) is $\mathfrak{e}_{\mathrm{H}}\left(\hat{Q}_{\mathcal{F}}\right)$, that is, the function approximation error with respect to the $l_{p}$-norm regarding the reduced parametric quantity of interest.

Comparing the function approximation error $\mathfrak{e}_{\mathrm{H}}(\mathrm{K})$ in (3.6) to the modeling error $\mathfrak{E}(Q)$ in $(2.22 b)$, it is apparent that one can control $\mathfrak{c}_{\mathrm{H}}(\mathrm{K})$ by a surrogate model's order $n$ where the positive integer $n+1$ can represent, for instance, the number of basis functions. Thus, assuming an $\mathbb{F}$-vector space structure on $Y^{X}$ and on $H$, a possible
general presentation of a surrogate $\tilde{\mathrm{K}}_{n}$ is given by

$$
\begin{equation*}
\tilde{\mathrm{K}}_{n}:=\sum_{i=0}^{n} \tilde{c}^{i} \cdot \mathbb{F} \tilde{\varphi}_{i}, \tag{3.7}
\end{equation*}
$$

where $\tilde{\varphi}_{i}: X \rightarrow Y$ signifies basis functions, i.e., members of a basis of a selected hypothesis space $H$ with dimension $n$, that is, $\operatorname{dim}(H) \equiv n$; and $\tilde{c}^{i} \in \mathbb{F}$ tags coefficients which are also referred to as components or coordinates of $\tilde{\mathrm{K}}_{n}$ with respect to the chosen basis. A prototypical hypothesis space is the space $\mathrm{P}_{\leq n}$, that is, the set of all univariate algebraic polynomials of degree at most $n$ on an interval $[a, b]$ equipped with an $\mathbb{R}$-vector space structure and the finite monomial basis. The degree $n$ provides a notion of a characteristic size of the hypothesis space $H$ (cf. [51, p. 13]). ${ }^{2}$

Supposing a Hilbert space structure on $Y^{X}$ and on $H$, one can define a notion of a best approximation in a least-squares sense. Thus, the best or the closest surrogate model $\hat{\mathrm{K}}_{n} \in \mathrm{H}$ to the high-fidelity model $\mathrm{K} \in Y^{X}$ is associated with the optimization problem

$$
\begin{align*}
\min _{\tilde{\mathrm{K}}_{n} \in \mathrm{H}} . \mathrm{R}\left(\tilde{\mathrm{~K}}_{n}\right): & :=\frac{1}{2}\left\|\mathrm{~K}-\tilde{\mathrm{K}}_{n}\right\|_{Y^{X}}^{2}  \tag{3.8a}\\
& \equiv \frac{1}{2}\left\langle\mathrm{~K}-\tilde{\mathrm{K}}_{n}, \mathrm{~K}-\tilde{\mathrm{K}}_{n}\right\rangle_{Y^{X}} \tag{3.8b}
\end{align*}
$$

where $R: H \rightarrow \mathbb{R}^{+}$denotes the residual objective functional. One can characterize the best surrogate model $\hat{\mathrm{K}}_{n}$ by the minimizer functional argmin $=\mathrm{R} \mapsto \hat{\tilde{\mathrm{K}}}_{n}:\left(\mathrm{H} \rightarrow \mathbb{R}^{+}\right) \rightarrow \mathrm{H}$ as a single-valued functional. If the residual objective functional $R$ possesses a unique global minimizer, then the minimizer functional returns $\hat{\mathrm{K}}_{n}$ as an output, i.e.,

$$
\begin{equation*}
\hat{\tilde{\mathrm{K}}}_{n}=\operatorname{argmin}(\mathrm{R}) \tag{3.9}
\end{equation*}
$$

is well-defined. If there are multiple global minimzers or if there is no global minimizer at all, the minimizer functional falls back to the common definition as a multivalued functional where the output is constituted by the set of global minimizers of R, i.e.,

$$
\begin{equation*}
\operatorname{argmin}(\mathrm{R}) \equiv \underset{\tilde{\mathrm{K}}_{n} \in \mathrm{H}}{\arg \min } \mathrm{R}\left(\tilde{\mathrm{~K}}_{n}\right):=\left\{\tilde{\mathrm{K}}_{n} \in \mathrm{H} \mid \mathrm{R}\left(\tilde{\mathrm{~K}}_{n}\right)=\inf _{\tilde{\mathrm{K}}_{n}^{\prime} \in \mathrm{H}} \mathrm{R}\left(\tilde{\mathrm{~K}}_{n}^{\prime}\right)\right\} \tag{3.10}
\end{equation*}
$$

In the case of (3.10), the surrogate model $\hat{\tilde{\mathrm{K}}}_{n}$ is $a$ best solution that reads as

$$
\begin{equation*}
\hat{\tilde{\mathrm{K}}}_{n} \in \underset{\tilde{\mathrm{~K}}_{n} \in \mathrm{H}}{\arg \min } \mathrm{R}\left(\tilde{\mathrm{~K}}_{n}\right) . \tag{3.11}
\end{equation*}
$$

In the present work, it is generally assumed that the expression (3.9) holds. This assumption is particularly reasonable for a least squares approach such as in (3.8a) (see, e.g., [201, p. 69ff]) where the best (or the closest) surrogate model $\hat{\mathrm{K}}_{n}$ is the orthogonal (pseudo-) projection of the high-fidelity model K onto the hypothesis

[^16]space $H$ such that
\[

$$
\begin{equation*}
\left(\mathrm{K}-\hat{\mathrm{K}}_{n}\right) \perp \mathrm{H} . \tag{3.12}
\end{equation*}
$$

\]

If we consider, e.g., the prototypical hypothesis space $\mathrm{P}_{\leq n}$ in the context of the space of all square-integrable functions $L^{2}$, then we have the basic continuous least squares $L^{2}$ approximation as an instance of (3.8a). ${ }^{3}$

## Sampling plan

If one would possess sufficient information in order to determine completely the high-fidelity model, then the previous considerations suffice for the discussion of corresponding surrogates. A standard example is the approximation of special functions such as the sine function. However, a basic assumption in the present work states that a single evaluation of a high-fidelity model is costly; hence, it is desired to keep the total number of evaluations as low as possible. Therefore, let us create a sample $\mathbf{s} \in W^{m}$ such that

$$
\begin{equation*}
\mathbf{s}:=\left(\left(x_{1}, y_{1}\right), \ldots,\left(x_{m}, y_{m}\right)\right) \in(X \times \underbrace{Y) \times \cdots \times(X}_{m} \times Y), \tag{3.13}
\end{equation*}
$$

where $m \in \mathbb{N} \backslash\{0\}$ and $\forall i \in\{1, \ldots, m-1, m\} . y_{i}=\mathrm{K}\left(x_{i}\right)$ and, furthermore, it is set that $W^{m}:=\left\{\left(s_{i}\right)_{i \in\{1, \ldots, m-1, m\}} \mid \forall i \in\{1, \ldots, m-1, m\} . s_{i} \in X \times Y\right\}$. Let us refer to $m$ as sample size. Mind that, in theoretical considerations, it is assumed that the pairs in $X \times Y$ are independently randomly chosen which involves utilizing means from the probability theory toolkit such as a probability measure (see, e.g., [51, p. 5]). For a more comprehensive discussion on the toolkit concerning measure and probability theory, see, e.g., [201, Ch. 2].

However, let us forgo using all of the corresponding theoretical toolkit: In the present work, we rather focus on the entities with respect to a sample, which are equipped with the attribute sample or empirical - such as the sample or empirical mean, the sample or empirical variance, and similar.

Given a sample s and given $X^{m}$ as the $m$-fold Cartesian product of $X$ with $X \subset$ $\mathbb{R}^{N_{\bar{\zeta}}}$, it is more common to deploy the notion of a sampling plan $X_{s} \subseteq X^{m}$ defined as

$$
\begin{equation*}
X_{\mathrm{s}}:=\left\{\left(x_{i}\right)_{i \in\{1, \ldots, m\}} \mid \forall i \in\{1, \ldots, m\} \cdot x_{i} \in X\right\}, \tag{3.14}
\end{equation*}
$$

where, concerning the implementation and the choice of a data structure, it is prevailing to identify the sampling plan $X_{\mathbf{s}}$ with an $m \times N_{\xi}$ matrix where $N_{\xi}$ denotes the number of parameters (see § 2.2.3).

Given a member $x \in X_{\mathrm{s}}$ and invoking the projection maps $\pi_{i}: X_{\mathbf{s}} \rightarrow X_{\mathrm{s}, i}$ with $i \in\{1, \ldots, m\}$ and $X_{\mathrm{S}, i} \equiv X$ where $\pi_{i}=x \mapsto \pi_{i}(x)=: x_{i}$, let us refer to $x_{i}$ as sampling plan points and to the corresponding $y_{i}$ in (3.14) as output points.

Let us discuss briefly some peculiarities regarding the design of a sampling plan. For a more elaborate discussion on the design of sampling plans, I refer to, e.g., [116, ch. 13], [70, ch. 1], [53, ch. 17], or [61, ch. 2].

Two desirable properties of a sampling plan $X_{s}$ are that it is:

[^17]- space-filling and
- non-collapsing (see, e.g., [100]).

The non-collapsing property requires that the coordinates of the sampling plan points are not identical. More precisely: Let $i \in\{1, \ldots, m\}$ be fixed, and let $\pi_{j}$ denote the coordinate projection maps such that $\pi_{j}: X_{\mathbf{s}, i} \rightarrow \mathbb{R}$ with $j \in\left\{1, \ldots, N_{\xi}\right\}$ where $\pi_{j}=$ $x_{i} \mapsto \pi_{j}\left(x_{i}\right)$, then $\forall x_{k} \in X_{\mathbf{s}, k}, x_{l} \in X_{\mathbf{s}, l} \cdot \pi_{j}\left(x_{k}\right) \neq \pi_{j}\left(x_{l}\right)$ where $k \in\{1, \ldots, m\}$ and $l \in\{1, \ldots, m\}$ and $k \neq l$. The rationale for this property is, in a strong sense, to exclude the pathological case where there are two identical sampling plan points; and, in a weak sense, to exclude the non-economical cases where two sampling plan points differ only in coordinates to which the high fidelity model is not very sensitive anyway, so that, in fact, the two points can be seen as equal.

The space-filling property requires to sample the domain of the high-fidelity model in such a way that the sampling plan error $\mathfrak{e}\left(X_{\mathbf{s}}\right) \in \mathbb{R}^{+}$is minimal which results in a maximal uniform scattering of the sampling plan points in the domain. Notice well that there are many ways to quantify the space-filling property (see, e.g., [53, p. 600]). In order to achieve an optimized sampling plan, a basic idea is to minimize some objective function involving a distance measure of the sampling plan points with respect to the $l_{p}$-norm. Pursuing this idea, the corresponding space-filling sampling plans are generally called Latin hypercube (LHC) sampling plans. Another kind of space-filling sampling plans are quasi-random sequences or low-discrepancy sequences ([53, p. 615ff]). They are discussed, for instance, in the context of the numerical integration of multivariate functions (cf. [116, p. 245]).

In [70, p. 17-27], the authors provide an implementation in the MATLAB ${ }^{\circledR}$ PL for creating an optimized LHC sampling plan based on the Morris-Mitchell criterion and an evolutionary operation. By exploiting the package MATLAB.jl (see [101]) which provides the capability to interact with the MATLAB ${ }^{\circledR}$ PL within the Julia PL, let us adapt the lines of code concerning this particular optimized LHC sampling plan to the Julia PL and label them (XSpkg1).

Additionally, let us invoke two Julia PL packages for the creation of sampling plans:
(XSpkg2) the package LatinHypercubeSampling.jl (see [215] and [216]) provides an implementation for creating an optimized LHC sampling plan based on the Audze-Eglais criterion and a genetic algorithm to solve the corresponding optimization problem,
(XSpkg3) the package Sobol.jl (see [106]) provides an implementation for creating a Sobol quasi-random sequence.

In Figure 3.1 and in Figure 3.2, there are representations of different sampling plans $X_{\mathbf{s}} \subseteq X^{m}$ where $X:=[0,1]^{2}$ denotes the unit 2-dimensional hypercube and the number of sampling plan points $m$ is given by $m \in\{10,25,50,100\}$ as well as by $m \in$ $\{10,25,100,1000\}$. Using (XSpkg1), an optimized Latin hypercube sampling plan is generated which is abbreviated to maximin LHC (cf. [100]). Using (XSpkg2), a random Latin hypercube and an optimized Latin hypercube sampling plan are generated which are abbreviated to Audze-Eglais LHC (cf. [100]), respectively. The unit hypercube is achieved by scaling the hypercube $[1, m] \times[1, m]$. Using (XSpkg3), a Sobol quasi-random sequence sampling plan is generated which is by default constructed for the unit hypercube. Other hypercubes can be achieved by scaling the unit hypercube.

Comparing the LHC sampling plans (see Figures 3.1a-3.1d)), one can observe that, already at a low number of sampling plan points, utilizing a random LHC can lead to a clustering instead of a uniform spreading. The comparison of the AudzeEglais LHC and the maximin LHC is intricate due to the different underlying criteria and the randomness in the corresponding stochastic optimization algorithms (see § 2.3.3). However, both optimized Latin hypercubes exhibit a highly uniform scattering of the sampling plan points as desired. For more details on a comparison of the Audze-Eglais LHC and the maximin LHC, see, e.g., (cf. [100]).

In Figure 3.2, the Sobol quasi-random sequence sampling plan is investigated which shows a highly uniform and nonrandom scattering of the sampling plan points. ${ }^{4}$

Interestingly, using (XSpkg1) for high numbers of sampling plan points such as $m>100$ and a high accuracy regarding the solving of the underlying optimization problem requires a significantly higher computational time than using (XSpkg2) and (XSpkg3).

From a modeling and simulation viewpoint, though, the fundamental premise is that a single evaluation of a high-fidelity model is expensive. Therefore, the aim is to construct a sufficiently space-filling sampling plan in a short amount of time at a low number of sampling plan points. With this aim in mind, all three presented space-filling sampling plans are well-suited.

From an optimization viewpoint, the sampling plan points should ideally be located in the vicinity of optimal points. In Figure 3.3, I illustrate this requirement by adapting the Audze-Eglais Latin hypercube sampling plan in Figure 3.2 for the contour representation in Figure 2.2b.

Notice well that a high number of sampling plan points can lower the sampling plan error $\mathfrak{e}\left(X_{\mathbf{s}}\right)$. Lowering this error can improve the local accuracy of a surrogate model built upon these points. A surrogate model's global accuracy, though, is also determined by the function approximation error in (3.6) which is independent of the sample.

## Empirical surrogate modeling error

In context of the local and global accuracy of a surrogate model, let us introduce another notion regarding the entity $\hat{Q}_{\overparen{\xi}}$ : the empirical surrogate modeling error $\mathfrak{e}_{\mathrm{H}, \mathrm{s}}\left(\hat{Q}_{\tilde{\xi}}\right)$ with respect to the sampling plan $X_{s}$.

Definition 3.1.2 (Empirical surrogate modeling error). Let us suppose a sampling plan $X_{\mathbf{s}}$ (such that $\mathfrak{e}\left(X_{\mathrm{s}}\right)$ is minimal) equipped with sampling plan points $x_{i} \in X_{\mathrm{s}, i}$ where $i \in\{1, \ldots, m\}$ and $X_{\mathrm{s}, i} \equiv X$. Furthermore, let us assume a high-fidelity model $\hat{Q}_{\xi}: X \rightarrow \mathbb{R}$ and a fixed surrogate model $\tilde{\hat{Q}}_{\xi, n}: X \rightarrow \mathbb{R}$ from a hypothesis space $\mathrm{H} \subseteq$ $\mathbb{R}^{X}$. Then, the empirical surrogate modeling error $\mathfrak{e}_{\mathrm{H}, \mathrm{s}}\left(\widehat{Q}_{\tilde{\xi}}\right) \in \mathbb{R}^{+}$with respect to the sampling plan $X_{\mathrm{s}}$ is constituted by

$$
\begin{equation*}
\mathfrak{e}_{\mathrm{H}, \mathrm{~s}}\left(\hat{Q}_{\tilde{\zeta}}\right):=\frac{1}{m} \sum_{i=1}^{m}\left(\hat{Q}_{\tilde{\zeta}}\left(x_{i}\right)-\tilde{\hat{Q}}_{\tilde{\xi}, n}\left(x_{i}\right)\right)^{2} .{ }^{5} \tag{3.15}
\end{equation*}
$$

[^18]
(A) The number of sampling plan points is given by $m:=10$.

(B) The number of sampling plan points is given by $m:=25$.

(C) The number of sampling plan points is given by $m:=50$.

(D) The number of sampling plan points is given by $m:=100$.

FIGURE 3.1: Representations of different sampling plans $X_{\mathbf{s}} \subseteq X^{m}$ where $X:=[0,1]^{2}$ denotes the unit 2-dimensional hypercube.
(i) Random LHC, (ii) Audze-Eglais LHC,
(iii) Maximin LHC.

(A) The number of sampling plan points is given by $m:=10$.

(B) The number of sampling plan points is given by $m:=50$.

(C) The number of sampling plan points is given by $m:=100$.

(D) The number of sampling plan points is given by $m:=1000$.

FIGURE 3.2: Representations of different sampling plans $X_{\mathbf{s}} \subseteq X^{m}$ where $X:=[0,1]^{2}$ denotes the unit 2-dimensional hypercube.
(i) Random LHC, (ii) Audze-Eglais LHC,
(iii) Sobol quasi-random sequence.

(A) The number of sampling plan points is given by $m:=10$.

(в) The number of sampling plan points is given by $m:=50$.

Figure 3.3: The Audze-Eglais LHC sampling plan in Figure 3.2 adapted for the contour representation in Figure 2.2b.

Remark 3.1.3. I point out that the size $m$ of the sample $\mathbf{s}$ in (3.15) is fixed. If $m$ would theoretically tend to infinity, then this asymptotic consideration would particularly affect the sampling plan error $\mathfrak{e}\left(X_{\mathbf{s}}\right)$. The function approximation error $\mathfrak{e}_{\mathrm{H}}\left(\hat{Q}_{\mathcal{G}}\right)$ in (3.6), though, would not be affected since it is independent of the sample.

Remark 3.1.4. If we recall the presentation of a surrogate in (3.7), then, technically, we face a family of surrogates parameterized by the coefficients $\tilde{c}^{i}$ and by the degree $n$. One can distinguish the kinds of parameters by utilizing the term hyperparameter. Let us comprehend all parameters to be determined as hyperparameters $\tilde{\chi}_{i}-$ except the coefficients $\tilde{c}^{i}$; one can organize these hyperparameters within an ordered set whose size depends on the given surrogate modeling problem.

Remark 3.1.5. Notice well that the presence of the coefficients and hyperparameters pressures us to translate our notation from $\tilde{\hat{Q}}_{\tilde{\xi}, n}\left(x_{i}\right)$ to, for instance, $\tilde{\hat{Q}}_{\tilde{\xi}}\left(x_{i_{1}}, c^{i_{2}}, \tilde{\chi}_{i_{3}}\right)$. Though, if the peril of confusion is low, in order to be consistent with the common literature, let us accept an abuse of notation, that is, let us put $\tilde{\hat{Q}}_{\mathcal{Z}, n}\left(x_{i}\right)$ to work which refers tacitly to, for instance, $\tilde{Q}_{\tilde{\xi}}\left(x_{i_{1}}, c^{i_{2}}, \tilde{\chi}_{i_{3}}\right)$.

Defining the error $\mathfrak{e}_{\mathrm{H}, \mathrm{s}}\left(\hat{Q}_{\tilde{\xi}}\right)$ has a conceptual and a practical value. At a conceptual level, it expresses the problem-dependency on $\hat{Q}_{\tilde{\xi}}$. Furthermore, this error encodes the dependency on the surrogate model's membership to a prescribed class of functions, that is, the hypothesis space $H$; and it encodes the dependency on the sample s .

At a practical level, the error $\mathfrak{e}_{\mathrm{H}, \mathrm{s}}\left(\widehat{Q}_{\xi}\right)$ serves as a starting point in order to define the empirical training error $\mathfrak{c}_{\mathrm{H}, s_{t}}\left(\hat{Q}_{\xi}\right)$ and, more importantly, the empirical generalization error $\mathfrak{e}_{\mathrm{H}, \mathrm{s}_{g}}\left(\hat{Q}_{\mathfrak{\xi}}\right)$. These errors require a partition $X_{\mathrm{s}, p}$ of the sampling plan $X_{\mathbf{s}}$, more precisely, $X_{s}$ is represented as the disjoint union of two subsets $X_{\mathrm{s}_{t}}$ and $X_{\mathrm{s}_{g}}$ - where $X_{s_{t}}$ denotes the training subset (or observed points subset) and $X_{s_{g}}$ denotes the testing subset (or prediction points subset). More formally, there exists an equivalence relation $\sim \chi_{X_{s, p}}$ on $X_{\mathrm{s}}$ associated with the given partition such that $X_{\mathrm{s}, p}:=X_{\mathrm{s}} / \sim \chi_{\mathrm{s}, p}$; and by demanding $X_{s_{t}} \cap X_{s_{g}}=\varnothing$, it is set that

$$
\begin{equation*}
X_{\mathrm{s}}:=X_{\mathbf{s}_{t}} \cup X_{\mathbf{s}_{g}}, \tag{3.16}
\end{equation*}
$$

where $X_{\mathbf{s}_{t}} \in X_{\mathbf{s}, p}$ and $X_{\mathbf{s}_{g}} \in X_{\mathbf{s}, p}$. Given the number of sampling plan points $m$, then the positive integer $m_{t}$ denotes the number of training points and the positive integer $m_{g}$ denotes the number of testing points such that $m \equiv m_{t}+m_{g}$ whereas mostly $m_{g} \ll m_{t}$. Moreover, given $m$ and a scalar $p_{m} \in(0,1)$ that encodes a fixed partition ratio for the sampling plan, then it is set that

$$
\begin{equation*}
m_{t}:=\lceil p \cdot m\rceil \text { such that } m_{g}:=m-\lceil p \cdot m\rceil . \tag{3.17}
\end{equation*}
$$

Let us suppose a scarcity of points in the sampling plan $X_{s}$. Hence, we do not define yet another subset, more precisely, we do not define a validating subset (see, e.g., [91, p. 222f]). A validating subset's purpose is to aid selecting a member from a parameterized family of surrogates (recall Remark 3.1.4) adapted to the training subset; before this preselected member is assessed by the testing subset.

The members of the training subset $X_{s_{t}}$ are deployed in (3.15) to determine or to estimate the parameters $\tilde{g} \in G$ such as the coefficients $\tilde{c}^{i}$ in (3.7) or the hyperparameters $\tilde{\chi}_{i}$ either by an interpolation problem (exactly fitting the given data) or by a regression problem (inexactly fitting the given data).

Determining the coefficients by interpolation results in an empirical training error $\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{t}}\left(\hat{Q}_{\mathcal{\xi}}\right)$ of zero - by virtue of the interpolation property: Supposing an evaluation functional $\operatorname{ev}_{i} \in \operatorname{hom}_{\mathbb{R}}\left(\mathbb{R}^{X}, \mathbb{R}\right)$ such that $\operatorname{ev}_{i}\left(\tilde{Q}_{\tilde{\xi}, n}\right):=\tilde{\hat{Q}}_{\mathcal{\xi}, n}\left(x_{i}\right)$, let us conceive the interpolation property in the sense of

$$
\begin{equation*}
\forall x_{i} \in X_{\mathbf{s}, i} \cdot \hat{Q}_{\tilde{\xi}}\left(x_{i}\right)=\tilde{\hat{Q}}_{\mathcal{\xi}, n}\left(x_{i}, \tilde{g}\right) \tag{3.18}
\end{equation*}
$$

In the next section, we discuss the radial basis surrogate model and the Gaussian process regression (or kriging) surrogate model, respectively. In this discussion, we encounter concisely the interpolation problem in a deterministic setting and in a stochastic setting, respectively. For an in-depth elaboration about deterministic and stochastic interpolation, I refer to, e.g., [201, ch. 13] or [188].

Recalling (2.35) and (3.8a), one can define an objective function $r: G \rightarrow \mathbb{R}$ in order to estimate the best parameters $\hat{\tilde{g}} \in G$ via a regression problem in the sense of a basic discrete least squares $l_{2}$ approximation

$$
\begin{equation*}
\underset{\tilde{g} \in G}{\operatorname{minimize}} \operatorname{r}(\tilde{g}):=\frac{1}{2} \sum_{i=1}^{m}\left\|\hat{Q}_{\tilde{\xi}}\left(x_{i}\right)-\tilde{Q}_{\tilde{\xi}}\left(x_{i}, \tilde{g}\right)\right\|_{l_{2}}^{2} \tag{3.19}
\end{equation*}
$$

If we estimate the coefficients by regression such as a least squares method, then, generically, the training error $\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{t}}\left(\widehat{Q}_{\boldsymbol{\xi}}\right)$ is greater than zero. A possible interpretation of this approach to parameter finding is the following: We know a priori that the sample $\mathbf{s}$ in (3.13) fits the surrogate model but the observed outputs within the sample are noisy, that is,

$$
\begin{equation*}
\forall i \in\{1, \ldots, m\} . y_{i}=\mathrm{K}\left(x_{i}\right)+\varepsilon_{i} \tag{3.20}
\end{equation*}
$$

where $\varepsilon_{i}$ indicate members of a vector $\varepsilon \in \mathbb{R}^{m}$ which are independent random numbers distributed with regard to the normal distribution with mean $\mu \equiv 0$ and constant variance $\sigma^{2}$, i.e., $\varepsilon \sim \mathcal{N}\left(0, \sigma^{2}\right) .{ }^{6}$ For ordinary least squares and a fixed standard deviation $\sigma>0$, this interpretation can be embedded in the more general parameter finding by maximum likelihood estimation (see, e.g., [91, p. 31], [155, p. 217ff] or [201, p. 96]). We look at the maximum likelihood estimation in the next section with regard to the Gaussian process regression (or kriging) low-fidelity model.

The training error $\mathfrak{e}_{\mathrm{H}, \boldsymbol{s}_{t}}\left(\hat{Q}_{\xi}\right)$ associated with a surrogate model $\tilde{\hat{Q}}_{\mathcal{\xi}, n}$ is inadequate to assess the surrogate model's predictive power concerning points not yet observed (see, e.g., [173, p. 108] or [91, p. 221]). Hence, derived from the empirical surrogate modeling error in (3.15), let us introduce the empirical generalization error $\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{g}}\left(\widehat{Q}_{\mathcal{\xi}}\right) \in$ $\mathbb{R}^{+}$with respect to the testing subset $X_{\mathbf{s}_{g}}$ such that

$$
\begin{equation*}
\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{g}}\left(\hat{Q}_{\xi}\right):=\frac{1}{m_{g}} \sum_{i=1}^{m_{g}}\left(\hat{Q}_{\tilde{\xi}}\left(x_{i}\right)-\tilde{Q}_{\xi, n}\left(x_{i}\right)\right)^{2}, \tag{3.21}
\end{equation*}
$$

where $\forall i \in\left\{1, \ldots, m_{g}\right\} . x_{i} \in X_{\mathbf{s}_{g}, i}:=\left(X_{\mathbf{s}} \backslash X_{\mathbf{s}_{t}}\right)_{i}$ and $\left(X_{\mathbf{s}} \backslash X_{\mathbf{s}_{t}}\right)_{i} \equiv X$. Often, it is convenient to normalize the error $\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{g}}\left(\hat{Q}_{\xi}\right)$ to the interval $\left[\hat{Q}_{\xi}^{\min }, \hat{Q}_{\xi}^{\max }\right]_{X_{s_{g}, i}}$, where $\hat{Q}_{\xi}^{\min } \in \mathbb{R}$

[^19]denotes the minimal output point with respect to $X_{s_{8}, i}$ and $\hat{Q}_{\xi}^{\max } \in \mathbb{R}$ denotes the maximal output point with respect to $X_{\mathrm{s}_{g}, i}$. Hence, one can define the normalized empirical generalization error (NEGE) $\mathfrak{c}_{\mathrm{H}, \mathrm{s}_{\xi}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right) \in \mathbb{R}^{+}$with respect to the testing subset $X_{\mathrm{s}_{g}}$ such that
\[

$$
\begin{equation*}
\mathfrak{e}_{\mathrm{H}, \boldsymbol{s}_{\xi}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right):=\frac{1}{\hat{\mathrm{Q}}_{\tilde{\xi}}^{\max }-\hat{Q}_{\tilde{\xi}}^{\min }} \mathfrak{e}_{\mathrm{H}, \mathbf{s}_{\boldsymbol{g}}}\left(\hat{Q}_{\tilde{\xi}}\right) . \tag{3.22}
\end{equation*}
$$

\]

As it has been pointed out in the commentary on the empirical surrogate modeling error in (3.15), one can additionally introduce the root empirical generalization error $\mathfrak{e}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{R}}\left(\hat{Q}_{\tilde{\xi}}\right) \in \mathbb{R}^{+}$such that

$$
\begin{equation*}
\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{R}}\left(\hat{Q}_{\tilde{\xi}}\right):=\left[\mathfrak{c}_{\mathrm{H}, \mathbf{s}_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)\right]^{\frac{1}{2}} \tag{3.23}
\end{equation*}
$$

and one can introduce the normalized root empirical generalization error (NREGE) $\mathfrak{e}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{NR}}\left(\widehat{Q}_{\tilde{\xi}}\right) \in \mathbb{R}^{+}$that reads as

$$
\begin{equation*}
\mathfrak{e}_{\mathrm{H}, \mathrm{~s}_{\mathcal{F}}}^{\mathrm{NR}}\left(\hat{Q}_{\tilde{\zeta}}\right):=\frac{1}{\hat{Q}_{\tilde{\zeta}}^{\max }-\hat{Q}_{\tilde{\zeta}}^{\min }} \mathfrak{e}_{\mathrm{H}, \mathrm{~s}_{\xi}}^{\mathrm{R}}\left(\hat{Q}_{\tilde{\zeta}}\right) . \tag{3.24}
\end{equation*}
$$

Note that both the error $\varepsilon_{H, s_{g}}^{R}\left(\hat{Q}_{\xi}\right)$ and the error $\varepsilon_{H, s_{g}}^{\mathrm{NR}}\left(\hat{Q}_{\xi}\right)$ can be seen as more conservative error measures than their counterparts in (3.21) and in (3.22), respectively.

Let us call a partition $X_{\mathrm{s}, p}$ - with, e.g., the cells $X_{\mathrm{s}_{t}}$ and $X_{\mathrm{s}_{g}}$ - randomly created if and only if the members of the sampling plan $X_{s}$ are permuted randomly and assigned to the partition's cells in accordance with the partition ratio $p_{m}$. Then, given a fixed number of sampling plan points $m$, a sampling plan point's membership to either the training subset or to the testing subset in (3.16) is random, i.e., the partition $X_{\mathrm{s}, p}$ is randomly created.

In order to average over this membership randomness, let us utilize the notion of a mean generalization error $\overline{\mathcal{E}}_{\mathrm{H}, \mathrm{s}_{g}}\left(\hat{Q}_{\xi}\right) \in \mathbb{R}^{+}$which we compute by deploying the hold-out with random sub-sampling method or the $k$-fold cross-validation method. These methods are computationally tractable since they are non-exhaustive in the sense that they do not consider all possible ways of partitioning the sampling plan. For further elaboration on the technical intricacies, see, e.g. the survey in [5].

The basic version of the hold-out (or simple validation) method computes the generalization error in (3.21) by assuming a fixed partition ratio $p_{m}$ and a randomly created partition $X_{s, p}$ that is constituted of the two cells $X_{s_{t}}$ and $X_{\mathrm{s}_{g}}$. An extended version of the hold-out method includes random sub-sampling (cf. [116, p. 267f]), viz. performing the basic hold-out method in a finite number of multiple independent runs where at each run the generalization error in (3.21) is computed; finally, the mean generalization error $\overline{\mathfrak{c}}_{\mathrm{H}, \mathbf{s}_{\mathcal{g}}}\left(\widehat{Q}_{\mathcal{\xi}}\right)$ is computed as the mean of all individual generalization errors. However, due to the random creation of the individual partitions, this method does not provide any guarantees that all sampling plan points will be exploited properly as testing points.

The $k$-fold cross-validation method supposes a randomly created partition $X_{\mathrm{s}, p}$ that is constituted of $k$ cells $X_{\mathrm{s}, p}^{(1)}, \ldots, X_{\mathrm{s}, p}^{(k-1)}, X_{\mathrm{s}, p}^{(k)}$ where the positive integer $k$ is selected as $k \ll m$. Either the cells are equal in size, i.e., given a positive integer $q$, then $\forall i \in\{1, \ldots, k\} .\left|X_{s, p}^{(i)}\right|=q$; or the cells are only approximately equal in size. Generally, there are numerous options to construct the required training and testing subsets. However, the construction principle underlying this method demands to define $k-1$
cells as the training subset and to define the $k$-th cell as the testing subset. Since the cells' ordering is not preserved when assigned to the subsets, there are $k$ different options to define the corresponding subsets. Therefore, one can define the $i$-th training subset and the $i$-th testing subset as

$$
\begin{align*}
& X_{\mathbf{s}_{t}}^{(i)}:=X_{\mathbf{s}, p} \backslash X_{\mathbf{s}, p}^{(i)},  \tag{3.25}\\
& X_{\mathbf{s}_{g}}^{(i)}:=X_{\mathbf{s}, p}^{(i)}, \tag{3.26}
\end{align*}
$$

where $i \in\{1, \ldots, k\}$. For each of the $k$ options, the generalization error in (3.21) is computed. Hence, the mean generalization error $\bar{\varepsilon}_{\mathrm{H}, s_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)$ is computed as the mean of all the $k$ individual generalization errors. In order to emphasize the dependency of the generalization error on the $k$-fold cross validation method, let us introduce the map $\mathfrak{e}_{\mathrm{H}, \mathrm{s}_{g}, \mathrm{cv}}$ that reads as

$$
\begin{equation*}
\mathfrak{e}_{\mathrm{H}, \mathrm{~s}_{8}, \mathrm{cv}}=\left.k \mapsto \mathfrak{e}_{\mathrm{H}, \mathbf{s}_{g}}\left(\hat{Q}_{\xi}\right)\right|_{k}:=\mathfrak{e}_{\mathrm{H}, \mathrm{~s}_{g}, \mathrm{cv}}(k): \mathbb{Z}^{+} \rightarrow \mathbb{R}^{+}, \tag{3.27}
\end{equation*}
$$

where $\left.\mathfrak{e}_{\mathrm{H}, \mathrm{s}_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)\right|_{k}$ and $\mathfrak{e}_{\mathrm{H}, \mathrm{s}_{g}, \mathrm{cv}}(k)$, respectively, denote the $k$-dependent generalization error. Hence, the mean $k$-dependent generalization error is denoted as $\left.\overline{\mathfrak{c}}_{\mathrm{H}, \mathrm{s}_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)\right|_{k}$ and $\overline{\mathcal{E}}_{\mathrm{H}, \mathrm{s}_{3}, \mathrm{cv}}(k)$, respectively.

In the context of (3.22), let us ease the notation for the sake of conciseness, that is, let us introduce the map $\mathrm{e}_{\mathrm{cv}}^{\mathrm{N}}$ that reads as

$$
\begin{equation*}
e_{\mathrm{cv}}^{\mathrm{N}}=\left.k \mapsto \mathfrak{e}_{\mathrm{H}, \mathrm{~s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)\right|_{k} \equiv \mathrm{e}_{\mathrm{cv}}^{\mathrm{N}}(k): \mathbb{Z}^{+} \rightarrow \mathbb{R}^{+}, \tag{3.28}
\end{equation*}
$$

where $\left.\mathfrak{e}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\zeta}}\right)\right|_{k}$ and $\mathrm{e}_{\mathrm{cv}}^{\mathrm{N}}(k)$, respectively, denote the $k$-dependent normalized generalization error. Hence, the mean $k$-dependent normalized generalization error is denoted as $\left.\overline{\bar{\varepsilon}}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right|_{k}$ and $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}(k)$, respectively. Notice well that one can analogously define the mean $k$-dependent normalized root generalization error $\left.\bar{\varepsilon}_{H, s_{\mathcal{g}}}^{N R}\left(\hat{Q}_{\mathcal{\xi}}\right)\right|_{k}$ and $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{NR}}(k)$, respectively.

In both the extended hold-out method and the $k$-fold cross-validation method the computational burden is dominated by determining the surrogate model via the training subset for the computation of the individual generalization errors. Hence, the number of runs (in the hold-out method) and the number of folds (in the crossvalidation method), respectively, have to be chosen in such a way that the burden is low while still producing a reliable mean generalization error. Let us choose the number of runs similar to the number of folds for which a computational reasonable choice is $k \equiv 5$ or $k \equiv 10$ (see, e.g., [91, p. 242ff]). ${ }^{7}$

These computational considerations are relevant for both the surrogate model assessment and the surrogate model selection - as already mentioned above regarding the non-utilization of a validating subset - which can be subsumed under the bias-variance problem (see, e.g., [51, p. 13f], [91, p. 223ff] or [155, p. 202]). It is a nontrivial task to find the optimal hyperparameters in the sense that there is an adequate tradeoff between the need for a small bias to avoid underfitting and the need for a small variance to avoid overfitting the points regarding the sample $\mathbf{s}$ in (3.15) where the size $m$ is fixed. Since there is a lack of a rigorously proven and computationally cost-efficient approach to finding the optimal hyperparameters, a common practice to emulate a validating subset's purpose is to specify some hyperparameters

[^20]by the user and to estimate the remaining hyperparameters by, e.g., cross-validation or maximum-likelihood (see, e.g., [69]).

Supplementary to the error in (3.21), some authors (see, e.g., [70, p. 37]) suggest to compute the squared sample Pearson correlation coefficient (SSPCC) $r_{\hat{y} \tilde{y}}^{2}$ with respect to the testing subset $X_{\mathrm{s}_{g}}$ where $r_{\hat{y} \hat{y}} \in[-1,1]$. For the sake of lucidity, let us apply partly the identifications

$$
\begin{align*}
& \hat{Y}:=\hat{Q}_{\tilde{\xi}}\left(x_{i}\right)  \tag{3.29a}\\
& \hat{\tilde{Y}}:=\hat{\hat{Q}}_{\xi, n}\left(x_{i}\right) . \tag{3.29b}
\end{align*}
$$

Since we consider solely the discrete setting, one can additionally set the sample means

$$
\begin{align*}
& \overline{\hat{Y}}:=\frac{1}{m_{g}} \sum_{i=1}^{m_{g}} \hat{Y}_{i}  \tag{3.30a}\\
& \overline{\hat{Y}}:=\frac{1}{m_{g}} \sum_{i=1}^{m_{g}} \tilde{Y}_{i} \tag{3.30b}
\end{align*}
$$

where the identifications $\hat{Y}_{i} \equiv \hat{Y}$ and $\tilde{Y}_{i} \equiv \tilde{Y}$ are invoked. Moreover, one can overload the meaning of the covariance map cov and the variance map var regarding random variables in a continuous setting. Hence, the coefficient $r_{\hat{y} \tilde{y}}^{2}$ reads as

$$
\begin{align*}
& r_{\hat{y} \hat{y}}^{2}:=\left(\frac{\operatorname{cov}(\hat{Y}, \tilde{\hat{Y}})}{\sqrt{\operatorname{cov}(\hat{Y}, \hat{Y}) \operatorname{cov}(\tilde{\hat{Y}}, \tilde{\hat{Y}})}}\right)^{2}  \tag{3.31a}\\
& \equiv\left(\frac{\operatorname{cov}(\hat{Y}, \tilde{\hat{Y}})}{\sqrt{\operatorname{var}(\hat{Y}) \operatorname{var}(\tilde{\hat{Y}})}}\right)^{2}  \tag{3.31b}\\
& \equiv\left(\frac{\frac{1}{m_{g}-1} \sum(\hat{Y}-\hat{Y})(\tilde{\hat{Y}}-\overline{\hat{Y}})}{\sqrt{\frac{1}{m_{g}-1} \sum(\hat{Y}-\hat{Y})^{2} \frac{1}{m_{g}-1} \sum(\tilde{\hat{Y}}-\overline{\hat{Y}})^{2}}}\right)^{2}  \tag{3.31c}\\
& \equiv\left(\frac{\sum(\hat{Y}-\bar{Y})(\tilde{\hat{Y}}-\overline{\tilde{Y}})}{\sqrt{\sum(\hat{Y}-\hat{Y})^{2} \sum(\tilde{\hat{Y}}-\overline{\tilde{Y}})^{2}}}\right)^{2}  \tag{3.31d}\\
& \equiv\left(\frac{m_{g} \sum \hat{Y} \tilde{\hat{Y}}-\sum \hat{Y} \sum \tilde{Y}}{\sqrt{\left(m_{g} \sum \hat{Y}^{2}-\left(\sum \hat{Y}\right)^{2}\right)\left(m_{g} \sum \tilde{Y}^{2}-\left(\sum \hat{Y}\right)^{2}\right)}}\right)^{2}  \tag{3.31e}\\
& \stackrel{(3.29)}{\equiv}\left(\frac{m_{g} \sum_{i=1}^{m_{g}}\left[\hat{Q}_{\xi}\left(x_{i}\right) \tilde{Q}_{\xi, n}\left(x_{i}\right)\right]-\left[\sum_{i=1}^{m_{g}} \hat{Q}_{\mathcal{\xi}}\left(x_{i}\right)\right]\left[\sum_{i=1}^{m_{g}} \tilde{Q}_{\xi, n}\left(x_{i}\right)\right]}{\sqrt{\left(m_{g} \sum_{i=1}^{m_{g}}\left[\hat{Q}_{\mathcal{\xi}}\left(x_{i}\right)\right]^{2}-\left[\sum_{i=1}^{m_{g}} \hat{Q}_{\xi}\left(x_{i}\right)\right]^{2}\right)\left(m_{g} \sum_{i=1}^{m_{g}}\left[\hat{Q}_{\mathcal{Q}, n}\left(x_{i}\right)\right]^{2}-\left[\sum_{i=1}^{m_{g}} \hat{Q}_{\xi, n}\left(x_{i}\right)\right]^{2}\right)}}\right)^{2} . \tag{3.31f}
\end{align*}
$$

In statistics parlance, if we regard the quantity $\hat{Y}$ as an encoding of the observed values, and if we regard the quantity $\tilde{\hat{Y}}$ as an encoding of the predicted (or computed or simulated) values, then the choice of the SSPCC $r_{\hat{y} \tilde{y}}^{2}$ in (3.31) rests on the assumption that all observed values and all predicted values are equally important. Hence, the weighting of the individual data points is one.

Geometrically speaking, assuming a list of abstract points from an Euclidean space, then the list of predicted values and the list of observed values can be interpreted as the Cartesian coordinates of the abstract points where $\mathbf{e}_{\tilde{Y}}$ and $\mathbf{e}_{\hat{Y}}$ refer to the unit vectors in $\mathbb{R}^{2}$ w.r.t. $\tilde{Y}$ and $\hat{Y}$, respectively. Therefore, the SSPCC $r_{\hat{y} \tilde{y}}^{2}$ indicates how well the relationship between the abstract points can be described by a linear equation in $\mathbb{R}^{2}$. Notice that, by definition, $r_{\hat{y} \tilde{y}}^{2} \in[0,1]$. Thus, if $r_{\hat{y} \tilde{y}}^{2}=1$, then there is a total positive linear correlation and the relationship between the abstract points can be described by the linear equation that reads as

$$
\begin{equation*}
\exists a \in \mathbb{R}^{+} . \exists b \in \mathbb{R} . \forall x_{i} \in X_{\mathbf{s}_{g}, i} \cdot \hat{Q}_{\tilde{S}}\left(x_{i}\right)=a \cdot \tilde{\hat{Q}}_{\mathcal{Z}, n}\left(x_{i}\right)+b . \tag{3.32}
\end{equation*}
$$

Ideally, one can provide the number $a$ and the number $b$ such that $a:=1$ and $b:=0$. However, the choice of the SSPCC $r_{\hat{y} \tilde{y}}^{2}$ in (3.31) is not capable of identifying the case where $a:=1$ and $b:=0$.

If $r_{\hat{y} \tilde{y}}^{2}=0$, then there is no linear correlation, more precisely, one cannot provide a number $a$ and a number $b$ such that (3.32) holds to be true at least for some $x_{i} \in X_{\mathrm{s}_{g}, i}$.

Observe that the geometrical consideration of the SSPCC $r_{\hat{y} \tilde{y}}^{2}$ reveals that the number $m_{g}$ in (3.31) has to satisfy the condition $m_{g}>2$ which can be translated into the requirement that there are at least three abstract points represented as members of the linear span constituted by the unit vectors $\mathbf{e}_{\tilde{\gamma}}$ and $\mathbf{e}_{\hat{\gamma}}$, i.e., $\operatorname{span}\left(\left\{\mathbf{e}_{\tilde{\gamma}}, \mathbf{e}_{\hat{\gamma}}\right\}\right)$; otherwise $r_{\hat{y} \tilde{y}}^{2}$ is immediately equal to one.

In order to assess a low-fidelity model more nuanced, as mentioned above, one can use $r_{\hat{y} \hat{y}}^{2}$ in combination with $\overline{\mathcal{E}}_{\mathrm{H}, \mathrm{s}_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)$. However, if we use the SSPCC in combination with the mean generalization error determined by the $k$-fold cross-validation method, then the condition $m_{g}>2$ requires a minimum number of sampling points $m_{k, \text { min }}$ depending on the number $k$, that is, the number of cells of the randomly created partition. Therefore, for instance, if $k:=5$, then $m_{k:=5, \min }:=15$; and if $k:=10$, then $m_{k:=10, \min }:=30$.

In order to emphasize this kind of dependency of the SSPCC on the $k$-fold cross validation method, let us introduce the map $r_{\hat{y} \tilde{y}, \mathrm{cv}}^{2}$ that reads as

$$
\begin{equation*}
r_{\hat{y} \tilde{y}, \mathrm{cv}}^{2}=\left.k \mapsto r_{\hat{y} \tilde{y}}^{2}\right|_{k}:=r_{\hat{y} \tilde{y}, \mathrm{cv}}^{2}(k): \mathbb{Z}^{+} \rightarrow[0,1], \tag{3.33}
\end{equation*}
$$

where $\left.r_{\tilde{y} \tilde{y}}^{2}\right|_{k}$ and $r_{\tilde{y} \tilde{y}, \mathrm{cv}}^{2}(k)$, respectively, denote the $k$-dependent SSPCC. In analogy to $\overline{\mathcal{e}}_{\mathrm{H}, \mathbf{s}_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)$, for each of the $k$ options in (3.25) and in (3.26), the SSPCC in (3.31) is computed. Hence, the mean $k$-dependent SSPCC (or short: mean SSPCC) $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k}$ is computed as the mean of all the $k$ individual SSPCCs.

Whereas the error in (3.21) focuses on the comparison of the values of the highfidelity model and the low-fidelity model, the SSPCC focuses on the comparison of the shapes (or landscapes) of the high-fidelity model and the low-fidelity model. Thus, if the SSPCC is close to the number 1, then it hints at a high geometrical similarity of the corresponding shapes. Therefore, the SSPCC can be suitable as a supplementary tool to assess quantitatively the similarity of the high-fidelity model and a low-fidelity model.

Notice well that describing shapes by exploiting information about derivatives is a common theme in languages such as, for instance, differential geometry (see, e.g., Detour 1 in § 2.1.2). Since we deploy a gradient-based interpretation of sensitivity measures (recall $\S 2.3 .3$ ), it is mathematically reasonable to associate the $k$-dependent

SSPCC with a low-fidelity model's normalized global first-order sensitivity measures $S_{\hat{y}, i}^{\mathbb{N}}(f)$ with $S_{\hat{y}, i}^{\mathrm{N}}(f) \equiv S_{i}^{\mathbb{N}}\left(\tilde{\hat{Q}}_{\tilde{\zeta}, n}\right)$. Hence, I propose to normalize the $k$-dependent
 cisely, one can define the normalized $k$-dependent SSPCC and the normalized mean SSPCC such that

$$
\begin{equation*}
r_{\hat{y} \tilde{y} \tilde{y}_{k}^{2}}^{\mathrm{N}}:=\frac{r_{\hat{y}}^{2} \tilde{\tilde{y}}_{k}}{\sum_{i=1}^{N_{\tilde{\xi}}} S_{\hat{y}, i}^{\mathrm{N}}(f)},\left.\quad \bar{r}_{\hat{y} \tilde{\tilde{y}}}^{2}\right|_{k} ^{\mathrm{N}}:=\frac{\left.\bar{r}_{\tilde{y} \tilde{y}}^{2}\right|_{k}}{\sum_{i=1}^{\mathrm{N}_{\tilde{\xi}}} S_{\tilde{\hat{y}}, i}^{\mathrm{N}}(f)} . \tag{3.34}
\end{equation*}
$$

Notice that $\left.\left.r_{\hat{y} \tilde{\tilde{y}}}^{2}\right|_{k} ^{\mathrm{N}}{ }_{[0,1]} r_{\hat{y} \tilde{y}}^{2}\right|_{k}$ and $\left.\bar{r}_{\tilde{y} \tilde{y}}^{2}\right|_{k} ^{\mathrm{N}}=[0,1] \bar{r}_{\left.\hat{y} \tilde{y}\right|_{k^{\prime}} ^{2}}$ that is, the corresponding entities are equal as numbers. However, they are conceptually different. A benefit of introducing the entity $\left.r_{\hat{y} \tilde{y}}^{2}\right|_{k} ^{\mathrm{N}}$ - and the entity $\left.\bar{r}_{\hat{y} \tilde{\hat{y}}}^{2}\right|_{k} ^{\mathrm{N}}$ as well - is to highlight the connection of various information sources for assessing the shape (or landscape) of the lowfidelity model with regard to the high-fidelity model. Another benefit is that $\left.r_{\hat{y} \tilde{y}}^{2}\right|_{k} ^{\mathrm{N}}$
 ity measures associated with a low-fidelity model, that is, $S_{\hat{\tilde{y}}, i}^{\mathrm{N}}(f)$, as proxies for the normalized global first-order sensitivity measures associated with the high-fidelity model, that is, $S_{\hat{y}, i}^{N}(f)$ with $S_{\hat{y}, i}^{N}(f) \equiv S_{i}^{N}\left(\hat{Q}_{\tilde{\xi}}\right)$. To put the conjecture in more formal words:

Conjecture (Trustworthiness of low-fidelity models' normalized global first-order sensitivity measures). Given $k$ and $m$ such that $m>m_{k, \text { min }}$, then there exist some lowfidelity models such that

$$
\begin{equation*}
\forall i \in\left\{1, \ldots, N_{\tilde{\xi}}\right\} . S_{\hat{y}, i}^{N}(f) \rightarrow S_{\hat{y}, i}^{N}(f) \text { as } m \rightarrow \infty \Longrightarrow \bar{r}_{\left.\tilde{y} \tilde{y}\right|_{k} ^{2}}^{N} \rightarrow 1 \text { as } m \rightarrow \infty .{ }^{8} \tag{3.35}
\end{equation*}
$$

Remark 3.1.6. If and only if the case

$$
\begin{equation*}
\forall i \in\left\{1, \ldots, N_{\tilde{\zeta}}\right\} \cdot S_{\hat{y}, i}^{N}(f)=S_{\hat{y}, i}^{N}(f) \tag{3.36}
\end{equation*}
$$

holds, then a low-fidelity model's sensitivity measures is considered as total trustworthy proxies for a high-fidelity model's sensitivity measures.

The contrapositive of the statement in (3.35) emphasizes that if $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k} ^{\mathrm{N}}$ is not asymptotically converging to one as $m$ tends to infinity, then one cannot expect that the low-fidelity model's sensitivity measures are trustworthy proxies at all. To my best knowledge, a thorough formal investigation of the above-mentioned conjecture is lacking. Mind that such a thorough formal investigation is out of the scope of the present work. The conjecture should be rather understood as an attempt to jot down formally an accumulation of experimental observations than an attempt to infer logically from a bundle of theoretical insights.

In § 3.2, I present a small data-driven investigation of the statement in (3.35) by means of numerical experiments with regard to the test functions in Table 2.1. Notice that, in this investigation, the sample size is limited, though. Hence, the asymptotic behavior is not examined but, primarily, the pre-asymptotic behavior - since, from

[^21]an application-driven viewpoint, the pre-asymptotic behavior is particular interesting.

In practical applications, the statement in (3.35) inspires to introduce for each $i \in\left\{1, \ldots, N_{\tilde{\xi}}\right\}$ a low-fidelity models' normalized global first-order sensitivity measures (LFSM) error $\mathfrak{e}_{m_{j}}\left(S_{\tilde{y}, i}^{N}\right)$ that reads as

$$
\begin{equation*}
\mathfrak{e}_{m_{j}}\left(S_{\hat{y}, i}^{\mathrm{N}}\right):=\frac{S_{i, m_{j}}^{\mathrm{N}}(f)-S_{i, m_{j-1}}^{\mathrm{N}}(f)}{S_{i, m_{j}}^{\mathrm{N}}(f)}, \tag{3.37}
\end{equation*}
$$

where $m_{j}$ and $m_{j-1}$ denote sample sizes such that $m_{j}>m_{j-1}$. The errors in (3.37) track the size as well as the orientation of the discrepancy between a low-fidelity model's normalized global first-order sensitivity measures w.r.t. a sample size $m_{j}$ and a sample size $m_{j-1}$. The notation $\mathfrak{e}_{m_{\infty}}\left(S_{\hat{y}, i}^{\mathrm{N}}\right)$ refers to the situation where a reference value, e.g., from an analytical calculation, is provided by a user.

### 3.1.2 Deterministic and probabilistic data-fit low-fidelity models

I only cover a small part of the vast territory of available deterministic and probabilistic surrogate models. For more examples from the zoo of surrogates, I refer to, e.g., [61] and [91] and references therein. Additionally, the new, still developing, Julia PL package surrogates. jl (see [23]) is recommended that is part of the larger open source software project called SciML: Scientific Machine Learning (see https://sciml.ai/).

The choice of surrogate models as well as the respective terminologies and the technicalities reflect partly a bias towards mesh-free and mesh-based numerical models (see, e.g., § 2.2) - albeit, these surrogate models are also connected to data-driven statistical models. All the surrogate models described in this subsection are essentially representable as an expansion of basis functions as shown in (3.7).

## Multivariate polynomials

In the previous subsection 3.1.1, we have encountered the prototypical hypothesis space $\mathrm{P}_{\leq n}$ which is the set of all univariate algebraic polynomials of degree at most $n$ on an interval $X \subset \mathbb{R}$ equipped with an $\mathbb{R}$-vector space structure and the finite monomial basis $B \subseteq P_{\leq n}$ that reads as

$$
\begin{equation*}
B:=\left\{1, x^{1}, \ldots, x^{n-1}, x^{n}\right\} \tag{3.38}
\end{equation*}
$$

such that $\mathrm{P}_{\leq n}$ can be regarded as the linear span of $B$, i.e., $\mathrm{P}_{\leq n} \equiv \operatorname{span}(B)$. A member of the space $\mathrm{P}_{\leq n}$ is a univariate polynomial $p=x \mapsto p(x): X \rightarrow \mathbb{R}$ which can be portrayed as

$$
\begin{equation*}
p(x):=c_{0} x^{0}+c_{1} x^{1}+\cdots+c_{n-1} x^{n-1}+c_{n} x^{n} \equiv \sum_{i=0}^{n} c_{i} x^{i}, \tag{3.39}
\end{equation*}
$$

where it holds that $x^{0} \equiv 1, n \in \mathbb{N}$ and the coefficients $c_{i} \in \mathbb{R}$ with $i \in\{0, \ldots, n\}$ and $c_{n} \neq 0$. Using the generic representation in (3.7), one can write (3.38) as

$$
\begin{equation*}
B:=\left\{\tilde{\varphi}_{i}(x) \equiv x^{i} \mid i \in\{0,1, \ldots, n-1, n\}\right\}, \tag{3.40}
\end{equation*}
$$

and one can define the linear span of $B$ as

$$
\begin{equation*}
\operatorname{span}(B):=\left\{\sum_{i=0}^{n} \tilde{c}^{i} \cdot \mathbb{R} \tilde{\varphi}_{i}(x) \mid n \in \mathbb{N} \wedge \tilde{\varphi}_{i}(x) \in B \wedge \tilde{c}^{i} \in \mathbb{R}\right\} .9 \tag{3.41}
\end{equation*}
$$

Following the path in [201, p. 154ff], let us wield the tensor product construction in order to articulate a space of multivariate polynomials. The construction's underlying principle is to express a $d$-variate polynomial with the arguments $\left(x_{1}, \ldots, x_{d}\right) \in$ $X \subseteq \mathbb{R}^{d}$ as a combination of $d$ univariate polynomials such as in (3.39).

Let $N:=\left(n_{1}, \ldots, n_{d}\right) \in \mathbb{N}_{0}^{d}$ be the ordered set of the degrees of the $d$ univariate polynomials. Furthermore, let $I:=\left(i_{1}, \ldots, i_{d}\right) \in \mathbb{N}_{0}^{d}$ denote an ordered set of $d$ indices, more precisely, a multi-index of $d$ members, and $|I|:=i_{1}+\cdots+i_{d}$ designates the degree of the multi-index $I$ or the total degree of the monomial $x^{I}$ which can be written as

$$
\begin{equation*}
x^{I}:=x_{1}^{i_{1}} x_{2}^{i_{2}} \cdots x_{d-1}^{i_{d-1}} x_{d}^{i_{d}}, \tag{3.42}
\end{equation*}
$$

where $x^{(0, \ldots, 0)} \equiv x_{1}^{0} \cdots x_{d}^{0}$ and $x^{(0, \ldots, 0)} \equiv 1$. Then, one can write a multivariate polynomial $p=x \mapsto p(x): X \rightarrow \mathbb{R}$ as

$$
\begin{equation*}
p(x):=\sum_{I \leq N} c_{I} x^{I}, \tag{3.43}
\end{equation*}
$$

where the coefficients $c_{I} \in \mathbb{R}$ are scalars and $I \leq N$ encodes ( $i_{1} \leq n_{1}, \ldots, i_{d} \leq n_{d}$ ). The maximal degree of the monomial $x^{I}$ in (3.42) can be expressed as $\max \{I\} \leq N$ that encodes $\left(\max \left\{i_{1}\right\} \leq n_{1}, \ldots, \max \left\{i_{d}\right\} \leq n_{d}\right)$; and the notion of a multivariate polynomial's degree $\operatorname{deg}(p)$ can be defined as

$$
\begin{equation*}
\operatorname{deg}(p):=\max \left\{|I| \mid c_{I} \neq 0\right\} \tag{3.44}
\end{equation*}
$$

Finally, the space of $d$-variate polynomials of total degree at most $k$ can be expressed as the direct sums of tensor products of $d$ spaces of univariate polynomials:

$$
\begin{equation*}
\mathrm{P}_{\leq k}^{d}:=\bigoplus_{|| | \leq k} \mathrm{P}_{i_{1}} \otimes \cdots \otimes \mathrm{P}_{i_{d}} \cdot{ }^{10} \tag{3.45}
\end{equation*}
$$

Let us restrict to the case in which $N \equiv(k, \ldots, k)$ with $k \in \mathbb{N}_{0}$, that is, each $x_{i}$ with $i \in\{1, \ldots, d\}$ is associated with a univariate polynomial of degree $k$. Then, a basis of the space $P_{s k}^{d}$ can be written as

$$
\begin{align*}
\bigotimes_{i=1}^{d} B_{i} & :=B_{1} \otimes \cdots \otimes B_{d}  \tag{3.46a}\\
& :=\left\{\tilde{\varphi}_{1 i_{1}}(x) \otimes \cdots \otimes \tilde{\varphi}_{d i_{d}}(x) \equiv x^{i_{1}} \otimes \cdots \otimes x^{i_{d}}\left|I \in\{0,1, \ldots, k\}^{d} \wedge\right| I \mid \leq k\right\}  \tag{3.46b}\\
& :=\left\{\tilde{\varphi}_{1 i_{1}}\left(x_{1}\right) \cdots \tilde{\varphi}_{d i_{d}}\left(x_{d}\right) \equiv x^{I}\left|I \in\{0,1, \ldots, k\}^{d} \wedge\right| I \mid \leq k\right\}  \tag{3.46c}\\
& =B^{\otimes d}, \tag{3.46d}
\end{align*}
$$

[^22]where $B_{i}$ is the basis $B$ in (3.40) with respect to $x_{i}$. Given the basis in (3.46), one can express the dimension of the space $\mathrm{P}_{\leq k}^{d}$ via the binomial coefficient $\binom{k+d}{d}$ such that
\[

$$
\begin{equation*}
\operatorname{dim}\left(\mathrm{P}_{\leq k}^{d}\right) \equiv\binom{k+d}{d} . \tag{3.47}
\end{equation*}
$$

\]

For example, the dimension of the space $P_{\leq 2}^{2}$ is $\operatorname{dim}\left(P_{\leq 2}^{2}\right):=6$ and a basis of this space can be written as

$$
\begin{align*}
\bigotimes_{i=1}^{2} B_{i} & :=B_{1} \otimes B_{2}  \tag{3.48a}\\
& :=\left\{\tilde{\varphi}_{1 i}(x) \otimes \tilde{\varphi}_{2 j}(x) \equiv x^{i} \otimes x^{j} \mid(i, j) \in\{0,1,2\}^{2} \wedge i+j \leq 2\right\}  \tag{3.48b}\\
& :=\left\{\tilde{\varphi}_{1 i}\left(x_{1}\right) \tilde{\varphi}_{2 j}\left(x_{2}\right) \equiv x_{1}^{i} x_{2}^{j} \mid(i, j) \in\{0,1,2\}^{2} \wedge i+j \leq 2\right\}  \tag{3.48c}\\
& =B^{\otimes 2}, \tag{3.48d}
\end{align*}
$$

where $B_{1}$ and $B_{2}$ are the basis $B$ in (3.40) for $x_{1}$ and $x_{2}$, respectively. A polynomial $p \in \operatorname{span}\left(B^{\otimes 2}\right)$ can be represented as

$$
\begin{equation*}
p(x):=c_{(0,0)}+c_{(1,0)} x_{1}+c_{(0,1)} x_{2}+c_{(2,0)} x_{1}^{2}+c_{(0,2)} x_{2}^{2}+c_{(1,1)} x_{1} x_{2}, \tag{3.49}
\end{equation*}
$$

where an ordering for the multi-index $I$ is chosen which does respect the degree $|I|$.
Occasionally, the basis associated with the space $P_{\leq k}^{d}$ is called a complete polynomial basis. The basis associated with the space $\mathrm{P}_{k}^{d}$, that is, the space of $d$-variate polynomials of maximal degree at most $k$, is called a tensor product polynomial basis. The space $\mathrm{P}_{k}^{d}$ is constructed by applying the condition $\max \{I\} \leq N$ in (3.45). Its dimension can be stated as

$$
\begin{equation*}
\operatorname{dim}\left(\mathrm{P}_{k}^{d}\right) \equiv(1+k)^{d} \tag{3.50}
\end{equation*}
$$

In Table 3.1, $\operatorname{dim}\left(\mathrm{P}_{\leq k}^{d}\right)$ and $\operatorname{dim}\left(\mathrm{P}_{k}^{d}\right)$ are listed for some pairs $(k, d) \in \mathbb{N}^{2}$.
TABLE 3.1: Given some pairs $(k, d) \in \mathbb{N}^{2}$, the dimension of $\mathrm{P}_{\leq k}^{d}$ and $\mathrm{P}_{k}^{d}$.

| $(k, d)$ | $\operatorname{dim}\left(\mathrm{P}_{\leq k}^{d}\right)$ | $\operatorname{dim}\left(\mathrm{P}_{k}^{d}\right)$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $(2,2)$ | 6 | 9 |  | $(3, d)$ | $\operatorname{dim}\left(\mathrm{P}_{\leq k}^{d}\right)$ | $\operatorname{dim}\left(\mathrm{P}_{k}^{d}\right)$ |
| $(2,3)$ | 10 | 27 |  | 10 | 16 |  |
| $(2,4)$ | 15 | 81 |  | $(3,4)$ | 20 | 64 |
| $(2,5)$ | 21 | 243 |  | $(3,5)$ | 56 | 256 |
|  |  |  |  |  |  |  |

For a fixed integer $k>0$ and a fixed integer $d>1$, in general, one can observe that

$$
\begin{equation*}
\operatorname{dim}\left(\mathrm{P}_{\leq k}^{d}\right)<\operatorname{dim}\left(\mathrm{P}_{k}^{d}\right) . \tag{3.51}
\end{equation*}
$$

For a given integer pair $(k, d)$, empirically, the computation time associated with the polynomials in the space $\mathrm{P}_{\leq k}^{d}$ is frequently lower than for polynomials in the space $\mathrm{P}_{k}^{d}$, while the approximation quality or accuracy is only slightly lower.

Recalling the test functions in Table (2.1), one cannot expect a globally sufficiently accurate approximation of those test functions that include periodic parts. From an optimization point of view, however, one can expect a locally sufficiently accurate approximation in the neighborhood of the global optimum. The rationale behind
these expectations is rooted in the relationship of the notion of approximation quality for the polynomials in the space $P_{\leq k}^{d}$ with the notion of approximation quality for the $d$-variate Taylor-kind polynomials of degree $k$.

Admittedly, I do not elaborate on this relationship; it triggers, though, an important special case regarding low-fidelity models which is the space of $d$-variate polynomials of total degree at most two. The corresponding polynomials $p \in P_{\leq 2}^{d}$ are called response surfaces (see, e.g., [61, p. 27]).

A $d$-variate polynomial of degree at most two can be presented in the form

$$
\begin{equation*}
p(x):=c_{(0, \ldots, 0)}+\sum_{i=1}^{d} e_{i} x_{i}+\sum_{i=1}^{d} \sum_{j=1}^{d} a_{i, j} x_{i} x_{j}, \tag{3.52}
\end{equation*}
$$

where $e_{i} \in \mathbb{R}$ and $a_{i, j} \in \mathbb{R}$ are scalars. Invoking Householder's notation for matrix operations (see, e.g., [99, p. 1ff]) and given $p: \mathbb{R}^{d \times 1} \rightarrow \mathbb{R}$, we can write (3.52) as

$$
\begin{equation*}
p(x):=\beta_{0}+e^{\mathrm{T}} x+x^{\mathrm{T}} A x, \tag{3.53}
\end{equation*}
$$

where $\beta_{0} \in \mathbb{R}$ represents a scalar, $x:=\left[x_{i}\right] \in \mathbb{R}^{d \times 1}$ and $e:=\left[e_{i}\right] \in \mathbb{R}^{d \times 1}$ represent column vectors (or column matrices) and $A:=\left[a_{i, j}\right] \in \mathbb{R}^{d} \times \mathbb{R}^{d}$ - with $\mathbb{R}^{d} \times \mathbb{R}^{d} \cong \mathbb{R}^{d \times d}$ - represents a quadratic matrix (or square matrix). ${ }^{11}$ By introducing the maps $l_{e}=x \mapsto e^{\mathrm{T}} x$ : $\mathbb{R}^{d \times 1} \rightarrow \mathbb{R}$ and $q_{A}=x \mapsto x^{\mathrm{T}} A x: \mathbb{R}^{d \times 1} \rightarrow \mathbb{R}$, a map-oriented presentation of (3.53) can be achieved:

$$
\begin{equation*}
p(x):=\beta_{0}+l_{e}(x)+q_{A}(x) . \tag{3.54}
\end{equation*}
$$

A possible matrix representation of the basis in (3.48c) is the column vector $\tilde{b} \epsilon$ $\mathbb{R}^{6 \times 1}$ with

$$
\tilde{b}:=\left[\begin{array}{llllll}
1 & x_{1} & x_{2} & x_{1}^{2} & x_{2}^{2} & x_{1} x_{2} \tag{3.55}
\end{array}\right]^{\mathrm{T}} .
$$

Given the order of the components in (3.55), the authors in [61, p. 133] show a general construction rule how to obtain the components of a column vector $\tilde{b} \in \mathbb{R}^{s \times 1}$ with $s:=d(d+3) / 2+1$ (cf. (3.47)) that represents the basis of a $d$-variate polynomial of degree at most two:

$$
\tilde{b}:=\left[\begin{array}{llllllllll}
\tilde{b}_{0} & \tilde{b}_{1} & \ldots & \tilde{b}_{d} & \tilde{b}_{d+1} & \ldots & \tilde{b}_{2 d} & \tilde{b}_{2 d+1} & \ldots & \tilde{b}_{s-1} \tag{3.56}
\end{array}\right]^{\mathrm{T}},
$$

where $\tilde{b}_{0}=1, \tilde{b}_{1}=x_{1}, \tilde{b}_{d}=x_{d}, \tilde{b}_{d+1}=x_{1}^{2}, \tilde{b}_{2 d}=x_{d}^{2}, \tilde{b}_{2 d+1}=x_{1} x_{2}$, and $\tilde{b}_{s-1}=x_{d-1} x_{d}$. Hence, if we introduce a column vector $\tilde{c} \in \mathbb{R}^{s \times 1}$ which encapsulates the coefficients with regard to the components of $\tilde{b}$, one can reformulate (3.53) as

$$
\begin{equation*}
p(x):=\tilde{b}^{\mathrm{T}} \tilde{c} \tag{3.57}
\end{equation*}
$$

Assuming a sample s such as in (3.13), one can employ the corresponding sampling plan points in (3.57). Thus, one can define a column vector $y \in \mathbb{R}^{m \times 1}$ whose components are the output points $y_{i}$ and one can succinctly define a matrix $B \in \mathbb{R}^{m} \times \mathbb{R}^{s}$

[^23]with respect to the sampling plan points $x_{i}$ :
\[

B:=\left[$$
\begin{array}{c}
\tilde{b}_{x_{1}}^{\mathrm{T}}  \tag{3.58}\\
\tilde{b}_{x_{2}}^{\mathrm{T}} \\
\vdots \\
\tilde{b}_{x_{m}}^{\mathrm{T}}
\end{array}
$$\right] .
\]

In a verbose mode, the matrix in (3.58) displays

$$
B:=\left[\begin{array}{cccccccccc}
1 & x_{11} & \ldots & x_{1 d} & x_{11}^{2} & \ldots & x_{1 d}^{2} & x_{11} x_{12} & \ldots & x_{1 d-1} x_{1 d}  \tag{3.59}\\
1 & x_{21} & \ldots & x_{2 d} & x_{21}^{2} & \ldots & x_{2 d}^{2} & x_{21} x_{22} & \ldots & x_{2 d-1} x_{2 d} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_{m 1} & \ldots & x_{m d} & x_{m 1}^{2} & \ldots & x_{m d}^{2} & x_{m 1} x_{m 2} & \ldots & x_{m d-1} x_{m d}
\end{array}\right],
$$

where $x_{m d}$ denotes the $d$-th coordinate of the $m$-th sampling plan point. Using the matrix $B$, one can define a function $h=\tilde{c} \mapsto B \tilde{c}: \mathbb{R}^{s} \rightarrow \mathbb{R}^{m}$ and one can state an inverse problem (cf. (3.1c) in §3.1.1) that reads as

$$
\begin{equation*}
\text { given } B \in \mathbb{R}^{m} \times \mathbb{R}^{s} \text { and } y \in \mathbb{R}^{m \times 1} \text {, find } \tilde{c} \in \mathbb{R}^{s \times 1} \text { such that } B \tilde{c}=y \text {. } \tag{3.60}
\end{equation*}
$$

If the condition $m=s$ holds and $\operatorname{rank}(B)$ is full, the left and the right inverse of the matrix $B$ exist. ${ }^{12}$ Hence, there is a unique solution $\tilde{c} \equiv B^{-1} \tilde{y}$ which is a determination of the coefficients by interpolation (cf. (3.18) in $\S$ 3.1.1).

In the applications of the present work, however, the condition $m>s$ usually holds that leads to an overdetermined system of linear equations in (3.60) such that $B^{-1}$ does not exist. In this case, the inverse's purpose is most adequately emulated by the pseudoinverse. For more details on the properties of the pseudoinverse, I refer to, e.g., [200, p. 618f].

Given the space-filling property and the non-collapsing property of a sampling plan, it is reasonable to assume that the column rank of $B$ is full such that the inverse $\left(B^{\mathrm{T}} B\right)^{-1}$ exists, then the pseudoinverse $B^{+} \in \mathbb{R}^{s} \times \mathbb{R}^{m}$ can be stated as

$$
\begin{equation*}
B^{+}:=\left(B^{\mathrm{T}} B\right)^{-1} B^{\mathrm{T}}, \tag{3.61}
\end{equation*}
$$

which can be computed efficiently by the singular value decomposition method. Using the pseudoinverse $B^{+}$results in a reformulation of the problem in (3.60) in terms of a projection matrix $P_{B} \in \mathbb{R}^{m} \times \mathbb{R}^{m}$ with $P_{B}:=B B^{+}$where $\operatorname{tr}\left(P_{B}\right) \equiv \operatorname{rank}(B)$ such that one can define a column vector $\hat{y} \in \mathbb{R}^{m \times 1}$ via $\hat{y}:=P_{B} y .{ }^{13}$ The corresponding coefficients column vector $\hat{\tilde{c}} \in \mathbb{R}^{q}$ such that

$$
\begin{equation*}
\hat{\tilde{c}}:=B^{+} y \tag{3.62}
\end{equation*}
$$

is the best solution in the sense of a linear multiple regression by the least squares

[^24]method. Recalling (3.19), the best solution $\hat{\tilde{c}} \in \mathbb{R}^{s \times 1}$ is associated with the optimization problem
\[

$$
\begin{equation*}
\operatorname{minimize}_{\tilde{c} \in \mathbb{R}^{\times \times 1}} \operatorname{rss}(\tilde{c}):=\frac{1}{2}(y-B \tilde{c})^{\mathrm{T}}(y-B \tilde{c}), \tag{3.63}
\end{equation*}
$$

\]

where the objective function rss: $\mathbb{R}^{s \times 1} \rightarrow \mathbb{R}$ is sometimes called the residual sum-ofsquares function (see, e.g., [91, p. 30]) with $r \in \mathbb{R}^{m \times 1}$ designating the residual column vector (abbreviated to residual) such that $r:=(y-B \tilde{c})$.

In order to ensure the well-posedness of the optimization problem, a fruitful generalization of the basic discrete least squares $l_{2}$ approximation problem in (3.63) is the Tikhonov regularized weighted least squares $l_{2}$ approximation problem that can be stated as

$$
\begin{equation*}
\underset{\tilde{c} \in \mathbb{R}^{s \times 1}}{\operatorname{minimize}} \operatorname{rss}(\tilde{c}):=\frac{1}{2}\|y-B \tilde{c}\|_{W}^{2}+\frac{1}{2}\left\|\tilde{c}-\tilde{c}_{0}\right\|_{R}^{2}, \tag{3.64}
\end{equation*}
$$

that leads to the normal equations and the best solution, respectively,

$$
\begin{equation*}
\hat{\tilde{c}}:=\left(B^{\mathrm{T}} W B+R\right)^{-1}\left(B^{\mathrm{T}} W y+R \tilde{c}_{0}\right), \tag{3.65}
\end{equation*}
$$

where the Tikhonov matrix $R \in \mathbb{R}^{s} \times \mathbb{R}^{s}$ and the residual variance-covariance matrix $W \in$ $\mathbb{R}^{m} \times \mathbb{R}^{m}$ denote symmetric positive definite diagonal matrices such that

$$
\begin{align*}
& \|\cdot\|_{W}=v \mapsto v^{\mathrm{T}} W v: \mathbb{R}^{m \times 1} \rightarrow \mathbb{R},  \tag{3.66}\\
& \|\cdot\|_{R}=v \mapsto v^{\mathrm{T}} R v: \mathbb{R}^{s \times 1} \rightarrow \mathbb{R}, \tag{3.67}
\end{align*}
$$

and $\tilde{c}_{0} \in \mathbb{R}^{s \times 1}$ denotes a column vector that represents the initial guess about the best solution (cf. [201, p. 71]).

Since the Tikhonov matrix $R$ encodes the regularization, it is set that $R:=L L^{T}$ where one can define $L \in \mathbb{R}^{s} \times \mathbb{R}^{s}$ as $L:=\sqrt{\lambda} I$ with the regularization parameter $\lambda \in$ [ 0,1 [ and $I \in \mathbb{R}^{s} \times \mathbb{R}^{s}$ being the identity matrix. The residual variance-covariance matrix $W$ encodes the weighting of the components of the squared residual in the sense that $W:=\operatorname{diag}\left(\sigma^{-2}, \ldots, \sigma^{-2}\right)$ where $\sigma$ denotes the constant conditional error variance in (3.20). ${ }^{14}$

Hence, let us consider all the components of the squared residual as uncorrelated (represented by setting all of W's off-diagonal entries to zero) and on a par with each other (represented by setting all of $W$ 's diagonal entries to the same positive number). If we set $W:=I$ with $I \in \mathbb{R}^{m \times m}$ being the identity matrix and $\lambda \equiv 0$ in the Tikhonov matrix, we recover the problem in (3.63) as a special case.

Mind, though, if the original problem in (3.63) is ill-conditioned, choosing $\lambda$ too small will not change much, choosing $\lambda$ too big leads much more to a detachment from the original problem. Thus, finding an optimal regularization parameter is not a trivial task and it depends highly on the problem at hand and the judgment of the user. Let us interpret the regularization parameter as a hyperparameter (recall Remark 3.1.4).

Given the optimal coefficients as floating-point numbers, a numerically stable approach to evaluate the function in (3.57) is Clenshaw's recurrence formula (see, e.g.,

[^25][171, p. 222f]) - which, in the case of a monomial sum, is the familiar Horner's method - that exploits the inherent recurrence relation and avoids the explicit evaluation of the polynomial functions in (3.56). For a more elaborate discussion on the propagation of the rounding error in the context of polynomial evaluation, see, e.g., [161]. In Listing 3.1, I present an example implementation of Clenshaw's algorithm for the evaluation of a univariate monomial sum in the Julia PL.

LISTING 3.1: An example implementation of Clenshaw's algorithm for the evaluation of a univariate monomial sum in the Julia PL.

```
function monomial_clenshaw_eval_1d(\tilde{c}:: Vector{T}, x::T) where {T<:Real}
    N = size(\tilde{c},1) - 1 # 1-based indexing
    d = zeros(N+2)
    d[N+2] = 0
    d[N+1] = \tilde{c}[N+1]
        for i in N:-1:2
            d[i] = x*d[i+1] + \tilde{c}[i]
        end
    return x*d[2] + c\tilde{c}[1]
end
```

In the multivariate case, one can apply a plain greedy approach in the sense that one can invoke multiple nested hierarchical univariate monomial sum evaluations. For instance, if we possess the space $\mathrm{P}_{\leq 2}^{2}$ with $\operatorname{dim}\left(\mathrm{P}_{\leq 2}^{2}\right) \equiv 6$, then one can introduce the column vectors $\tilde{\phi}\left(x_{2}\right) \in \mathbb{R}^{6 \times 1}$ and $\tilde{\psi}\left(x_{1}\right) \in \mathbb{R}^{6 \times 1}$ and the diagonal matrix $\tilde{\Sigma} \in \mathbb{R}^{6} \times \mathbb{R}^{6}$ such that

$$
\begin{align*}
& \tilde{\phi}\left(x_{2}\right):=\left[\begin{array}{llllll}
x_{2}^{0} & x_{2}^{0} & x_{2}^{0} & x_{2}^{1} & x_{2}^{2} & x_{2}^{1}
\end{array}\right]^{\mathrm{T}},  \tag{3.68a}\\
& \tilde{\psi}\left(x_{1}\right):=\left[\begin{array}{llllll}
x_{1}^{0} & x_{1}^{1} & x_{1}^{2} & x_{1}^{0} & x_{1}^{0} & x_{1}^{1}
\end{array}\right]^{\mathrm{T}} \text {, }  \tag{3.68b}\\
& \tilde{\Sigma}:=\operatorname{diag}\left(\tilde{c}_{1}, \tilde{c}_{2}, \tilde{c}_{3}, \tilde{c}_{4}, \tilde{c}_{5}, \tilde{c}_{6}\right),  \tag{3.68c}\\
& p(x):=\tilde{\phi}\left(x_{2}\right)^{\mathrm{T}} \tilde{\Sigma} \tilde{\psi}\left(x_{1}\right) \text {. } \tag{3.68d}
\end{align*}
$$

A display that is favorable for the application of the multiple nested hierarchical evaluations is

$$
\begin{equation*}
p(x):=\sum_{i=1}^{\operatorname{dim}\left(\mathrm{P}_{\leq 2}^{2}\right)} \tilde{c}_{j} \tilde{\phi}_{j}\left(x_{2}\right) \tilde{\psi}_{j}\left(x_{1}\right), \tag{3.69}
\end{equation*}
$$

where, firstly, the terms $\tilde{c}_{j} \tilde{\phi}_{j}\left(x_{2}\right)$ are evaluated, and, secondly, these evaluated terms are used as the coefficients for the evaluation of the terms $\tilde{\psi}_{j}\left(x_{1}\right)$. The proper generalization of Horner's method to multivariate polynomials is still an active research area (see, e.g., [130]). Furthermore, notice well that the display in (3.69) hints at the connection to the vivid research area of the computationally efficient representation of multivariate functions using low-rank tensor approximation techniques (see, e.g., [206], [90]). However, in the present work, let us leave it at that adumbration.

I have argued that due to a sampling plan's space-filling property and its noncollapsing property, it is reasonable to assume the generic case in which the matrix $B^{\mathrm{T}} B$ is invertible (or non-singular or non-degenerate). However, this premise could be challenged. Hence, let us glance briefly at the influence of the arrangement of a
sampling plan on the condition number $\kappa\left(B^{\mathrm{T}} B\right)$ with the property

$$
\begin{equation*}
\kappa\left(B^{\mathrm{T}} B\right) \equiv(\kappa(B))^{2} \cdot{ }^{15} \tag{3.70}
\end{equation*}
$$

Thus, the condition number with respect to $B^{\mathrm{T}} B$ is always worse than the condition number with respect to $B$. Let us focus on instances of a multicollinearity with respect to the chosen basis - where, in a theoretical absence of numerical errors, one could spot that the column rank of $B$ is not full.

Observing Figure 3.1 and Figure 3.2, a situation is conceivable where a sampling plan could be constructed as, for instance,

$$
\begin{align*}
X_{\mathrm{s}, 1}:=\{ & (0.1,0.6),(0.2,0.3),(0.4,0.7),(0.5,0.8), \\
& (0.6,0.1),(0.3,0.2),(0.7,0.4),(0.8,0.5)\}, \tag{3.71}
\end{align*}
$$

where $X_{\mathrm{s}, 1}$ is represented by a $\mathbb{R}^{m} \times \mathbb{R}^{d}$ matrix with $m=8$ and $d=2$ such that the condition number is $1.06 \times 10^{4}$ (see (i) in Figure 3.4). The underlying construction principle is based on the Householder reflection matrix $H \in \mathbb{R}^{d} \times \mathbb{R}^{d}$ with $H:=I-2 \frac{v v^{\mathrm{T}}}{v^{\mathrm{T}}}$ where $I \in \mathbb{R}^{d} \times \mathbb{R}^{d}$ denotes the identity matrix and $v \in \mathbb{R}^{d \times 1}$ denotes a column vector. Let us choose $v$ such that it is orthogonal to the vector $v_{t}:=\sum_{i=0}^{d} e_{i}$ where $e_{i}$ denote the standard basis vectors of $\mathbb{R}^{d}$.

Another illustration is a sampling plan $X_{\mathrm{s}, 2}$ such that

$$
\begin{equation*}
\forall i \in\{1, \ldots, m\} \cdot x_{i d-1}=x_{i d} \Longrightarrow \forall i \in\{1, \ldots, m\} \cdot b_{i, s-1}=b_{i, 2 d} . \tag{3.72}
\end{equation*}
$$

For an example regarding the case $m=8$ where the condition number is $8.67 \times 10^{49}-$ which, numerically, indicates a singular matrix -, consult (ii) in Figure 3.4. Let us compare the sampling plan $X_{\mathbf{s}, 1}$ in (3.71) and $X_{\mathbf{s}, 2}$ in (3.72) with a sampling plan $X_{\mathbf{s}, 3}$ based on the Sobol quasi-random sequence where the condition number is $3.23 \times 10^{3}$ (see (iii) in Figure 3.4). Even in the case of a space-filling and non-collapsing sam-


Figure 3.4: Sampling plan $X_{\mathbf{s}}$ [condition number $\left.\kappa\left(B^{\mathrm{T}} B\right)\right]$.
(i) $X_{s, 1}\left[1.06 \times 10^{4}\right]$,
(ii) $X_{\mathrm{s}, 2}\left[8.57 \times 10^{49}\right]$, (ii)
(iii) $X_{s, 3}\left[3.23 \times 10^{3}\right]$.

[^26]pling plan such as the Sobol quasi-random sequence, one can observe a high condition number. It is an indication that the familiar ill-conditioned behavior of a monomial basis for the space $P_{\leq 2}$ is mimicked by the monomial basis for the space $P_{\leq 2}^{d}$.

In Appendix A, numerical experiments are conducted with regard to a reparametrization using mean-centered arguments, Bernstein polynomials, and Chebyshev polynomials.

In Figure 3.5, the monomial basis in (3.46), the Bernstein basis in (A.3), and the Chebyshev basis in (A.10) for the space $\mathrm{P}_{\leq 2}^{1}$ are exhibited. Note that there are Julia PL packages such as MultivariatePolynomials.jl (see doi:10.5281/zenodo.3839754) that, in their definition of various kinds of polynomials, utilize intensely several language features of the Julia PL, e.g., its type system or its metaprogramming capabilities. Note further that there are MATLAB ${ }^{\circledR}$ PL toolboxes such as Chebfun3 (see [90]) that are purely written in MATLAB. These toolboxes avoid using, e.g., MEX files, i.e., MATLAB executables. Let us not dwell on language-related issues because there is a lack of comprehensive studies (cf. § 2.3.3) that, for such particular use cases, compare thoroughly the pros and cons of each programming language in terms of performance, readability or maintainability - to name but a few criteria. For example, it is difficult to determine whether potential performance differences are due to different language designs or different implementations, or both.


Figure 3.5: Basis under consideration for the space $P_{\leq 2}^{1}$.
(i) Monomial basis, (ii) Bernstein basis, (iii) Chebyshev basis.

Mind that the Chebyshev grid (see Figure A.1) is closely linked to the numerical technique of sparse grids (see, e.g., [39]). This technique is particularly useful when one is dealing with the space $\mathrm{P}_{k}^{d}$ where high values are chosen for $k$ and $d$ such that $\operatorname{dim}\left(\mathrm{P}_{k}^{d}\right)$ in (3.50) is high as well. Thus, the dimensionality $d$ is a curse in the sense that the computational costs in terms of memory and time grow exponentially in the dimensionality $d$.

Based on ideas from information-based complexity theory (see, e.g., [207]), the authors in [39] invoke a formal encoding of the curse of dimensionality by the complexity estimate $O\left(\varepsilon^{-\alpha d}\right)$ where the non-negative real number $\varepsilon$ denotes a desired accuracy of an approximate solution and the non-negative real number $\alpha$ is dependent on the properties of the high-fidelity model and the low-fidelity model, and the concrete implementation as well. It is assumed that $0<\varepsilon<1$ and $0<\alpha$ and that the big $O$ notation refers to the worst-case time complexity estimate.

The authors in [39] provide a link between the approximation error and the complexity estimate which is adapted to the high-fidelity function approximation error
in (3.6) such that

$$
\begin{equation*}
\left\|\mathrm{K}-\tilde{\mathrm{K}}_{n}\right\|_{Y^{X}}=O\left(n^{-r / d}\right) \text { as } n \rightarrow \infty: \Leftrightarrow\left\|\mathrm{K}-\tilde{\mathrm{K}}_{n}\right\|_{Y^{X}} \in O\left(n^{-r / d}\right), \tag{3.73}
\end{equation*}
$$

where $r$ denotes the isotropic smoothness of the high-fidelity model and $O\left(n^{-r / d}\right)$ can be interpreted as the corresponding complexity class.

Even though the sparse grids technique is a tool to alleviate the curse of dimensionality to some extent, this technique is ignored in the present work. A reason is that the focus is on low-fidelity models associated with the space $P_{\leq 2}^{d}$ where, for a fixed $d$, the curse of dimensionality does not appear as heavily as for the space $P_{2}^{d}$ (cf. Table 3.1). By focusing on these spaces, a potential loss of accuracy is acceptable which is also partly due to the global character of these low-fidelity models.

Another tool to alleviate the curse of dimensionality to some extent are radial basis functions - with their local character - which we encounter next. Mind that, in favor of radial basis functions, a discussion on multivariate splines (see, e.g., [91, ch. 5.7]) is skipped. However, radial basis functions are partly related to splines (see, e.g., [62, p. 311]).

## Radial basis functions

Using a radial basis function as a low-fidelity model assumes that its corresponding hypothesis space (recall Definition 3.1.1) is a reproducing kernel Hilbert space $\mathrm{H}_{K}$. For an elaborate treatment of reproducing kernel Hilbert spaces, I refer to the literature (see, e.g., [187], [51, ch. 2.4], [91, ch. 5.8]).

Let us regard the space $H_{K}$ solely in the context of radial basis functions where a generic kernel $\phi=(x, t) \mapsto \phi(x, t): X \times X \rightarrow \mathbb{R}$ is specified as a radial kernel by setting $\phi=r \mapsto \phi(r): \mathbb{R}^{+} \rightarrow \mathbb{R}$ where $r:=\|x-t\|_{l_{2}}$ and $t \in X$ denotes a center point. Technically, it is tacitly assumed that the basis functions are radially symmetric on $\mathbb{R}^{d}$.

By invoking a function $\phi_{x}=t \mapsto \phi(x, t): X \rightarrow \mathbb{R}$, a member of the space $H_{K}$ is a function $\psi=x \mapsto \psi(x): X \rightarrow \mathbb{R}$ which can be portrayed as

$$
\begin{align*}
\psi(x) & :=c_{1} \phi_{x}\left(r_{1}\right)+\ldots+c_{n-1} \phi_{x}\left(r_{n-1}\right)+c_{n} \phi_{x}\left(r_{n}\right)  \tag{3.74a}\\
& \equiv \sum_{i=1}^{n} c_{i} \phi_{x}\left(r_{i}\right), \tag{3.74b}
\end{align*}
$$

where $r_{i}:=\left\|x-t_{i}\right\|_{l_{2}}$ with $i \in\{1, \ldots, n\}$ and it holds that $n \in \mathbb{N}$ and the coefficients $c_{i} \in \mathbb{R}$.
Analogously to (3.41), one can state that

$$
\begin{equation*}
\psi \in \operatorname{span}\left(\left\{\phi_{x}\left(r_{i}\right) \mid x \in X \wedge i \in\{1, \ldots, n-1, n\}\right\}\right) . \tag{3.75}
\end{equation*}
$$

Notice well that if there is no peril of confusion, let us reduce the amount of notation by omitting the subscript of $\phi_{x}$.

In Table 3.2, six different assignment definitions for a generic radial basis function $\phi=r \mapsto \phi(r): \mathbb{R}^{+} \rightarrow \mathbb{R}$ are provided. Their description follows the notational convention in [70, p. 46] and [116, p. 262], respectively; thus, let us call the parameter $\sigma$ the shape parameter. Furthermore, let us interpret the shape parameter as a hyperparameter (recall Remark 3.1.4). For more elaborations on smoothness properties or convergence properties of radial basis functions, I refer to, e.g., [38].

Despite favorable smoothness and convergence properties associated with radial basis functions, the selection of an appropriate radial basis functions for a task at

TABLE 3.2: Given a generic radial basis function $\phi=r \mapsto \phi(r)$ with
function signature $\mathbb{R}^{+} \rightarrow \mathbb{R}$, six different definitions for the assignment $\phi(r)$.

| Linear | Cubic | Thin plate spline |
| :---: | :---: | :---: |
| $r$ | $r^{3}$ | $r^{2} \log (r)$ |
| Gaussian | Multiquadratic | Inverse multiquadratic |
| $e^{-\frac{r^{2}}{2 \sigma^{2}}}$ | $\left(r^{2}+\sigma^{2}\right)^{\frac{1}{2}}$ | $\left(r^{2}+\sigma^{2}\right)^{-\frac{1}{2}}$ |

hand is heavily problem-dependent such that a heuristic approach to the selection is common.

In Figure 3.6, I illustrate the six different radial basis function definitions from Table 3.2. For those radial basis functions involving the shape parameter $\sigma$, I sketch the members of the corresponding family where $\sigma \in\{0.2,0.4,0.6\}$.


Figure 3.6: The six radial basis functions in Table 3.2.
(i) Linear, (ii) Cubic, (iii) Thin plate spline
(iv) Gaussian, (v) Multiquadratic, (vi) Inverse multiquadratic.

If we identify the number of basis functions $n$ with the number of sampling plan points, that is, $n \equiv m$, one can execute a determination of the coefficients in (3.74) by interpolation (cf. (3.18)). A common choice regarding the center points $t_{i}$ is to identify the center points with the sampling plan points $x_{i}$, that is, $t_{i} \equiv x_{i}$. The floatingpoint arithmetic operations complexity is $O\left(n^{3}\right)$, the storage costs are $O\left(n^{2}\right)$, and the evaluation costs are $O(n)$ (see, e.g., [181], [19], [63]). Since there is no particular sparsity pattern associated with the corresponding interpolation matrix that could be exploited during the solving, the solving's arithmetic complexity is in the same class as the Gaussian elimination algorithm's arithmetic complexity of $O\left(n^{3}\right)$.

If the number of basis functions becomes too large, for instance, $n \gg 1 \times 10^{4}$, then this low-fidelity model becomes impractical. However, since the number of sampling plan points is kept rather small, i.e., $m \ll 1 \times 10^{4} \Longrightarrow n \ll 1 \times 10^{4}$, let us not dwell on schemes for the evaluation of radial basis functions. For more details on this active area of research, I refer to, e.g., [181], [19], [63].

By definition, the corresponding interpolation matrix - or, more suited to the given context, the Gram matrix - is for all radial basis functions in Table 3.2 at least positive semi-definite (see, e.g., [187], [116, p. 262]). In the case of the Gaussian radial basis function, the corresponding interpolation matrix is even positive definite (see, e.g., [187], [70, p. 46]) such that, given the output points, it is guaranteed that a coefficient column vector exits. Nevertheless, for all radial basis functions, there is still a governing trade-off principle or uncertainty principle which states that if we increase the accuracy, e.g., by increasing the number of sampling plan points $m$ or by increasing the shape parameter $\sigma$, then the condition number - as an indicator for numerical stability - grows as well (see, e.g. [62, ch. 16], [187]). Note that using a sampling plan such as the Sobol quasi-random sequence can have a moderately beneficial influence on the condition number (see, e.g., [34]).

Observe that if we make the choice $n<m$, more precisely, if we do not take every sampling plan point as a center point, then one can execute a determination of the coefficients in (3.74) by regression - analogously to (3.62). However, to my best knowledge, there is no complete theory to explain the optimal selection of sampling plan points as center points in the case of regression (see, e.g., [62, p. 168ff], [70, p. 49]). Due to some relationship between radial basis functions and splines (see, e.g., [91, p. 36]), the center point selection problem is mostly solved by heuristic selection - similarly to the heuristic approach to the knot selection problem in multivariate spline regression.

For the sake of completeness, I mention briefly two kinds of extensions of the low-fidelity model in (3.74) which are discussed in the literature for reasons such as ensuring well-posedness, increasing accuracy, and the like.

The first kind of extension is related to the thin plate spline and the radial powers - such as the linear or the cubic - (see Table 3.2). More precisely, the low-fidelity model $\psi$ in (3.74) is extended linearly by a $d$-variate polynomial $p$ of total degree at most one, i.e., $p \in \mathrm{P}_{\leq 1}^{d}$ and, usually, in a monomial basis setting, such that one can represent an extended radial basis low-fidelity model $\hat{\psi}$ as

$$
\begin{equation*}
\hat{\psi}(x):=p(x)+\psi(x), \tag{3.76}
\end{equation*}
$$

where the signature of the operation + is $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. For more details on the determination of the coefficients of $\hat{\psi}(x)$ and for some applications of $\hat{\psi}(x)$ in engineering, I refer to [27] and references therein. ${ }^{16}$

The second kind of extension is related to a statistical setting (see, e.g., [61, p. 177]), that is, the low-fidelity model $\psi$ in (3.74) is extended linearly by a $d$-variate polynomial $\hat{\mu}$ of total degree at most zero, i.e., $\hat{\mu} \in \mathrm{P}_{\leq 0}^{d}$ such that one can define an extended radial basis low-fidelity model $\hat{\psi}$ as

$$
\begin{equation*}
\hat{\psi}(x):=\hat{\mu}(x)+\psi(x), \tag{3.77}
\end{equation*}
$$

[^27]where the term $\hat{\mu}(x)$ can be interpreted as the estimate of the mean of the highfidelity model. In order to discuss the relationship between $\psi(x)$ and $\hat{\psi}(x)$, one would have to dwell on affine spaces which we forgo since, frequently, the term $\hat{\mu}(x)$ is considered as incorporated in the coefficients and in the basis functions of $\psi(x)$ (see, e.g., [91, p. 11f]). Otherwise, the determination of the coefficients column vector in (3.62) has to be adapted such that
\[

$$
\begin{equation*}
\hat{c}:=B^{+}\left(y-\underline{\mu}_{y}\right), \tag{3.78}
\end{equation*}
$$

\]

where $B^{+}$has to be customized to the corresponding radial basis function and $\underline{\mu}_{y}$ denotes the constant unknown mean column vector with respect to $y$. Let us postpone, though, further discussions on the expression in (3.77) to the elaboration on stochastic interpolation via kriging low-fidelity models.

## Kriging

Supposing a sampling plan $X_{s}$ such as in (3.14), a kriging low-fidelity model can be interpreted as an extended radial basis low-fidelity model $\hat{\psi}$ such as in (3.77) (see, e.g., [70, p. 60] or [109]), that is,

$$
\begin{equation*}
\hat{\psi}(x):=\hat{\mu}(x)+\sum_{i=1}^{n} c_{i} \phi_{x}\left(r_{i}\right), \tag{3.79}
\end{equation*}
$$

where $n \equiv m$, i.e., all sampling plan points are utilized as center points. Following the Gaussian in Table 3.2, the assignment $\phi_{x}\left(r_{i}\right)$ in (3.79) is commonly chosen as

$$
\begin{equation*}
\phi_{x}\left(r_{i}\right):=\exp \left(-r_{i}\right), \tag{3.80}
\end{equation*}
$$

where the radii $r_{i}$ are not defined via an $l_{2}$-norm such as in (3.74), but via a metric such that $r_{i}$ can be written as

$$
\begin{equation*}
r_{i}:=\sum_{j=1}^{d} \theta_{j}\left|x^{j}-x_{i}^{j}\right|^{p_{j}}, \tag{3.81}
\end{equation*}
$$

where $x_{i}^{j}$ refers to the $j$-th component of the $i$-th sampling plan point, $d$ refers to the total number of components of the $i$-th sampling plan point, i.e., $d \equiv N_{\xi}$, and $\theta_{j} \in \mathbb{R}^{+}$and $p_{j} \in[0,2]$ refer to parameters that have to be determined. Mind that the coefficients $c_{i}$ in (3.79) depend on the parameters $\theta_{j}, p_{j}$, and $\hat{\mu}(x)$. These unknown quantities are determined by means of a statistical machinery which we sketch out next.

Building upon (3.20), an approach to a kriging low-fidelity model is to consider

$$
\begin{equation*}
\forall i \in\{1, \ldots, m\} . z\left(x_{i}\right)=y_{i}-\mu\left(x_{i}\right), \tag{3.82}
\end{equation*}
$$

where $\mu$ indicates a constant unknown mean function with respect to $y$ and the residual $y_{i}-\mu\left(x_{i}\right)$ indicates a realization of a Gaussian process $z(x)$ (cf. [61, p. 146]). Note that we only consider a constant unknown mean function. This choice of $\mu$ is associated with the so-called ordinary kriging. For other kinds of kriging, I refer to [221, p. 154].

From a statistical viewpoint, $z(x)$ is associated with a random error (recall the discussion on noise in § 3.1.1). However, from an interpolation viewpoint, we do not regard any errors in the output points $y_{i}$ (cf. [70, p. 55]).

Therefore, the sketch is inspired mainly by the approach to a kriging low-fidelity model presented in [108] and in [70, ch. 2.4]. For more details regarding the sketch, I refer to, e.g., [184], [108], [61, ch. 5.4], [173, ch. 5], [70, ch. 2.4] or [116, ch. 15].

With regard to a given sampling plan $X_{s}$, let us encode in matrix representation the output points $y_{i}$ as a column vector $y \in \mathbb{R}^{m \times 1}$ with

$$
y:=\left[\begin{array}{lllll}
y_{1} & y_{2} & \cdots & y_{m-1} & y_{m} \tag{3.83}
\end{array}\right]^{\mathrm{T}} .
$$

Furthermore, with regard to a given sample, one can define the probability density of an $m$-dimensional Gaussian distribution at $y$ as

$$
\begin{equation*}
\mathcal{N}_{m}\left(y \mid \underline{y}_{y^{\prime}}, \Sigma\right):=\frac{1}{(2 \pi)^{m / 2}|\Sigma|^{1 / 2}} \exp \left(-\frac{1}{2}\left(y-\underline{-}_{y}\right)^{\mathrm{T}} \Sigma^{-1}\left(y-\underline{-}_{y}\right)\right), \tag{3.84}
\end{equation*}
$$

where, technically, $y$ is associated with a corresponding random vector $\underline{Y}, \Sigma \in \mathbb{R}^{m \times m}$ denotes the covariance matrix and $\underline{\mu}_{y} \in \mathbb{R}^{m \times 1}$ denotes the constant unknown mean column vector with respect to $y$ that, given a scalar $\mu_{y} \in \mathbb{R}$, is defined as

$$
\begin{align*}
\underline{\mu}_{y} & :=\mu_{y} \cdot\left[\begin{array}{lllll}
1 & 1 & \ldots & 1 & 1
\end{array}\right]^{\mathrm{T}}  \tag{3.85}\\
& \equiv \mu_{y} \cdot \underline{1}, \tag{3.86}
\end{align*}
$$

where $\underline{1}:=\left[\begin{array}{lllll}1 & 1 & \ldots & 1 & 1\end{array}\right]^{\mathrm{T}}$ with $\underline{1} \in \mathbb{R}^{m \times 1} .{ }^{17}$ Observe that $\mathcal{N}_{m}\left(y \mid \underline{\mu}_{y^{\prime}}, \Sigma\right)$ in (3.84) is a slight abuse of notation in order to emphasize the sample-oriented viewpoint in the present work (recall § 3.1.1).

If we consider the constant unknown variance with respect to $y$, i.e., $\sigma_{y}^{2} \in \mathbb{R}$, that is associated with the constant unknown standard deviation w.r.t. $y$, i.e., $\sigma_{y} \in \mathbb{R}$, by

$$
\begin{equation*}
\sigma_{y}:=\sqrt{\sigma_{y}^{2}}, \tag{3.87}
\end{equation*}
$$

then the covariance matrix can be expressed by the correlation matrix $\Psi \in \mathbb{R}^{m \times m}$ such that

$$
\begin{equation*}
\Sigma \equiv \sigma_{y}^{2} \cdot \Psi . \tag{3.88}
\end{equation*}
$$

In the context of a kriging low-fidelity model, the entries of the correlation matrix $\Psi:=\left[\psi_{i, l}\right]$ are commonly defined by the radial basis function in (3.80) such that

$$
\begin{equation*}
\psi_{i, l} \equiv \exp \left(-\sum_{j=1}^{d} \theta_{j}\left|x_{i}^{j}-x_{l}^{j}\right|^{p_{j}}\right) . \tag{3.89}
\end{equation*}
$$

The choice of entries in (3.89) reveals that if $i=l$, then $\psi_{i, l}=1$, and if $\left\|x_{i}-x_{l}\right\|_{l_{2}}$ grows exponentially, then $\psi_{i, l}$ tends asymptotically to zero.

Instead of a parametrization by its mean vector and its covariance matrix, one can parameterize the $m$-dimensional Gaussian distribution in (3.84) by its mean vector, its variance scalar, and its correlation matrix such that (3.84) can be rewritten as

$$
\begin{equation*}
\mathcal{N}_{m}\left(y \mid \underline{-}_{y^{\prime}}, \sigma_{y}^{2}, \Psi\right):=\frac{1}{\left(2 \pi \sigma_{y}^{2}\right)^{m / 2}|\Psi|^{1 / 2}} \exp \left(-\frac{1}{2 \sigma_{y}^{2}}\left(y-\underline{\mu}_{y}\right)^{\mathrm{T}} \Psi^{-1}\left(y-\underline{-}_{y}\right)\right), \tag{3.90}
\end{equation*}
$$

[^28]where it holds that $\forall \Psi \in \mathbb{R}^{m \times m} . \forall \sigma_{y}^{2} \in \mathbb{R} \cdot \operatorname{det}\left(\sigma_{y}^{2} \Psi\right)=\left(\sigma_{y}^{2}\right)^{m} \operatorname{det}(\Psi)$.
Due to the definition of the entries of $\Psi$ in (3.89), the matrix $\Psi$ is positive definite and all of its eigenvalues are positive, respectively. Therefore, the matrix $\Psi$ is non-singular and the inverse matrix $\Psi^{-1}$ exists, respectively. Furthermore, one can discern that $|\Psi|>0 .{ }^{18}$

Abstractly, one can define the likelihood function $L=\vartheta \mapsto L(\vartheta) \equiv L(y \mid \vartheta)$ with the signature $\Theta \rightarrow[0,1]$ and $\vartheta:=\left(\mu_{y}, \sigma_{y}^{2}\right)$ such that the assignment $L(\vartheta)$ reads as

$$
\begin{equation*}
L(\vartheta):=\frac{1}{\left(2 \pi \sigma_{y}^{2}\right)^{m / 2}|\Psi|^{1 / 2}} \exp \left(-\frac{1}{2 \sigma_{y}^{2}}\left(y-\underline{\mu}_{y}\right)^{\mathrm{T}} \Psi^{-1}\left(y-\underline{\mu}_{y}\right)\right) \tag{3.91}
\end{equation*}
$$

Roughly speaking: Given any $y$, the aim is to find the parameter $\vartheta$ such that the likelihood of observing $y$ is maximized. ${ }^{19}$ Hence, the maximum likelihood estimate (MLE) for $\vartheta$, i.e., $\hat{\vartheta}_{M L E}$ or $\hat{\vartheta}$, is characterized by

$$
\begin{equation*}
\hat{\vartheta}:=\underset{\vartheta \in \Theta}{\arg \max } L(\vartheta) . \tag{3.92}
\end{equation*}
$$

Though, computationally more amenable regarding (3.92) is to consider the lnlikelihood function $L_{\ln }=\vartheta \mapsto \ln (L(\vartheta))$ with the signature $\left.\left.\Theta \rightarrow\right]-\infty, 0\right]$ such that the assignment $L_{\ln }(\vartheta)$ reads as

$$
\begin{equation*}
L_{\ln }(\vartheta):=-\frac{m}{2} \ln (2 \pi)-\frac{m}{2} \ln \left(\sigma_{y}^{2}\right)-\frac{1}{2} \ln (|\Psi|)-\frac{1}{2 \sigma_{y}^{2}}\left(y-\underline{\hat{\mu}}_{y}\right)^{\mathrm{T}} \Psi^{-1}\left(y-\underline{\hat{\mu}}_{y}\right) \tag{3.93}
\end{equation*}
$$

where, due to the definition of the natural logarithm, one has to suppose that $\sigma_{y}^{2} \geq 0$ and $|\Psi| \geq 0$.

The maximum likelihood estimates for $\mu_{y}$ and $\sigma_{y}^{2}$ can be described by

$$
\begin{align*}
& \hat{\mu}_{y}:=\frac{\underline{1}^{\mathrm{T}} \Psi^{-1}}{\underline{1}^{\mathrm{T}} \Psi^{-1} \underline{1}} y,  \tag{3.94a}\\
& \hat{\sigma}_{y}^{2}:=\frac{1}{m}\left(y-\hat{\mu}_{y}\right)^{\mathrm{T}} \Psi^{-1}\left(y-\hat{\mu}_{y}\right), \tag{3.94b}
\end{align*}
$$

where $\underline{\underline{\mu}}_{y}$ is defined analogously to (3.85), that is,

$$
\begin{equation*}
\underline{\underline{\mu}}_{y}:=\hat{\mu}_{y} \cdot \underline{1} \tag{3.95}
\end{equation*}
$$

By evaluating the ln-likelihood function in (3.93) at the estimates in (3.94) and truncating those terms of the assignment $L_{\ln }(\vartheta)$ that represent solely numbers, one can define the concentrated ln-likelihood function $L_{c \ln }=(\theta, p) \mapsto L_{c \ln }(\theta, p)$ with the signature $\left[0,+\infty\left[{ }^{d} \times[0,2]^{d} \rightarrow\right]-\infty, 0\right]$ such that the assignment $L_{c} \ln (\theta, p)$ reads as

$$
\begin{equation*}
L_{c \ln }(\theta, p):=-\frac{m}{2} \ln \left(\hat{\sigma}_{y}^{2}\right)-\frac{1}{2} \ln (|\Psi|) . \tag{3.96}
\end{equation*}
$$

In order to determine the maximum likelihood estimates of $(\theta, p)$ numerically by utilizing a suitable optimization algorithm (recall $\S 2.3 .3$ ), it is common to associate

[^29]$(\hat{\theta}, \hat{p})$ with the expression
\[

$$
\begin{equation*}
(\hat{\theta}, \hat{p}):=\underset{(\theta, p) \in\left[0,+\infty\left[\left[^{d} \times[0,2]^{d}\right.\right.\right.}{\arg \min }-L_{c} \ln (\theta, p) . \tag{3.97}
\end{equation*}
$$

\]

In [70, p. 55-58], the authors mention further aspects regarding the numerical treatment of the expression in (3.97). For instance, due to the definition of the entries of the correlation matrix $\Psi$ in (3.89), the maximum likelihood estimates $(\hat{\theta}, \hat{p})$ are sensitive to the scaling of a given sampling plan $X_{s}$. Therefore, it is advisable to consider the normalized sampling plan $X_{\mathrm{s}}$, that is, the unit $d$-dimensional hypercube (recall the case $d=2$ in Figure 3.1 and Figure 3.2).

Furthermore, it is preferable to consider the entity $\theta$ rather on a closed logarithmic interval such as $\theta \in\left[10^{-3}, 10^{2}\right]^{d}$.

Additionally, it is common to alleviate the computational burden in (3.97) by setting heuristically the entity $\hat{p}$ to a fixed value in advance. A usual choice is $\hat{p}=\left[\begin{array}{lllll}2 & 2 & \ldots & 2 & 2\end{array}\right]^{\mathrm{T}}$ with $\hat{p} \in \mathbb{R}^{d \times 1}$.

It should be recalled that the kriging low-fidelity models are closely related to Gaussian radial basis functions. Hence, it is reasonable to assume that the kriging low-fidelity models suffer from ill-conditioning issues as well which could be mitigated by simple regularization techniques (see the discussion concerning (3.65)) such that, for instance, instead of the correlation matrix $\Psi$, the regularized correlation matrix $\Psi+\lambda I$ is considered with $\lambda \in\left[\epsilon, 1 \times 10^{-6}\right]$ and $I \in \mathbb{R}^{m \times m}$. Hence, the matrix $\Psi+\lambda I$ has to be take into account in (3.94) and in (3.96) as well. Technically, the regularized correlation matrix implicitly supposes a regression problem as opposed to an interpolation problem. Pragmatically, due to regarding the range of values for the hyperparameter $\lambda$ as marginal, this specific regularized correlation matrix is still treated within an interpolation problem (see, e.g., [70, p. 152]).

It is also reasonable to assume that the kriging low-fidelity models exhibits a similar character regarding the floating-point arithmetic operations complexity, the storage costs, and the evaluation costs such as the Gaussian radial basis functions. Thus, it possible to exploit the matrix structure of the correlation matrix $\Psi$ which is a square positive-definite matrix. More precisely, a matrix inversion based on Cholesky decomposition can be performed in order to reduce the number of floating point arithmetic operations compared to a lower-upper (LU) decomposition.

Notice well that if we apply a singular value decomposition (SVD) to the correlation matrix $\Psi$, then we obtain a least-squares Kriging regression (cf. [70, p. 152]). The determination of the inverse $\Psi^{-1}$ of a non-singular matrix $\Psi$ in SVD can be seen as computationally equivalent to the determination of the pseudoinverse $\Psi^{+}$of the matrix $\Psi$ by the SVD method such as in (3.61).

After the determination of the maximum likelihood estimates $(\hat{\theta}, \hat{p})$, one can specify the kriging low-fidelity model in (3.79) as

$$
\begin{equation*}
\hat{y}(x):=\hat{\mu}_{y}+r^{\mathrm{T}} \Psi^{-1}\left(y-\hat{\underline{\hat{n}}}_{y}\right), \tag{3.98}
\end{equation*}
$$

where $\hat{y}: X \rightarrow \mathbb{R}$ such that $\hat{y}(x)$ indicates the prediction ${ }^{20}$ at an arbitrary point $x$ and $r:=\left[r_{i}\right] \in \mathbb{R}^{m \times 1}$ denotes the correlation column vector that reads as

$$
r:=\left[\begin{array}{lllll}
r_{1} & r_{2} & \ldots & r_{m-1} & r_{m} \tag{3.99}
\end{array}\right]^{\mathrm{T}},
$$

[^30]where the components $r_{i}$ are defined such as in (3.81). For an in-depth derivation of the maximum likelihood estimate $\hat{y}(x)$, I refer to [70, p. 59-62].

### 3.1.3 Simplified-physics low-fidelity models

Simplified-physics low-fidelity models depend on a user's domain-specific knowledge regarding the mathematical description of the physics associated with the highfidelity model and regarding the numerical software associated with the high-fidelity model (recall chapter 2).

Depending on the degree of intervention in the implementation of the numerical software regarding the high-fidelity model, the low-fidelity models are intrusive or non-intrusive. Let us consider all simplified-physics low-fidelity models as nonintrusive, especially if a low-fidelity model is based on, e.g., a coarse-grid discretization or a weakened termination criteria of an iterative solver or a combination of both.

Recalling $\S$ 3.1.1, a basic postulate concerning a high-fidelity model K and a lowfidelity model $\tilde{\mathrm{K}}$ is that $\mathrm{K} \in Y^{X}$ and $\tilde{\mathrm{K}} \in Y^{X}$. Unlike the deterministic and probabilistic data-fit low-fidelity models, one cannot generally provide a hypothesis space $H$. Furthermore, the computational costs and the degrees of fidelity linked to the lowfidelity models under consideration are prescribed by the user who, abstractly speaking, imposes implicitly some kind of lexicographic ordering or lexicographic preference on the class that encompasses all models.

Besides the low-fidelity model based on, e.g., a coarse-grid discretization, another example of a simplified-physics low-fidelity model is a one-dimensional, linear boundary value problem (1D-LBVP) that is, in some sense, related to a twodimensional, linear boundary value problem (2D-LBVP). Notice well that the twodimensional, linear boundary value problem can be seen as a simplified-physics low-fidelity model which in turn, in some sense, is related to a three-dimensional, non-linear boundary value problem (3D-NLBVP).

Hence, one can construct a hierarchy of low-fidelity models where the highfidelity model corresponds to a 3D-NLBVP. In Figure 3.7, there is a schematic depiction of such a possible user-prescribed hierarchy. ${ }^{21}$

If we invoke the Figure 2.1a, then one can concretize the Figure 3.7 with regard to a user-prescribed hierarchy of magnetoquasistatic and magnetostatic problems, respectively, by means of the Figure 3.8.

In (i) of Figure 3.8, a single conducting subdomain as a common representative of a magnetoquasistatic subsystem's domain of application is depicted. In order to emphasize the subdomain's three-dimensionality, let us utilize the superscript $3 D$.

In (ii) of Figure 3.8, a single conducting subdomain exhibiting two-dimensionality $(2 D)$ is shown. The cross sectional area indicated by $\Omega_{\mathrm{nc}}$ can be regarded as topologically equivalent to the closed 2-ball that is topologically equivalent to the closed unit 2 -cube $[0,1]^{2}$. Hence, the cross-sectional area's boundary $\partial \Omega_{\mathrm{nc}}$ can be seen as topologically equivalent to the 1 -sphere that is topologically equivalent to the boundary of the closed unit 2 -cube $[0,1]^{2}$. In applications, assuming an appropriate metric structure, a round and a rectangular cross-sectional area are commonly utilized for geometrically modeling a round conductor and a foil conductor, respectively. These conductor kinds are usually the building blocks of an inductive components winding of varying complexity. For instance, the round conductor constitutes a basic building block of a litz wire winding (see, e.g., [154, p. 110-113]).

[^31](a)

(b)


Figure 3.7: A schematic depiction of a user-prescribed hierarchy of problems which are associated with simplified-physics low-fidelity models. The problem 3D-NLBVP is associated with the high-fidelity model. (a) An arrangement of the user-prescribed hierarchy with regard to the degree of fidelity and the computational costs. (b) An encoding of the user-prescribed hierarchy as a relationships diagram in which the arrow points from a model with higher degree of fidelity and higher computational costs to a model with lower degree of fidelity and lower computational costs.

A corresponding two-dimensional boundary value problem can be associated with a three-dimensional boundary value problem where the cross-sectional area is assumed to be spatially longitudinally homogeneous.

In (iii) of Figure 3.8, an ohmic resistor from electric circuit components is depicted as a symbolic representation of a function $R$. In applications, the function $R$ is often associated with special mathematical functions such as the natural logarithm or Bessel functions. Prevalently though, the function $R$ is associated with a multivariate rational function.

For instance, a conductor's ohmic resistance at the frequency 0 Hz can be expressed as a multivariate rational function depending on a parameter point $\xi$ (recall § 2.2.3) in which geometrical parameters are incorporated that define the conductor's length and its cross-sectional area. Assuming a spatially constant electric conductivity (recall $\S 2.1 .2$ ), that is, $\sigma(\mathbf{x}):=\sigma_{0}$ with $\sigma_{0} \in \mathbb{R}^{+}$, then, technically, one could include the material parameter $\sigma_{0}$ in the parameter point $\boldsymbol{\xi}$ as well. However, a good conductor is usually assumed in the sense that the material characteristics of plain copper are utilized such that $\sigma_{0}$ is fixed as $\sigma_{0}:=\sigma_{\mathrm{Cu}}$ with $\sigma_{\mathrm{Cu}}:=5.96 \times 10^{7} \mathrm{~S} / \mathrm{m}$.

Another example is an impedance which can be interpreted as a representative of a real inductive component in a circuit theory context. This impedance can be used, e.g., in the computation of a two-port S-parameter matrix. The components of the parameter point $\boldsymbol{\xi}$ adhere to the physical interpretation as an angular frequency, a capacitance, an inductance, and an electrical resistance.

Abstractly, one can state that $R \in{ }_{q}^{p} \mathrm{P}_{(m, n)}^{d}$ where, analogous to (3.50), the space ${ }_{q}^{p} \mathrm{P}_{(m, n)}^{d}$ denotes the space of $d$-variate rational polynomials of total degree at most $m$ in the numerator polynomial $p$, i.e., $p \in \mathrm{P}_{\leq m}^{d}$ and total degree at most $n$ in the denominator polynomial $q$, i.e., $q \in P_{\leq n}^{d}$. Hence, the function $R$ can be called a multivariate rational polynomial function of type ( $m, n$ ) with $m, n \in \mathbb{Z}_{0}^{+}$. Finally, one can read the


FIGURE 3.8: A schematic depiction of a user-prescribed hierarchy of magnetoquasistatic and magnetostatic problems which are associated with simplified-physics low-fidelity models. The three-dimensional domains in (i) are associated with the high-fidelity model's underlying magnetoquasistatic or magnetostatic problem. The twodimensional domains in (ii) are associated with a low-fidelity model's underlying magnetoquasistatic or magnetostatic problem. The ohmic resistors in (iii) are associated with a multivariate rational polynomial derived from a magnetoquasistatic or a magnetostatic problem. It is assumed that the user prefers the low-fidelity models in (ii) over those low-fidelity models in (iii).
assignment of the ( $m, n$ ) multivariate rational function $R$ as

$$
\begin{equation*}
R(x):=\frac{p(x)}{q(x)} \text { such that } p \in \mathrm{P}_{\leq m}^{d}, q \in \mathrm{P}_{\leq n}^{d} . \tag{3.100}
\end{equation*}
$$

Unlike deterministic data-fit low-fidelity models, the function $R$ is derived from the system of Maxwell's equations (recall $\S$ 2.1.2) and its coefficients are fixed to known values. It might be beneficial to examine the usefulness of multivariate rational polynomials as deterministic data-fit low-fidelity models, however, I ignore them in the present work and they - as well as associated methods within the electromagnetics context such as vector fitting (see, e.g., [85]) - are left for future investigations.

Notice well that the rationale for ignoring them is driven by potential difficulties in handling properly spurious poles of a multivariate rational polynomial in an optimization context. ${ }^{22}$ Additionally, it is supposed that the potential benefits of a multivariate rational polynomial's localized behavior steered by the poles are comparable with a radial basis function's localized behavior steered by the choice of the center points such that the radial basis functions are preferred over the multivariate rational polynomials. A thoroughly elaborated juxtaposition of these two kinds of deterministic data-fit low-fidelity models is out of the scope of this work, though.

Notice well that the dotted arrows in Figure 3.8 are semantically overloaded in the sense that their vertical reading and their horizontal reading differ.

Considering the first level and the second level within (i) of Figure 3.8, the dotted arrows indicate a relationship between a fine-grid discretization and a coarse-grid discretization, more precisely, the respective simplicial triangulations (recall § 2.2.2) $\mathcal{T}_{h_{1}}^{3 D}$ and $\mathcal{T}_{h_{2}}^{3 D}$ are governed by the characteristic $h_{1}<h_{2}$. Furthermore, considering the second and the third level, the dotted arrows indicate a relationship between a higher threshold ${ }^{1} \Delta_{A x=b}^{3 D}$ and a lower threshold ${ }^{2} \Delta_{A x=b}^{3 D}$ for a termination criterion of an iterative solver, more precisely, the thresholds exhibit the characteristic ${ }^{1} \Delta_{A x=b}^{3 D}<{ }^{2} \Delta_{A x=b}^{3 D}$ with ${ }^{1} \Delta_{A x=b}^{3 D} \in \mathbb{R}^{+}$and ${ }^{2} \Delta_{A x=b}^{3 D} \in \mathbb{R}^{+}$.

The explanation for the levels within (ii) of Figure 3.8 is analogous to the previous one. By contrast, the dotted arrows for the levels within (iii) of Figure 3.8 hint at a change such as the domain transformations in (A.1) or in (A.14). Or the arrows hint at a change, for instance, from a multivariate rational polynomial function of type ( $m, n$ ) to a multivariate rational polynomial function of type $(m, 0)$ and leading coefficient of one, that is, to a multivariate polynomial function from the space $\mathrm{P}_{\leq k}^{d}$.

The horizontal reading of the dotted arrows in Figure 3.8 reflects the relationships diagram within (b) of Figure 3.7. In the vertical reading, it is partially conceivable how a formal encoding of the arrows could look like, but it is not straightforward to conceive such a formal encoding with regard to the horizontal reading. If we employ a structural perspective to the Figure 3.8 - similarly to the structural perspectives in ch. 2 , then one can extract exemplarily the formal encodings by a map-oriented representation in (Diagrams of Fig. 3.8) where, for the sake of clarity, it is omitted to extract the inverse maps from the Figure 3.8.

From (Diagrams of Fig. 3.8), one can conclude that, theoretically, a less preferred problem can be constructed by a composition of maps associated with more preferred problems. Then, one can make statements such as

$$
\begin{gather*}
l_{1} \circ f_{3} \circ f_{2} \circ f_{1}=g_{3} \circ g_{2} \circ g_{1}  \tag{3.101a}\\
l_{1} \circ f_{3} \circ f_{2} \circ f_{1}=g_{3} \circ g_{2} \circ g_{1} \circ i_{1}  \tag{3.101b}\\
l_{2} \circ l_{1} \circ f_{3} \circ f_{2} \circ f_{1}=h_{3} \circ h_{2} \circ h_{1} \circ i_{2} \circ i_{1} . \tag{3.101c}
\end{gather*}
$$

Note that the statements in (3.101) are valid under the assumption that all maps are set functions and their domains and co-domains are sets. However, this assumption does not take into account adequately the different algebraic characters of, e.g., $\mathcal{T}_{h_{1}}^{3 D}$,

[^32]${ }^{1} \Delta_{A x=b}^{3 D}$ and $R_{1}$.


(Diagrams of Fig. 3.8)

Moreover, the statements in (3.101) do not adequately capture the idea that, for instance, moving from a high-fidelity model's underlying three-dimensional magnetoquasistatic problem to a low-fidelity model's underlying two-dimensional magnetoquasistatic problem corresponds technically to a loss of problem information, e.g., with regard to the boundary conditions. Therefore, it is more appropriate to consider the problems associated with the low-fidelity models as forgetful interpretations of the problem associated with the high-fidelity model. This viewpoint can be mediated by, for example, the maps $F_{1}$ and $F_{2}$ such that

$$
\begin{align*}
F_{1}\left(\mathcal{T}_{h_{1}}^{3 D}\right):=\mathcal{T}_{h_{1}}^{2 D} & F_{1}\left(\mathcal{T}_{h_{2}}^{3 D}\right):=\mathcal{T}_{h_{2}}^{2 D} & F_{1}\left(f_{1}\right):=g_{1}  \tag{3.102a}\\
\left(F_{2} \circ F_{1}\right)\left(\mathcal{T}_{h_{1}}^{3 D}\right):=R_{1} & \left(F_{2} \circ F_{1}\right)\left(\mathcal{T}_{h_{2}}^{3 D}\right):=R_{2} & \left(F_{2} \circ F_{1}\right)\left(f_{1}\right):=h_{1}, \tag{3.102b}
\end{align*}
$$

where the map $F_{1}$ and the map $F_{2}$ are overloaded in order to deal with the different algebraic characters. We elaborate on the corresponding formal approach based on the category theoretical language in chapter 4.

To adjust the expectations correctly, notice well that the category theoretical language is not a panacea at all. Its merits stem from the fact that, in a nutshell, there is an absence of a cohesive theory to express in formal terms the relationships between different problems associated with a high-fidelity model and corresponding low-fidelity models (see Figure 3.8).

This absence, though, is the Achilles' heel of the mathematical analysis of any optimization approach that exploits simplified-physics low-fidelity models and relies on a, in some sense, benign resemblance between these low-fidelity models and the high-fidelity model (see, e.g., [49, p. 76]).

In order to mitigate the ramifications of the absence, the authors in [121] suggest to assess the resemblance based on an observed points subset, i.e., the training subset, and some quality factors that are derived in the context of the space mapping paradigm.

In a more general context, the author in [201, p. 76] suggests the assessment and ordering of different problems by analyzing their explanatory power based on an
observed points subset in a Bayesian setting.
Mind that these approaches are not pursued in the present work, although these approaches are promising endeavors towards a quantification of the relationships between different problems associated with the various models.

However, these approaches do not seem widely adopted in practical applications - because, presumably, the user-prescribed hierarchy of problems, which reflects the user's preferences, outranks other conceivable orderings of the problems.

Nevertheless, in § 3.3.2, I contribute partly to this overall discussion by elaborating briefly on the potential role of the NREGE in (3.24) and the SSPCC in (3.31) regarding the quantitative assessment of the quality of a low-fidelity model and a surrogate model within the space-mapping paradigm.

In § 3.3.1, we discuss the efficient global optimization (or sequential kriging optimization) technique as a subtype of the model management strategy adaptation. This technique exploits solely a kriging low-fidelity model.

In $\S 3.3 .2$ and in $\S 3.3 .2$, we discuss the space mapping paradigm and the cokriging approach as subtypes of the two model management strategies adaptation and fusion, respectively (recall § 1.3). Both the space mapping paradigm and the cokriging approach are designed such that they exploit especially simplified-physics low-fidelity models.

A notable distinction between the space mapping paradigm and the co-kriging approach is how they deal with the statement in (3.2). Abstracting from the authors' perspective in [70, p. 167], it can be argued that the co-kriging approach focuses on an instance of the generic statement

$$
\begin{equation*}
\forall x \in X . \mathrm{K}(x)={ }_{\gamma} \mathbf{Z}_{\rho}(x) \cdot{ }_{\gamma} \tilde{\mathrm{K}}(x)+_{\gamma} \mathbf{Z}_{\Delta}(x), \tag{3.103}
\end{equation*}
$$

where $Z_{\rho} \in Y^{X}$ and $Z_{\Delta} \in Y^{X}$ denote correction maps and the map $\cdot \gamma: Y \times Y \rightarrow Y$ and the map $+_{Y}: Y \times Y \rightarrow Y$ denote a suitable multiplication on $Y$ and a suitable addition on $Y$, respectively. We elaborate on the instance associated with the co-kriging approach in §3.3.2.

Taking into account in an abstract manner the stance of the authors in [194, p. 32f], [56, ch. 2.5], and [49, p. 110ff], it can be argued that the space-mapping paradigm focuses on instances of the generic statements

$$
\begin{align*}
& \forall x \in X . \mathrm{K}(x)==_{Y}\left(\tilde{\mathrm{~K}} \circ_{X} \tilde{\mathrm{P}}\right)(x),  \tag{3.104a}\\
& \forall x \in X . \mathrm{K}(x)==_{Y}\left(\tilde{\mathrm{R}} \circ_{Y} \tilde{\mathrm{~K}}\right)(x),  \tag{3.104b}\\
& \forall x \in X . \mathrm{K}(x)={ }_{\gamma}\left(\tilde{\mathrm{R}} \circ_{Y} \tilde{\mathrm{~K}} \circ_{X} \tilde{\mathrm{P}}\right)(x), \tag{3.104c}
\end{align*}
$$

where $\tilde{P} \in X^{X}$ denotes a domain-oriented correction map and $\tilde{R} \in Y^{Y}$ denotes a co-domain-oriented correction map and the map $\circ_{X}: Y^{X} \times X^{X} \rightarrow Y^{X}$ and the map $\circ_{Y}$ : $Y^{Y} \times Y^{X} \rightarrow Y^{X}$ denote suitable composition maps. Notice well that, in (3.104), the maps $\tilde{\mathrm{K}} \circ_{X} \tilde{\mathrm{P}}: X \rightarrow Y, \tilde{\mathrm{R}} \circ_{Y} \tilde{\mathrm{~K}}: X \rightarrow Y$, and $\tilde{\mathrm{R}} \circ_{Y} \tilde{\mathrm{~K}} \circ_{X} \tilde{\mathrm{P}}: X \rightarrow Y$ designate surrogate models. Hence, we encounter the conceptional distinction between the notion of a low-fidelity model and a surrogate model that has been mentioned in §3.1.1.

By using the defect correction principle of numerical analysis as a scaffolding, the author in [56, ch. 2.5] investigates the space mapping paradigm. With regard to the corresponding numerical iteration schemes, the defect correction principle permits to interpret implementations of the map $\tilde{R}$ and the map $\tilde{P}$ as a left-preconditioner and a right-preconditioner, respectively. In $\S$ 3.3.1, I dwell on algorithmic instances associated with the space mapping paradigm.

In § 4, the category theoretical language is employed as an algebraic modeling scaffolding in order to assess its capability to complement the primarily numerical analytic narrative on simplified-physics low-fidelity models and the space mapping paradigm.

### 3.2 Surrogate-based optimization

A basic premise in the present work is that the acquisition of pairs of sampling plan points and output points with respect to a sample $\boldsymbol{s}$ in (3.13) is computationally expensive. Thus, it forces a user to be parsimonious with regard to the sample size $m$.

From an engineering application viewpoint, though, imagine the use case in which an effortless interplay between hardware and software enables a much faster acquisition of a sample $\mathbf{s}$ than without using this interplay. More concretely, imagine that a user can, without much ado, exploit opportunities for parallel computing and GPU (graphics processing unit) computing.

Recalling the Figure 1.4, this use case shifts rather the attention from the level of algorithms and the level of programs to a level of hardware technologies which is out of the scope of the present work.

However, if we abstract from the aforementioned concrete use case, then it reveals this section's main aim as gaining some insights about the degree of similarity between a high-fidelity model and a low-fidelity model without the usage of a model management strategy. This consideration corresponds to the assessment of the global and the local accuracy of a low-fidelity model with regard to a highfidelity model.

Building upon a certain degree of established similarity between a high-fidelity model and a low-fidelity model, the basic idea underlying the surrogate-based optimization (cf. § 1.2) is, first, to find a minimum associated with the low-fidelity model and, second, either to accept this low-fidelity model's minimum as a proxy to some extent - of a minimum associated with the high-fidelity model or to use this low-fidelity model's minimum as a starting point of the search for a minimum within the high-fidelity model.

In the subsequent subsections, let us examine the optimization with the test functions in Figure 2.2 by data-fit low-fidelity models and by emulated simplifiedphysics low-fidelity models.

### 3.2.1 Optimization with test functions by data-fit low-fidelity models

Using the Sobol quasi-random sequence sampling plan in Figure 3.2, let us invoke a 2 -variate monomial polynomial model $p(x) \in \mathrm{P}_{\leq 2}^{2}$ via regression of the test functions (and high-fidelity models, respectively) in Figure 2.2 where the matrices $W$ and $R$ in (3.65) are chosen such that $W:=I$ with $I \in \mathbb{R}^{m \times m}$ and $R:=0$ with $0 \in \mathbb{R}^{6 \times 6}$. In Figure 3.9, the corresponding contour representations are depicted in the cases where the number of sampling plan points is given by $m:=10$ and by $m:=50$.

In Figure 3.10, a Sobol quasi-random sequence sampling plan is utilized with the number of sampling plan points set to $m:=10$ and to $m:=50$ as well. A radial basis function $\phi=r \mapsto \phi(r)$ with thin plate spline assignment (see Table 3.2) is invoked via interpolation of the test functions in Figure 2.2. Let us choose the thin plate spline assignment as a representative of radial basis functions without additional hyperparameters to be adjusted. Mind that, for visualization purposes, in-house Julia code
is combined with the Julia PL package ScatteredInterpolation.jl (see [141] and [216]).

Finally, utilizing the Sobol quasi-random sequence sampling plan with $m:=10$ and $m:=50$ again, let us invoke the last data-fit low-fidelity model, that is, a kriging low-fidelity model via interpolation of the test functions in Figure 2.2. Since the computational burden of finding numerically the optimal parameters ( $\hat{\theta}, \hat{p}$ ) in (3.97) scales with the dimensionality $d$, let us choose a compromise with regard to the parameters ( $\hat{\theta}, \hat{p}$ ) in the sense that it is set that ( $\hat{\theta}_{1}, \hat{\theta}_{2}, \hat{p}_{1}, \hat{p}_{2} \equiv \hat{p}_{1}$ ). Thus, we compensate slightly the computational burden of finding numerically the optimal parameters whilst taking into account the benefit of a numerical search for optimal parameters compared to a manually predefined set of parameters. Recalling § 2.3.3, notice that the Nelder-Mead simplex algorithm is primarily employed to the optimization problem in (3.97). Compared loosely to an adaptive differential evolution algorithm, the NMS algorithm's results differ mostly in one or two decimal places from the ADE algorithm's results but, on the average, the NMS algorithm's results are achieved faster than the ADE algorithm's results. Though, it is hard to generalize this observation and to detect a firm preference for an optimization algorithm for the task at hand in (3.97). Mind that, for visualization purposes, in-house Julia code is combined with the Julia PL package Surrogates. jl .

At a qualitative level, given the number of sampling plan points by $m:=10$, one can observe that the 2 -variate monomial polynomial model best recovers the contour, more precisely, the value and the shape, of the Unit sphere and the Booth test function. Furthermore, it satisfactorily recovers the contour of the Rosenbrock and the modified Branin test function, and it worst recovers the contour of the Ackley and the Michalewicz test function. Due to the known definitions of the test functions, these observations are plausible. Mind that the polynomial model is invoked in a regression context, hence, the influence of another kind of polynomial model on the quality of the low-fidelity model with respect to the high-fidelity model is subdued. However, the influence of a higher number of sampling plan points is slightly bigger - especially, if we consider the modified Branin test function which can be seen as a rather protypical function within an engineering applications' context (cf. [70, p. 196]).

In the case of $m:=10$, if compared to the monomial polynomial model within the regression context, then the thin plate spline radial basis function and the kriging low-fidelity model within the interpolation context recover moderately, for instance, the modified Branin function. If we contrast the thin plate spline radial basis function (or short TPS RBF) with the kriging low-fidelity model, then one can observe that the kriging low-fidelity model tends to retrieve more accurately the values of the test functions whereas the thin plate spline radial basis function tends to retrieve more accurately the shape of the test functions. However, in the case of $m:=50$, both low-fidelity models are able to recover the values and the shapes of the highfidelity models satisfactorily albeit the kriging low-fidelity performs the recovery slightly better. Recall, though, that the thin plate spline radial basis function does not involve any hyperparameters, thus, a computationally intensive hyperparameters optimization step is omitted.

At a quantitative level, let us look at the normalized mean generalization error $\overline{\mathfrak{c}}_{\mathrm{H}, s_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \tilde{\tilde{y}}}^{2}\right|_{k}$ within the $k$-fold cross validation method w.r.t. deterministic and probabilistic data-fit low-fidelity models. ${ }^{23}$

[^33]
(A) The number of sampling plan points is given by $m:=10$.

The red cross refers to the global minimum of the high-fidelity models in Figure 2.2b.

(в) The number of sampling plan points is given by $m:=50$.

The red cross refers to the global minimum of the high-fidelity models in Figure 2.2b.

Figure 3.9: Using the Sobol quasi-random sequence sampling plan in Figure 3.2, a 2-variate monomial polynomial model $p(x) \in \mathrm{P}_{\leq 2}^{2}$ via regression ( $W:=I$ with $I \in \mathbb{R}^{m \times m}$ and $R:=0$ with $0 \in \mathbb{R}^{6 \times 6}$ in (3.65)) of the test functions (and high-fidelity models, respectively) in Figure 2.2 (solely in contour representation).

(A) The number of sampling plan points is given by $m:=10$.

The red cross refers to the global minimum of the high-fidelity models in Figure 2.2b.

(B) The number of sampling plan points is given by $m:=50$.

The red cross refers to the global minimum of the high-fidelity models in Figure 2.2b.

Figure 3.10: Using the Sobol quasi-random sequence sampling plan in Figure 3.2, a radial basis function $\phi=r \mapsto \phi(r)$ with thin plate spline assignment (see Table 3.2) via interpolation of the test functions (and
high-fidelity models, respectively) in Figure 2.2
(solely in contour representation).

(A) The number of sampling plan points is given by $m:=10$.

The red cross refers to the global minimum of the high-fidelity models in Figure 2.2b.

(в) The number of sampling plan points is given by $m:=50$.

The red cross refers to the global minimum of the high-fidelity models in Figure 2.2b.

Figure 3.11: Using the Sobol quasi-random sequence sampling plan in Figure 3.2, a kriging low-fidelity model via interpolation (where $\left(\hat{\theta}_{1}, \hat{2}_{2}, \hat{p}_{1}, \hat{p}_{2} \equiv \hat{p}_{1}\right)$ in (3.97)) of the test functions (and high-fidelity models, respectively) in Figure 2.2 (solely in contour representation).

As it has been already uttered in § 3.1.1, we consider the 5 -fold case and the 10 -fold case. The sample size is set to $m:=50$ since we consider the results associated with this sample size as sample-based best case error estimates and samplebased lower error bounds, respectively. The wording concerning the estimates and bounds is rather a pragmatism-driven ad-hoc artifice and it should not be regarded too tightly through the formally well crafted glasses in the context of numerical simulations (see § 2.2).

Moreover, let us consider the normalized global first-order sensitivity measures $S_{\tilde{\hat{y}}, i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the data-fit low-fidelity models.

Table 3.3: The normalized mean generalization error $\bar{\varepsilon}_{\mathrm{H}, \mathbf{s}_{\xi}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \tilde{\tilde{y}}}^{2}\right|_{k}$ within the $k$-fold cross validation method w.r.t. the 2 -variate monomial polynomial in Figure 3.9 with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\left.\overline{\mathfrak{e}}_{\mathrm{H}, s_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right\|_{k:=5}$ | 0.5098 | $<1.0 \times 10^{-16}$ | $\ll 1.0 \times 10^{-16}$ | $\gg 1.0$ | 0.0657 | $>1.0$ |
| $\left.\bar{r}_{\hat{y} \tilde{\hat{y}}}^{2}\right\|_{k=5}$ | 0.5461 | 1.0 | 1.0 | 0.9003 | 0.3369 | 0.6558 |
| $\left.\overline{\mathfrak{c}}_{\mathrm{H}, \mathrm{s}_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)\right\|_{k:=10}$ | 0.4477 | $<1.0 \times 10^{-16}$ | $\ll 1.0 \times 10^{-16}$ | $\gg 1.0$ | 0.0582 | $>1.0$ |
| $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right\|_{k:=10}$ | 0.7217 | 1.0 | 1.0 | 0.9352 | 0.3951 | 0.8137 |

TABLE 3.4: The normalized mean generalization error $\overline{\mathfrak{e}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\mathcal{\xi}}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k}$ within the $k$-fold cross validation method w.r.t. the radial basis function with thin plate spline assignment in Figure 3.10 with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\left.\overline{\mathfrak{e}}_{\mathrm{H}, s_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{Z}}\right)\right\|_{k=5}$ | 0.3377 | 0.0149 | 0.4173 | $\gg 1.0$ | 0.0935 | 0.3707 |
| $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right\|_{k:=5}$ | 0.6166 | 0.9999 | 0.9973 | 0.9610 | 0.1772 | 0.8905 |
| $\left.\overline{\mathfrak{e}}_{\mathrm{H}, s_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right\|_{k:=10}$ | 0.2330 | 0.0111 | 0.4064 | $\gg 1.0$ | 0.0677 | 0.2898 |
| $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right\|_{k:=10}$ | 0.6963 | 0.9999 | 0.9963 | 0.9780 | 0.4328 | 0.9732 |

In Table 3.3, the normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. the 2 -variate monomial polynomial in Figure 3.9 is presented. It supports the observations at the qualitative level. If we pick the modified Branin test function as an example, then the Table hints additionally at the monomial polynomial's convenience for recovering at least partly the shape of such a test function.

In Table 3.4, the normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. a radial basis function with thin plate spline assignment is listed. Compared to the monomial polynomial, the thin plate spline radial basis function recovers better the values and the shape of the modified Branin test function.

TABLE 3.5: The normalized mean generalization error $\overline{\mathfrak{c}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\mathcal{\xi}}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right|_{k}$ within the $k$-fold cross validation method w.r.t. the kriging low-fidelity model in Figure 3.11 with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\left.\overline{\mathfrak{e}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right\|_{k:=5}$ | 0.2517 | 0.2818 | 0.0043 | 0.9956 | 0.0443 | 0.0242 |
| $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right\|_{k:=5}$ | 0.5367 | 0.9987 | 0.9999 | 0.9998 | 0.5188 | 0.9983 |
| $\left.\overline{\mathfrak{e}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)\right\|_{k:=10}$ | 0.2555 | 0.0925 | 0.0011 | 0.3290 | 0.0130 | 0.0022 |
| $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right\|_{k:=10} ^{2}$ | 0.5838 | 0.9999 | 0.9999 | 0.9998 | 0.7665 | 0.9998 |

TABLE 3.6: The normalized global first-order sensitivity measure $S_{\tilde{\hat{y}}, i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the 2 -variate monomial polynomial in Figure 3.9b with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $S_{\tilde{y}, 1}^{\mathrm{N}}(f)$ | 0.4922 | 0.5000 | 0.4894 | 0.9935 | 0.8633 | 0.4455 |
| $S_{\tilde{y}, 2}^{N}(f)$ | 0.5078 | 0.5000 | 0.5106 | 0.0065 | 0.1367 | 0.5545 |
| $\sum_{i=1}^{2} S_{\tilde{\hat{y}}, i}^{\mathrm{N}}(f)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

TABLE 3.7: The normalized global first-order sensitivity measure $S_{\tilde{y}, i}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the radial basis function with thin plate spline assignment in Figure 3.10b with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $S_{\tilde{y}, 1}^{\mathrm{N}}(f)$ | 0.5041 | 0.5000 | 0.4851 | 0.9733 | 0.5351 | 0.6226 |
| $S_{\tilde{y}, 2}^{\mathrm{N}}(f)$ | 0.4959 | 0.5000 | 0.5149 | 0.0267 | 0.4649 | 0.3774 |
| $\sum_{i=1}^{2} S_{\tilde{\hat{y}}, i}^{\mathrm{N}}(f)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

TABLE 3.8: The normalized global first-order sensitivity measure $S_{\tilde{\hat{y}}, i}^{N}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the kriging low-fidelity model in Figure 3.11b with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $S_{\tilde{y}, 1}^{\mathrm{N}}(f)$ | 0.8994 | 0.5000 | 0.4893 | 0.9963 | 0.2609 | 0.7221 |
| $S_{\tilde{y}, 2}^{\mathrm{N}}(f)$ | 0.1006 | 0.5000 | 0.5107 | 0.0037 | 0.7391 | 0.2779 |
| $\Sigma_{i=1}^{2} S_{\tilde{\hat{y}}, i}^{\mathrm{N}}(f)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

In Table 3.5, the normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. a kriging low-fidelity model. Compared to the thin plate spline radial basis function and the monomial polynomial, the kriging low-fidelity model shows a relatively high mean SSPCC for all test functions. Observe that the kriging low-fidelity model is capable to recover the values
and the shape of the modified Branin test function with a high degree of accuracy.
However, all data-fit low-fidelity model exhibits some difficulties in recovering the values of (iv) the Rosenbrock function. This observation indicates that there is a need for a normalization of the test function's values. Thus, let us determine automatically the decimal power of the test function's maximal value and normalize all the test function's values regarding this decimal power. Notice well that the shaperelated entities such as the SSPCC or the normalized global first-order sensitivity measures are not affected by a normalization of the test function's values.

In Figure 3.12, in Figure 3.13 and in Figure 3.14, let us extend in a self-explanatory manner the definition of $\left.\overline{\mathcal{\varepsilon}}_{\mathrm{H}, \mathrm{s}_{\delta}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right|_{k}$ in (3.28) - and the definition of $\left.\overline{\mathcal{\varepsilon}}_{\mathrm{H}, \mathrm{s}_{\xi}}^{\mathrm{NR}}\left(\hat{Q}_{\xi}\right)\right|_{k}$ as well - and the definition of $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right|_{k}$ in (3.33) in order to depict the corresponding courses of the paths regarding the number of testing points $m_{g}$.

Observe that the number of testing points is truncated at $m_{g}:=10$ which corresponds to the total number of points $m$ by $m:=k m_{g}$. If $k:=10$, then $m:=100$. Thus, from an application-driven viewpoint with, e.g., weaker requirements w.r.t. the accuracy, let us regard such a sample size as an upper limit. An immediate potential drawback of such an upper limit is that, in general, one cannot expect to detect some kind of monotonicity which is usually associated with an asymptotic consideration.

Inspecting Figure 3.12b, one can observe that the paths associated with the error $\left.\bar{\varepsilon}_{H, s_{g}}^{N R}\left(\hat{Q}_{\xi}\right)\right|_{k}$ reflect values that are order of magnitudes higher than the values corresponding to the paths associated with the error $\left.\overline{\mathcal{e}}_{\mathrm{H}, s_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)\right|_{k}$. This observation echos the more conservative behavior of $\left.\overline{\mathcal{\varepsilon}}_{\mathrm{H}, s_{g}}^{\mathrm{NR}}\left(\hat{Q}_{\xi}\right)\right|_{k}$ (cf. § 3.1.1).

Observing Figure 3.12 and Figure 3.14, a valuable insight is that the continuous deformation of the courses of paths corresponding to the 5 -fold cross validation into the courses of paths corresponding to the 10 -fold cross validation is relatively marginal - except for (i) the Ackley test function and (v) the Michalewicz test function which, from an application-driven viewpoint, one can consider as extreme use cases as opposed to the common use cases (ii), (iii), (iv), and (vi). Another valuable insight is that if the requirements regarding the values and the shapes of the lowfidelity models are weakened, then, for the common use cases, the deterministic data-fit low-fidelity models can provide a computationally less expensive alternative for the probabilistic data-fit low-fidelity models.

In Table 3.6, in Table 3.7, and in Table 3.8, I present the results regarding the normalized global first-order sensitivity measure $S_{\tilde{y}, i}^{N}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the 2 -variate monomial polynomial in Figure 3.9 b, the radial basis function with thin plate spline assignment in Figure 3.10b, and the kriging low-fidelity model in Figure 3.11b, respectively. In all cases the sample size $m$ is set to $m:=50$.

In Figure 3.15, let us extend in an obvious way the definition of $S_{\tilde{y}, i}^{\mathrm{N}}(f)$ in (3.34) in order to depict the corresponding courses of the paths regarding the number of training points $m_{t}$.

A valuable insight is that if we only focus on limit considerations, then we are not able to tell some test functions apart (see, e.g., (i), (ii) or (iii) in Figure 3.15). However, the courses of the paths furnish us with some hints about the landscapes of the test functions, therefore, they are helping us partially to tell the test functions apart.

Another valuable insight is that, in the case of the interpolation-focused data-fit low-fidelity models, the entity $S_{\tilde{y}, i}^{\mathbb{N}}(f)$ exhibit some kind of correlation with the entity $\bar{r}_{\hat{y} \tilde{y}, \mathrm{cv}}^{2}\left(k, m_{\mathrm{g}}\right)$. This insight drives the conjecture about the trustworthiness of lowfidelity models' normalized global first-order sensitivity measures in (3.35). Mind
that this insight should rather be understood as an intriguing starting point for future investigations where the behavior of regression-focused and interpolationfocused data-fit low-fidelity models is elaborated more thoroughly, and where the number of test functions under consideration is increased substantially. To my best knowledge, there is currently a lack of such extensive benchmarking in a similar style of Figure 3.12, Figure 3.13, Figure 3.14, and Figure 3.15.

Since the elaborations are without loss of generality w.r.t. the dimensionality $d$, let us dwell briefly on this issue by exploring the Rosenbrock test function on the smaller domain $[-2.0,2.0]^{d}$ (for the case $d=2$, see the Figure 2.3) where the dimensionality $d$ is governed by $d \in\{2,3,4,5,6,7\}$. Thus, we explore the evolution of the normalized mean generalization error, the mean SSPCC, and the normalized global first-order sensitivity measure with the dimensionality.

Recalling § 2.3.3, the Rosenbrock test function in Table 2.1 permits to consider immediately an arbitrary dimensionality with $d \in\{2,3,4,5,6,7\}$, more precisely, $f_{\mathrm{R}_{N_{\xi}}}$ from (2.52) with $N_{\xi} \equiv d$, where, in each case, the global minimum is at the point $(1,1, \ldots, 1) \in \mathbb{R}^{d}$. In Table 3.9, I depict the normalized mean generalization error and the mean SSPCC within the $k$-fold cross validation method w.r.t. a kriging lowfidelity model of a generalized version of the Rosenbrock test function where the sample size $m$ is set to $m:=50$. Essentially, the Table 3.9 reveals how strongly the curse of dimensionality kicks in: For instance, the error $\left.\overline{\mathfrak{c}}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)\right|_{k:=10}$ increases by approximately five orders of magnitude from $d:=2$ to $d:=7$, and the SSPCC $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k:=10}$ drops by almost $70 \%$ from $d:=2$ to $d:=7$. Geometrically, this observations translates into shifts of the corresponding courses of paths in Figure 3.13 and in Figure 3.14. Thus, the findings in Table 3.9 can be conceived as a worst-case estimate of the evolution of the normalized mean generalization error and the mean SSPCC with the dimensionality regarding the findings in Figure 3.13 and in Figure 3.14.

TAble 3.9: Given the sample size $m:=50$, the normalized mean generalization error $\bar{\varepsilon}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right|_{k}$ within the $k$ fold cross validation method w.r.t. a kriging low-fidelity model of the Rosenbrock test function in Table 2.1 (with normalized values) generalized to the domain $[-2.0,2.0]^{d}$ with $d \in\{2,3,4,5,6,7\}$.

| Dimensionality $d$ | $\left.\overline{\mathfrak{e}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right\|_{k:=5}$ | $\left.\bar{r}_{\hat{y} \tilde{\hat{y}}}^{2}\right\|_{k:=5}$ | $\left.\overline{\mathfrak{e}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right\|_{k:=10}$ | $\left.\bar{r}_{\hat{y} \tilde{y} \tilde{\hat{y}}}^{2}\right\|_{k:=10}$ |
| :---: | :---: | :---: | :---: | :---: |
| 2 | $1.5808 \times 10^{-6}$ | 0.9998 | $7.9288 \times 10^{-7}$ | 0.9996 |
| 3 | $3.2861 \times 10^{-4}$ | 0.9652 | $9.0496 \times 10^{-5}$ | 0.9915 |
| 4 | $4.7462 \times 10^{-3}$ | 0.6747 | $3.5632 \times 10^{-3}$ | 0.6093 |
| 5 | $6.5692 \times 10^{-3}$ | 0.6483 | $5.7308 \times 10^{-3}$ | 0.6387 |
| 6 | $9.7461 \times 10^{-3}$ | 0.2742 | $8.7787 \times 10^{-3}$ | 0.3653 |
| 7 | $1.3586 \times 10^{-2}$ | 0.0393 | $1.3579 \times 10^{-2}$ | 0.3013 |

Determining the normalized mean generalization error $\left.\overline{\mathcal{E}}_{\mathrm{H}, \mathrm{s}_{g}}\left(\hat{Q}_{\tilde{\xi}}\right)\right|_{k:=5}$ for the dimensionality $d:=7$ takes, roughly, about 23 times longer than for the dimensionality $d:=2$. Furthermore, determining the normalized mean generalization error $\left.\overline{\mathcal{E}}_{\mathrm{H}, s_{\xi}}\left(\widehat{Q}_{\xi}\right)\right|_{k:=10}$ for the dimensionality $d:=7$ takes, roughly, about 17 times longer than for the dimensionality $d:=2$. In both cases, the longer computation time for the dimensionality $d:=7$ is primarily dominated by the more involved optimization in (3.97). However, these factors should merely be conceived as raw estimates or raw proxies of

(B) $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{NR}}\left(5, m_{\mathrm{g}}\right)$ vs. $m_{\mathrm{g}}$.

Figure 3.12: The value of $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}$ and $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{NR}}$ evaluated at the number of folds $k$ and the number of testing points $m_{g}$ w.r.t. the test functions in Figure 2.2 with normalized values. The number $m_{\mathrm{g}}$ corresponds to the number $m_{t}$ by $m_{t}:=(k-1) m_{\mathrm{g}}$ and to the number $m$ by $m:=k m_{\mathrm{g}}$. The data-fit low-fidelity models are ( $\bullet$ ) a 2-variate monomial polynomial via regression, ( $\mathbf{\bullet})$ a TPS RBF, and ( $\uparrow$ ) a kriging model.
In (ii) and (iii), $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}\left(k, m_{\mathrm{g}}\right)$ w.r.t. ( $(\stackrel{)}{ }$ is below machine precision.

(B) $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}\left(10, m_{\mathrm{g}}\right)$ vs. $m_{\mathrm{g}}$.

Figure 3.13: The value of $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}$ evaluated at the number of folds $k$ and the number of testing points $m_{g}$ w.r.t. the test functions in Figure 2.2 with normalized values. The number $m_{\mathrm{g}}$ corresponds to the number $m_{t}$ by $m_{t}:=(k-1) m_{\mathrm{g}}$ and to the number $m$ by $m:=k m_{\mathrm{g}}$.
The data-fit low-fidelity models are ( $\bullet$ ) a 2-variate monomial polynomial via regression, ( $\boldsymbol{\bullet})$ a TPS RBF, and $(\star)$ a kriging model.
In (ii) and (iii), $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}\left(k, m_{\mathrm{g}}\right)$ w.r.t. ( () is below machine precision.

(B) $\bar{r}_{\hat{y} \tilde{y}, \mathrm{cv}}^{2}\left(10, m_{\mathrm{g}}\right)$ vs. $m_{\mathrm{g}}$.

Figure 3.14: The value of $\bar{r}_{\hat{y} \tilde{y}, \mathrm{cv}}^{2}$ evaluated at the number of folds $k$ and the number of testing points $m_{g}$ w.r.t. the test functions in Figure 2.2. The number $m_{\mathrm{g}}$ corresponds to the number $m_{t}$ by $m_{t}:=(k-1) m_{\mathrm{g}}$ and to the number $m$ by $m:=k m_{\mathrm{g}}$.
The data-fit low-fidelity models are ( $\bullet$ ) a 2-variate monomial polynomial via regression, ( $\mathbf{\bullet})$ a TPS RBF, and $(\uparrow)$ a kriging model.

(A) $S_{\hat{y}, 1}^{\mathrm{N}}\left(f, m_{t}\right)$ vs. $m_{t}$.

(B) $S_{\hat{y}, 2}^{\mathrm{N}}\left(f, m_{t}\right)$ vs. $m_{t}$.

Figure 3.15: The value of $S_{\tilde{\hat{y}}, i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at $f$ w.r.t. the test functions in Figure 2.2 and the number of training points $m_{t}$.
The data-fit low-fidelity models are ( $\bullet$ ) a 2-variate monomial polynomial via regression, ( $\boldsymbol{\bullet})$ a TPS RBF, and $(\star)$ a kriging model.
The gray dotted line indicates the position $m_{t}:=12$ and the gray dashed line indicates the position $m_{t}:=27$. The thick black line refers to the value of $S_{i}^{\mathrm{N}}(f)$ of the respective test function in Table 2.2.
the overhead in the optimization in (3.97) due to the increase in the dimensionality.
TABLE 3.10: The normalized global first-order sensitivity measure $S_{i}^{\mathrm{N}}$ evaluated at $f_{\mathrm{R}_{\mathrm{N}_{\xi}}}$ from (2.52) w.r.t. the data in Table 3.9.

|  | $S_{\tilde{y}, i}^{\mathrm{N}}(f)$ | $i:=1$ | $i:=2$ | $i:=3$ | $i:=4$ | $i:=5$ | $i:=6$ | $i:=7$ | $\Sigma_{i=1}^{d} S_{\tilde{y}, i}^{\mathrm{N}}(f)$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2 |  | 0.9081 | 0.0919 | - | - | - | - | - | 1.0000 |
| 3 | 0.4125 | 0.5475 | 0.0400 | - | - | - | - | 1.0000 |  |
| 4 | 0.2004 | 0.4122 | 0.3568 | 0.0306 | - | - | - | 1.0000 |  |
| 5 | 0.2133 | 0.2161 | 0.2386 | 0.3227 | 0.0093 | - | - | 1.0000 |  |
| 6 | 0.1228 | 0.1842 | 0.2792 | 0.3246 | 0.0861 | 0.0031 | - | 1.0000 |  |
| 7 | 0.0614 | 0.2896 | 0.0592 | 0.1496 | 0.0158 | 0.2985 | 0.1259 | 1.0000 |  |

Table 3.11: Using the reference values in Table 2.4, the low-fidelity models' normalized global first-order sensitivity measures (LFSM) error in (3.37) evaluated at $f_{\mathrm{R}_{\tilde{\xi}}}$ from (2.52) w.r.t. the data in Table 3.9.

|  | $\mathfrak{e}_{m_{\infty}}\left(S_{\tilde{y}, i}^{N}\right)$ | $i:=1$ | $i:=2$ | $i:=3$ | $i:=4$ | $i:=5$ | $i:=6$ | $i:=7$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2 |  | +0.0031 | -0.0314 | - | - | - | - | - |
| 3 | -0.0292 | +0.0223 | -0.0204 | - | - | - | - |  |
| 4 | +0.2199 | -0.1482 | +0.0061 | -0.2191 | - | - | - |  |
| 5 | -0.1274 | +0.1817 | +0.0966 | -0.2219 | +0.4973 | - | - |  |
| 6 | +0.1780 | +0.1187 | -0.3359 | -0.5531 | +0.5880 | +0.7877 | - |  |
| 7 | +0.5044 | -0.6759 | +0.6574 | +0.1343 | +0.9086 | -0.7274 | -9.405 |  |

In Table 3.10, I present the normalized global first-order sensitivity measure $S_{i}^{\mathrm{N}}$ evaluated at $f_{\mathrm{R}_{N_{\xi}}}$ from (2.52) w.r.t. the data in Table 3.9. Additionally, in Table 3.11, I present the LFSM error in (3.37) w.r.t. the Table 2.4. The increase of the unsigned LFSM error in Table 3.11 from $d:=2$ to $d:=7$ correlates with the decrease of the SSPCC in Table 3.9 from $d:=2$ to $d:=7$. Hence, this relationship supports in higher dimensions the similar observations in Figure 3.15 and in Figure 3.14 for the case $d:=2$.

A benefit of the previously discussed tables and figures with regard to the deterministic and probabilistic data-fit low-fidelity models is that they provide us with some quantitative hints about the values and the shape (or landscape) associated with the high-fidelity models in the physics-oriented context of the applications in the preset work such that one can roughly assign these high-fidelity models' behavior to one or more test functions from Table 2.1.

Thus, this by no means complete list of indicators or properties (i.e., tables and figures) enables us some kind of raw classification of application-driven high-fidelity models' behaviors. More interestingly, if a high-fidelity model does not possess certain properties of a selected function from the list of test functions, then one can make an educated guess that the high-fidelity model is probably not a representative of the selected function from the list of test functions.

However, there are a couple of caveats regarding such a benchmark-focused classification. First, it is not clear whether there exists a reliable complete list of
indicators or properties. Second, it is not clear how big the list of test functions has to be. Nevertheless, from an application-oriented viewpoint as well as from a theory-oriented viewpoint, such an attempt of a benchmark-focused classification is a worthwhile endeavor.

To conclude the discussion with respect to the data-fit low-fidelity models, let us perform a surrogate-based optimization concerning the modified Branin test function according to the following schematic procedure that I refer to as SBO-DFLF:

1) create a sample w.r.t. the high-fidelity model (§3.1.1);
2) construct a data-fit low-fidelity model (§3.1.2) w.r.t. the sample from step 1 );
3) invoke a global optimization algorithm (§ 2.3.3) w.r.t. the constructed data-fit low-fidelity model from step 2);
4) use the minimizer from step 3) as a starting point for a local optimization algorithm (§ 2.3.3) w.r.t. the high-fidelity model from step 1 ).

Notice that the proposed schematic procedure SBO-DFLF builds upon the common canon of optimization algorithms (recall § 2.3.3) such as the Nelder-Mead simplex (NMS) algorithm and the adaptive differential evolution (ADE) algorithm.

Since the present work's core mantra is that a function evaluation of the highfidelity model is computationally expensive, let us count the number of high-fidelity model function evaluations in order to find the modified Branin test function's global minimizer and its corresponding value presented in Table 2.1 within a certain accuracy.

We start with ten function evaluations because the sample size $m$ is set to $m:=10$ such as in Figure 3.9a, and Figure 3.10a, and Figure 3.11a in order to construct the respective data-fit low-fidelity model.

Next, let use invoke an adaptive differential evolution (ADE) algorithm to determine the global minimizer and its corresponding value of the respective data-fit low-fidelity model. Mind that we only consider box constraints. Depending on the degree of rigor needed for a task at hand, in addition, a guaranteed deterministic global optimization algorithm using interval arithmetic can be employed in order to certify the result of the ADE algorithm, i.e., the global minimizer and its corresponding value of the respective data-fit low-fidelity model.

Finally, let us use the minimizer from the ADE algorithm as a starting point for a Nelder-Mead simplex (NMS) algorithm with regard to the modified Branin test function. Note that, instead of the NMS algorithm from (Opkg1), the NMS algorithm from (Opkg3) is invoked since, in addition to the common classic implementation, it provides a modern implementation in the sense that, e.g., it defaults to an adaptive tuning parameters scheme and it focuses on keeping the number of necessary function evaluations low. ${ }^{24}$

In Table 3.12, I present the results from the proposed schematic surrogate-based optimization procedure SBO-DFLF. The quantity $\mathrm{er}_{x^{*}}$ refers to the relative error with respect to the global minimizer listed in Table 2.1 and the quantity $\mathrm{er}_{f\left(x^{*}\right)}$ refers to the relative error with respect to the global minimizer's function value listed in Table 2.1 as well. The quantity $i^{\text {NMS }}$ refers to the total number of function evaluations in the NMS algorithm.

[^34]Table 3.12: Building upon the Figure 3.9a, and the Figure 3.10a, and the Figure 3.11a, surrogate-based optimization according to the proposed schematic procedure SBO-DFLF w.r.t. the modified Branin function using data-fit low-fidelity models.

| Low-fidelity model | $m$ | $\mathrm{it}_{\text {NMS }}$ | $\mathrm{er}_{x^{*}}$ | $\mathrm{er}_{f\left(x^{*}\right)}$ |
| :---: | :---: | :---: | :---: | :---: |
| Polynomial | 10 | 11 | 0.0086 | 0.0018 |
| TPS RBF | 20 | 17 | 0.0703 | 0.0400 |
| Kriging | 10 | 11 | 0.0067 | 0.0023 |

The Table 3.12 reveals that the procedures based on the monomial polynomial model and the kriging low-fidelity model consume in total 21 function evaluations of the high-fidelity model in order to produce a relative error $\mathrm{er}_{x^{*}}$ and a relative error $\operatorname{er}_{f\left(x^{*}\right)}$ that, encoding these errors in percentage, are less than one percent. Choosing the thin plate spline radial basis function requires to increase the sample size, otherwise the TPS RBF is not capable to produce a sufficiently good starting point. Ultimately, the procedure based on the TPS RBF consumes in total 37 function evaluations of the high-fidelity model in order to produce a relative error $\mathrm{er}_{x^{*}}$ and a relative error $\operatorname{er}_{f\left(x^{*}\right)}$ that, encoding these errors in percentage, are less than eight percent.

In practical applications, usually, there is no apriori knowledge about the minimizers of a high-fidelity model. Hence, the proposed schematic surrogate-based optimization procedure SBO-DFLF offers an approach to find relatively quickly and sufficiently accurately a global minimizer of a high-fidelity model.

An advantage of the Table 3.12 is that it serves as a quantitative hint at the potential quality of a surrogate-based optimization procedure that is useful for the assessment of a surrogate-guided optimization procedure.

### 3.2.2 Optimization with test functions by emulated simplified-physics low-fidelity models

Finally, let us emulate simplified-physics low-fidelity models for the modified Branin test function via the assignment rule

$$
\begin{equation*}
\tilde{\mathrm{K}}(x):=\gamma+_{\mathbb{R}} \delta \cdot \mathbb{R} \mathrm{K}\left(\alpha+_{\mathbb{R}^{2}} \beta \odot x\right) \tag{3.105}
\end{equation*}
$$

where $\gamma, \delta \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}^{2 \times 1}$ denote adjustment parameters and the map $\odot$ is comprehended as the Hadamard product or the element-wise product.

Table 3.13: The choice of the 4 -tuple of parameters $(\alpha, \beta, \gamma, \delta)$ in Figure 3.16a and in Figure 3.16b.

|  | Figure 3.16a | Figure 3.16b |
| :---: | :---: | :---: |
| (i) | $\left(\left[\begin{array}{ll}0.0 & 10.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.0 & 1.0\end{array}\right]^{\mathrm{T}}, 0.0,1.0\right)$ | $\left(\left[\begin{array}{ll}0.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.0 & 1.0\end{array}\right]^{\mathrm{T}}, 1.0 \times 10^{3}, 1.0\right)$ |
| (ii) | $\left(\left[\begin{array}{ll}10.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.0 & 1.0\end{array}\right]^{\mathrm{T}}, 0.0,1.0\right)$ | $\left(\left[\begin{array}{ll}10.0 & 10.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.0 & 1.0\end{array}\right]^{\mathrm{T}}, 1.0 \times 10^{3}, 1.0\right)$ |
| (iii) | $\left(\left[\begin{array}{lll}10.0 & 10.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.0 & 1.0\end{array}\right]^{\mathrm{T}}, 0.0,1.0\right)$ | $\left(\left[\begin{array}{ll}0.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.5 & 1.5\end{array}\right]^{\mathrm{T}}, 1.0 \times 10^{3}, 1.0\right)$ |
| (iv) | $\left(\left[\begin{array}{ll}0.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{lll}0.5 & 1.0\end{array}\right]^{\mathrm{T}}, 0.0,1.0\right)$ | $\left(\left[\begin{array}{ll}0.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.0 & 1.0\end{array}\right]^{\mathrm{T}}, 0.0,1.0 \times 10^{3}\right)$ |
| (v) | $\left(\left[\begin{array}{ll}0.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.0 & 0.5\end{array}\right]^{\mathrm{T}}, 0.0,1.0\right)$ | $\left(\left[\begin{array}{ll}10.0 & 10.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{lll}1.0 & 1.0\end{array}\right]^{\mathrm{T}}, 0.0,1.0 \times 10^{3}\right)$ |
| (vi) | $\left(\left[\begin{array}{lll}0.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.5 & 1.5\end{array}\right]^{\mathrm{T}}, 0.0,1.0\right)$ | $\left(\left[\begin{array}{ll}0.0 & 0.0\end{array}\right]^{\mathrm{T}},\left[\begin{array}{ll}1.5 & 1.5\end{array}\right]^{\mathrm{T}}, 0.0,1.0 \times 10^{3}\right)$ |

The assignment in (3.105) is a variation and a generalization to the two-dimensional case of the one-dimensional cases in [49, p. 86] and in [70, p. 195]. In Table 3.13, several choices of the 4 -tuple of parameters $(\alpha, \beta, \gamma, \delta)$ are listed whose influences on the low-fidelity model are depicted in Figure 3.16.

In Figure 3.17, I illustrate $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ as a projection on the contour representation of the emulated simplified-physics low-fidelity models.

For the sake of completion, in Table 3.14 and in Table 3.15, I provide the normalized mean generalization error $\overline{\mathcal{E}}_{\mathrm{H}, s_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right|_{k}$ within the $k$-fold cross validation method w.r.t. the emulated simplified-physics low-fidelity models in Figure 3.16 with sample size $m:=50$. For all combinations of the adjustment parameters in Table 3.13, the deterioration in the value is comprehensibly large. The deterioration in the shape is largest for those combinations of the adjustment parameters in which the argument of the high-fidelity model are shifted.

TABLE 3.14: The normalized mean generalization error $\bar{\varepsilon}_{H, S_{g}}^{N}\left(\hat{Q}_{\xi}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \tilde{j}}^{2}\right|_{k}$ within the $k$-fold cross validation method w.r.t. emulated simplified-physics low-fidelity models in Figure 3.16a with normalized values and with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\left.\overline{\mathfrak{e}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\xi}}\right)\right\|_{k:=5}$ | $6.2 \times 10^{-2}$ | $4.2 \times 10^{-2}$ | $9.6 \times 10^{-2}$ | $4.9 \times 10^{-3}$ | $1.7 \times 10^{-2}$ | $1.6 \times 10^{-2}$ |
| $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right\|_{k:=5}$ | 0.6391 | 0.0391 | 0.1092 | 0.7466 | 0.3434 | 0.8684 |
| $\left.\overline{\mathfrak{e}}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\tilde{\zeta}}\right)\right\|_{k:=10}$ | $6.2 \times 10^{-2}$ | $4.2 \times 10^{-2}$ | $9.6 \times 10^{-2}$ | $4.9 \times 10^{-3}$ | $1.7 \times 10^{-2}$ | $1.6 \times 10^{-2}$ |
| $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right\|_{k:=10}$ | 0.6238 | 0.2691 | 0.3007 | 0.7224 | 0.3611 | 0.7565 |

TABLE 3.15: The normalized mean generalization error $\overline{\mathfrak{e}}_{\mathrm{H}, \mathbf{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)$ and the mean SSPCC $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right|_{k}$ within the $k$-fold cross validation method w.r.t. emulated simplified-physics low-fidelity in Figure 3.16b with normalized values and with sample size $m:=50$.

|  | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\left.\overline{\mathfrak{c}}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)\right\|_{k:=5}$ | $1.8 \times 10^{-1}$ | $1.4 \times 10^{-1}$ | $8.6 \times 10^{-2}$ | $<2.2 \times 10^{-16}$ | $9.6 \times 10^{-2}$ | $1.6 \times 10^{-2}$ |
| $\left.\bar{r}_{\hat{y} \tilde{y}}\right\|_{k:=5}$ | 1.0 | 0.1558 | 0.8367 | 1.0 | 0.1558 | 0.8367 |
| $\left.\overline{\mathfrak{c}}_{\mathrm{H}, s_{g}}\left(\hat{Q}_{\xi}\right)\right\|_{k:=10}$ | $1.8 \times 10^{-1}$ | $1.4 \times 10^{-1}$ | $8.6 \times 10^{-2}$ | $<2.2 \times 10^{-16}$ | $9.6 \times 10^{-2}$ | $1.6 \times 10^{-2}$ |
| $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right\|_{k=10}$ | 1.0 | 0.3522 | 0.7545 | 1.0 | 0.3522 | 0.7545 |

Let us skip a discussion about the sensitivity measures since the procedure is analogous to the corresponding procedure in the previous section. However, an import insight from the Table 3.14 and the Table 3.15 is that if the mean SSPCC $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k}$ is above a threshold of 0.75 , then it might be useful to adapt the proposed procedure SBO-DFLF to the case of simplified-physics low-fidelity models as well. Notice that the threshold of 0.75 is rather user-dependent. Based on anecdotal evidence, however, the authors in [70, p. 37] allege that a threshold above 0.80 corresponds to a good low-fidelity model in terms of predictive power.

(A) The concrete 4 -tuple of parameters $(\alpha, \beta, \gamma, \delta)$ are listed in the first column of the Table 3.13. The red cross refers to the global minimum of the high-fidelity model in Figure 2.2b.

(в) The concrete 4 -tuple of parameters $(\alpha, \beta, \gamma, \delta)$ are listed in the second column of the Table 3.13. The red cross refers to the global minimum of the high-fidelity model in Figure 2.2b.

FIgURE 3.16: Emulated simplified-physics low-fidelity models for the modified Branin test function (and high-fidelity model, respectively) in Figure 2.2 via the assignment rule in (3.105) (solely in contour representation).

(A) Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ within Figure 3.16a.

Same scaling from (i) to (vi).

(в) Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ within Figure 3.16b.

Same scaling from (i) to (iii) and from (iv) to (vi). The scaling in (iv), in (v), and in (v) is $1 \times 10^{-3}$ of the scaling in (i), in (ii), and in (iii).

Figure 3.17: Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ as a projection on the contour representation of the emulated simplified-physics low-fidelity models for the modified Branin test function in Figure 3.16.

Let us apply the following schematic procedure that I refer to as SBO-SPLF:

1) create a sample w.r.t. the high-fidelity model (\$3.1.1);
2) provide a simplified-physics low-fidelity model (§3.1.3) and compute $\bar{r}_{\left.\hat{y} \tilde{y}\right|_{k} ^{2}}$ w.r.t. the sample from step 1) and if $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k}<0.75$, break off the procedure, otherwise continue the procedure;
3) invoke a global optimization algorithm (§ 2.3.3) w.r.t. the provided simplifiedphysics low-fidelity model from step 2);
4) use the minimizer from step 3) as a starting point for a local optimization algorithm ( $\$ 2.3 .3$ ) w.r.t. the high-fidelity model from step 1 ).

Similarly to the proposed procedure SBO-DFLF, the proposed schematic procedure SBO-SPLF builds upon the common canon of optimization algorithms such as the Nelder-Mead simplex (NMS) algorithm and the adaptive differential evolution (ADE) algorithm.

Depending on the desired degree of scrutiny and rigor, it is also conceivable to combine the proposed procedures SBO-DFLF and SBO-SPLF in sequence or in parallel and to compare, e.g., the respective minimizers that serve as starting points. Furthermore, the procedure SBO-DFLF suits well as a fallback branch for the procedure SBO-SPLF in the case that $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k}$ is below the threshold 0.75.

Finally, let us observe that, in order to determine the quantities $\left.\overline{\mathcal{e}}_{\mathrm{H}, \mathrm{s}_{g}}^{\mathrm{N}}\left(\hat{Q}_{\xi}\right)\right|_{k}$ and $\left.\bar{r}_{\hat{y} \tilde{\tilde{y}}}^{2}\right|_{k}$ in Table 3.14 and in Table 3.15, there are 50 evaluations of the simplified-physics low-fidelity model and 50 evaluations of the high-fidelity model needed. Thus, in practical applications, the sample size has to be lowered to, e.g., the number 10 that corresponds to 10 evaluations of the high-fidelity model and 10 evaluations of the simplified-physics low-fidelity model. However, in order to compute $\left.\bar{r}_{\hat{y} \hat{y}}^{2}\right|_{k}$ with, e.g., $k:=5$, there are at least 15 evaluations of the high-fidelity model and 15 evaluations of the simplified-physics low-fidelity model needed (recall § 3.1.1).

Under the common assumption that, compared to the number of evaluations of the high-fidelity model, the number of evaluations of the simplified-physics lowfidelity model is negligible, results from the proposed procedure SBO-SPLF are comparable to results such as in Table 3.12 from the proposed procedure SBO-DFLF.

### 3.3 Surrogate-guided optimization

The fundamental philosophy in the present work is to keep the sample size $m$ as low as possible since, with regard to the high-fidelity model, the acquisition of pairs of sampling plan points and output points is computationally expensive.

In order to determine the high-fidelity model's optimal solution, a reasonable sample size $m$ and a reasonable number of high-fidelity function evaluations, respectively, are unknown apriori. Furthermore, as elucidated in the elaborations of the previous section $\S 3.2$, these numbers are problem-dependent as well.

If we suppose that the positive integer number of high-fidelity function evaluations $m_{\text {DSO }}$ w.r.t. a direct solving of a high-fidelity optimization problem is higher than the positive integer number of high-fidelity function evaluations $m_{\text {SBO }}$ w.r.t. a
surrogate-based optimization, then, hopefully, the positive integer number of highfidelity function evaluations $m_{\text {SGO }}$ w.r.t. a surrogate-guided optimization is lower than $m_{\text {SBO }}$ such that the transitive relation

$$
\begin{equation*}
\forall m_{\mathrm{DSO}}, m_{\mathrm{SBO}}, m_{\mathrm{SGO}} \in \mathbb{N} \backslash\{0\} . m_{\mathrm{DSO}}>m_{\mathrm{SBO}} \wedge m_{\mathrm{SBO}}>m_{\mathrm{SGO}} \Longrightarrow m_{\mathrm{DSO}}>m_{\mathrm{SGO}} \tag{3.106}
\end{equation*}
$$

holds to be true. From an application-driven viewpoint, the additional value of checking the relation

$$
\begin{equation*}
\forall m_{\mathrm{SBO}}, m_{\mathrm{SGO}} \in \mathbb{N} \backslash\{0\} . m_{\mathrm{SBO}}>m_{\mathrm{SGO}} \tag{3.107}
\end{equation*}
$$

for a given problem is, for instance, comprehensible in the context of validation and verification.

The relation in (3.106) as well as the relation in (3.107) possess the hidden assumption that the corresponding solving procedures converge to the same optimal solution w.r.t. a user-defined tolerance. Let us encode this hidden assumption by the pre-condition

$$
\begin{equation*}
\exists m_{\mathrm{DSO}}, m_{\mathrm{SBO}}, m_{\mathrm{SGO}} \in \mathbb{N} \backslash\{0\} . m_{\mathrm{DSO}}, m_{\mathrm{SBO}}, m_{\mathrm{SGO}}>0, \tag{3.108}
\end{equation*}
$$

and let us encode the implicit order structure in (3.106) by the post-condition

$$
\begin{equation*}
\forall m_{\mathrm{DSO}}, m_{\mathrm{SBO}}, m_{\mathrm{SGO}} \in \mathbb{N} \backslash\{0\} . m_{\mathrm{DSO}}>m_{\mathrm{SBO}}>m_{\mathrm{SGO}}>0 . \tag{3.109}
\end{equation*}
$$

In those cases in which the hidden assumption is not satisfied or the relation in (3.107) does not hold, it might be more preferable to perform a surrogate-based optimization instead of a surrogate-guided optimization.

However, mind that, generally, one cannot know with certainty in advance whether the relation in (3.107) holds for a task at hand; therefore, from an applicationdriven viewpoint embedded in the context of validation and verification, it might be reasonable to perform a surrogate-based optimization as well as a surrogate-guided optimization for a task at hand.

The basic idea of surrogate-guided optimization is to provide some kind of interaction between a high-fidelity model and a low-fidelity model, that is, to provide some model management strategy (cf. § 1.2). In the subsequent subsections, let us focus on the model management strategies adaptation and fusion (cf. § 1.3).

The essential idea underlying adaptation is to exploit information about a lowfidelity model in each step of the solving procedure regarding the high-fidelity optimization problem in order to adapt the procedure according to the low-fidelity model information.

The key idea of fusion is to combine or to fuse information about the high-fidelity model and information about the low-fidelity model into a single model that is exploited to constitute a proxy for the high-fidelity optimization problem.

For the sake of completeness of the discussion about the kriging low-fidelity model in $\S 3.1 .2$, let us elaborate briefly on the so-called sequential kriging optimization which is a subtype of the model management strategy adaptation.

Next, we discuss different algorithms from the context of the space-mapping paradigm. Their underlying optimization procedures are subtypes of the model management strategy adaptation as well.

Finally, we examine the co-kriging optimization that can be conceived as a subtype of the model management strategy fusion.

Regarding all optimization approaches under consideration, we elaborate some convergence-related issues.

### 3.3.1 Sequential kriging optimization

An important feature of the kriging low-fidelity model in (3.98) is that one can provide a mean squared prediction error $\left(\hat{s}_{y}(x)\right)^{2}$ at an arbitrary point $x$. More precisely, the error $\left(\hat{s}_{y}(x)\right)^{2}$ that is associated with $\hat{y}(x)$ in (3.98) can be stated as

$$
\begin{equation*}
\left(\hat{s}_{y}(x)\right)^{2}:=\hat{\sigma}_{y}^{2}\left(1-r^{\mathrm{T}} \Psi^{-1} r+\frac{\left(1-\underline{1}^{\mathrm{T}} \Psi^{-1} r\right)^{2}}{\underline{1}^{\mathrm{T}} \Psi^{-1} \underline{1}}\right) \tag{3.110}
\end{equation*}
$$

where $\hat{s}_{y}: X \rightarrow \mathbb{R}$. Note that the term involving the fraction is negligibly small, thus, let us consider the mean squared prediction error $\left(\hat{s}_{y}(x)\right)^{2}$ as

$$
\begin{equation*}
\left(\hat{s}_{y}(x)\right)^{2}:=\hat{\sigma}_{y}^{2}\left(1-r^{\mathrm{T}} \Psi^{-1} r\right) . \tag{3.111}
\end{equation*}
$$

In addition to the numerical justification for omitting the fraction term in (3.110), the authors in [70, p. 84] provide a method-based justification, more specifically, they argue that, from a Bayesian viewpoint, the fraction term is not mentioned at all (such as in, e.g., [116, p. 280ff]).

Instead of $\left(\hat{s}_{y}(x)\right)^{2}$, it is common to employ the root mean squared prediction error $\hat{s}_{y}(x)$ as the measure of the uncertainty in the prediction $\hat{y}(x)$. The error $\hat{s}_{y}(x)$ is linked with the error $\left(\hat{s}_{y}(x)\right)^{2}$ by

$$
\begin{align*}
\hat{s}_{y}(x) & :=\sqrt{\left|\left(\hat{s}_{y}(x)\right)^{2}\right|}  \tag{3.112a}\\
& \equiv \sqrt{\left|\hat{\sigma}_{y}^{2}\left(1-r^{\mathrm{T}} \Psi^{-1} r\right)\right|} . \tag{3.112b}
\end{align*}
$$

Observe that, if a sampling plan point is given, then the error is zero. More formally, if we provide an auxiliary map $a_{\hat{s}_{y}}=x \mapsto r^{T} \Psi^{-1} r: X \rightarrow \mathbb{R}$, then one can state that

$$
\begin{equation*}
\forall i \in\{1, \ldots, m\} \cdot a_{\hat{s}_{y}}\left(x_{i}\right)=1 \Longrightarrow \forall i \in\{1, \ldots, m\} . \hat{s}_{y}\left(x_{i}\right)=0 . \tag{3.113}
\end{equation*}
$$

This observation encodes the intuitive expectation that the prediction is exact at a given sampling plan point where the corresponding output point is given as well.

Generally, if we utilize a data-fit low-fidelity model in an interpolation context, then one can reduce the empirical surrogate modeling error (recall Definition 3.1.2) by increasing sufficiently the sample size and by positioning appropriately the sampling plan points.

However, for reasons of computational thrift, the total number of sampling plan points has to be kept as low as possible. Hence, a usual economical approach is to start with a sampling plan and to add sequentially new sampling plan points in a guided way.

Ideally, this kind of adaptive interpolation balances the error-based global exploration for generating an overall accurate low-fidelity model and the predictionbased local exploitation for determining an optimal value of the high-fidelity model.

Using the error $\hat{s}_{y}(x)$, one can define the acquisition function (or infill criterion
function or update point function) expected improvement $E I: X \rightarrow \mathbf{R}^{+}$whose assignment is defined by the conditional expression

$$
E I(x):= \begin{cases}0, & \text { if } \hat{s}_{y}(x) \equiv 0  \tag{3.114}\\ (\min (y)-\hat{y}(x))\left(\frac{1}{2}+\frac{1}{2} \operatorname{erf}\left(\frac{\min (y)-\hat{y}(x)}{\hat{s}_{y}(x) \sqrt{2}}\right)\right), & \text { if } \hat{s}_{y}(x)>0, \\ +\frac{\hat{s}_{y}(x)}{\sqrt{2 \pi}} \exp \left(-\frac{(\min (y)-\hat{y}(x))^{2}}{2\left(\hat{s}_{y}(x)\right)^{2}}\right) & \end{cases}
$$

where a signature of the map $\min$ reads as $\mathbb{R}^{m \times 1} \rightarrow \mathbb{R}$ and $\min (y)$ returns the minimal output point within the current column vector $y$ in (3.83), thus, it is set that $y_{\min } \equiv \min (y)$; and the map erf : $\mathbb{R} \rightarrow[-1,1]$ denotes the Gauss error function. Notice that there is a slight abuse of notation in the sense that the expected improvement acquisition function $E I$ is regarded as an independent map instead of as the expected value of an improvement function I, i.e., $\mathbb{E}[I(x)]$. For more details, see, e.g., [108], [70, p. 89ff] or [116, p. 294ff].

The acquisition function $E I$ enables us to assert how much utility or improvement is to be expected from a potentially new sampling plan point. For other kinds of acquisition function, see, e.g., [70, ch. 3.2] or [116, ch. 16] and references therein.

After determining $y_{\min }$ and the initial kriging low-fidelity model in (3.98) w.r.t. a given sampling plan, one can identify the new $m+1$ sampling plan point as the optimal value that maximizes the corresponding expected improvement acquisition function in (3.114) - or, equivalently, as the minimizer of the optimization problem

$$
\begin{equation*}
x_{m+1}:=\underset{x \in X}{\arg \min }-E I(x) .{ }^{25} \tag{3.115}
\end{equation*}
$$

Observe that the statement in (3.115) follows a common notation in which it is not emphasized that the assignment definition $E I(x)$ in (3.114) depends on the sampling plan points and the corresponding output points as well as on the parameters of the kriging low-fidelity model.

Ideally, the iteration procedure in (3.115) of determining $y_{\text {min }}$ and the kriging low-fidelity model, and finding a new sampling plan point $x_{m+1}$ terminates after finitely many steps at the global minimum such that $\hat{s}_{y}(x) \equiv 0$, and, consequently, $E I(x) \equiv 0$.

However, the experimental rate of convergence of the iteration procedure in (3.115) might be very low or the iteration procedure might not be convergent at all (cf. [70, p. 91]). Thus, the condition in (3.107) for a given problem does not necessarily hold to be true or the condition cannot be applied because its assumption is violated.

The iteration procedure in (3.115) is associated with the so-called efficient global optimization (or sequential kriging optimization) technique which can be considered as a subtype of the model management strategy adaptation (see [166, p. 555]).

Regarding the zoo of possible acquisition functions, it is intricate to find the most appropriate embedding of concepts associated with iteration procedures such as in (3.115) into the context of numerical optimization with the magnetoquasistatic

[^35]model (recall § 2.3). Hence, concepts such as, for instance, constrained expected improvement (see, e.g., [70, ch. 5.4]) within the electromagnetics context demand a thorough examination all on their own - which is out of the scope of the present work.

In the present work, the previous considerations concerning sequential kriging optimization are primarily employed as internal checking tools at the level of programs (recall Figure 1.4).

### 3.3.2 Optimization within the space-mapping paradigm

Recalling § 2.3.2 and § 3.1.3, one can adapt the problem formulation in (2.40) to the generic statements in (3.104) in the sense that

$$
\begin{equation*}
\min _{x \in X_{0}}(\hat{j} \circ \mathrm{~K})(x), \tag{3.116}
\end{equation*}
$$

where $\mathrm{K} \in \operatorname{hom}\left(X_{0}, Y_{0}\right)$ denotes the high-fidelity model and $\hat{j} \in \operatorname{hom}\left(Y_{0}, Z_{0}\right)$ denotes the objective functional and the composition map $\circ$ is overloaded such that its signature reads as $Z_{0}^{Y_{0}} \times Y_{0}^{X_{0}} \rightarrow Z_{0}^{X_{0}}$. Let us think of (3.116) as the high-fidelity optimization problem.

Referring to the low-fidelity model as $\tilde{\mathrm{K}} \in \operatorname{hom}\left(X_{1}, Y_{1}\right)$ and overloading the objective functional $\hat{j} \in \operatorname{hom}\left(Y_{1}, Z_{1}\right)$ and the composition map $\circ \in \operatorname{hom}\left(Z_{1}^{Y_{1}} \times Y_{1}^{X_{1}}, Z_{1}^{X_{1}}\right)$ in (3.116), one can think of the low-fidelity optimization problem as

$$
\begin{equation*}
\min _{\tilde{x} \in X_{1}}(\hat{\dot{j}} \circ \tilde{K})(\tilde{x}) . \tag{3.117}
\end{equation*}
$$

Furthermore, if we refer to the domain-oriented correction map as $\tilde{\mathrm{P}} \in \operatorname{hom}\left(X_{0}, X_{1}\right)$, and to the codomain-oriented correction map as $\tilde{\mathrm{R}} \in \operatorname{hom}\left(Y_{1}, Y_{0}\right)$, then the generic statements in (3.104) reduce to one legitimate generic statement from the perspective of function extensionality such as in (2.28), that is,

$$
\begin{equation*}
\forall x \in X_{0} \cdot \mathrm{~K}(x)={\gamma_{0}}\left(\tilde{\mathrm{R}} \circ_{Y_{10}} \tilde{\mathrm{~K}} \circ_{X_{01}} \tilde{\mathrm{P}}\right)(x), \tag{3.118}
\end{equation*}
$$

where the map ${ }^{\circ}{ }_{X_{01}}: Y_{1}^{X_{1}} \times X_{1}^{X_{0}} \rightarrow Y_{1}^{X_{0}}$ and the map ${ }^{\circ} Y_{10}: Y_{0}^{Y_{1}} \times Y_{1}^{X_{1}} \rightarrow Y_{0}^{X_{1}}$ denote suitable composition maps which adhere to right-associativity. Thus, the ideal property regarding the maps $\tilde{\mathrm{R}}, \tilde{\mathrm{K}}, \tilde{\mathrm{P}}$ reads as

$$
\begin{equation*}
\forall \tilde{\mathrm{R}} \in \operatorname{hom}\left(Y_{1}, Y_{0}\right) \cdot \forall \tilde{\mathrm{K}} \in \operatorname{hom}\left(X_{1}, Y_{1}\right) \cdot \forall \tilde{\mathrm{P}} \in \operatorname{hom}\left(X_{0}, X_{1}\right) \cdot \mathrm{K}=X_{0} \rightarrow Y_{0}\left(\tilde{\mathrm{R}} \circ \bigcirc_{10} \tilde{\mathrm{~K}} \circ_{X_{01}} \tilde{\mathrm{P}}\right) . \tag{3.119}
\end{equation*}
$$

Recalling the commentary on (3.104), let us introduce the map $\tilde{\mathrm{K}}_{\mathrm{s}}$ such that

$$
\begin{equation*}
\tilde{\mathrm{K}}_{\mathrm{s}}=(\tilde{\mathrm{R}}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}) \mapsto \tilde{\mathrm{K}}_{\mathrm{s}}(\tilde{\mathrm{R}}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}):=\tilde{\mathrm{R}} \circ_{Y_{10}} \tilde{\mathrm{~K}} \circ_{X_{01}} \tilde{\mathrm{P}}: Y_{0}^{Y_{1}} \times Y_{1}^{X_{1}} \times X_{1}^{X_{0}} \rightarrow Y_{0}^{X_{0}}, \tag{3.120}
\end{equation*}
$$

in order to conceptually discriminate the notion of a low-fidelity model $\tilde{K}$ and the notion of a surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}$ within the context of the space-mapping paradigm. For the sake of notational ease, let us define

$$
\begin{equation*}
\mathrm{K}_{\mathrm{s}}^{\tilde{R}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}}:=\tilde{\mathrm{K}}_{\mathrm{s}}(\tilde{\mathrm{R}}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}) . \tag{3.121}
\end{equation*}
$$

If we add another assumption to the list of assumptions regarding the maps $\tilde{\mathrm{R}}$, $\tilde{K}$, and $\tilde{P}$, more specifically, if we assume that there is some kind of sub-structure between $X_{0}$ and $X_{1}$ as well as between $Y_{1}$ and $Y_{0}$, then it is meaningful to define the
inclusion map $\iota_{\tilde{R}}$ and the inclusion map $\iota_{\tilde{p}}$ such that

$$
\begin{align*}
& \iota_{\tilde{\mathrm{P}}}=x \mapsto \iota_{\tilde{\mathrm{p}}}(x):=x: X_{0} \rightarrow X_{1},  \tag{3.122a}\\
& \iota_{\tilde{\mathrm{R}}}=\tilde{y} \mapsto \iota_{\tilde{\mathrm{R}}}(\tilde{y}):=\tilde{y}: Y_{1} \rightarrow Y_{0} . \tag{3.122b}
\end{align*}
$$

However, if we assume that $X \equiv X_{0}$ and $X_{0} \equiv X_{1}$ as well as $Y \equiv Y_{1}$ and $Y_{1} \equiv Y_{0}$, then we receive $\iota_{\tilde{R}} \equiv \mathrm{id} Y_{Y}$ and $\iota_{\tilde{\mathrm{P}}} \equiv \mathrm{id} \mathrm{X}_{X}$. Another possible assumption is that the corresponding entities are isomorphic regarding some prescribed algebraic structure, i.e., $X_{0} \cong X_{1}$ and $Y_{1} \cong Y_{0}$. We dwell on this assumption in ch. 4, though.

Observe that if and only if the case $\iota_{\tilde{R}} \equiv \operatorname{id}_{Y}$ and $\iota_{\tilde{\mathrm{P}}} \equiv \mathrm{id}_{X}$ is given, then

$$
\begin{equation*}
\forall \tilde{\mathrm{K}} \in \operatorname{hom}(X, Y) \cdot \mathrm{K}_{\mathrm{s}}^{\mathrm{id}_{Y}, \tilde{K}, \mathrm{id}_{X}}=X \rightarrow Y \tilde{\mathrm{~K}} \tag{3.123}
\end{equation*}
$$

holds to be true such that the notion of a low-fidelity model $\tilde{K}$ and the notion of a surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}$ collapse within the context of the space-mapping paradigm.

If the condition in (3.118) and in (3.119), respectively, hold to be true, then one can substitute $K$ with $K_{s}^{\tilde{R}, \tilde{K}, \tilde{P}}$ in (3.116) such that

$$
\begin{equation*}
\min _{x \in X_{0}}\left(\hat{j} \circ\left(\tilde{\mathrm{R}} \circ \Upsilon_{10} \tilde{\mathrm{~K}} \circ_{X_{01}} \tilde{\mathrm{P}}\right)\right)(x) . \tag{3.124}
\end{equation*}
$$

However, according to the remarks regarding (2.28), one cannot expect that the condition in (3.118) and in (3.119), respectively, are satisfied for real-world applications. Thus, we have to incorporate the information about the high-fidelity model into the surrogate model to reduce a potential discrepancy between the two.

If we assume that the low-fidelity model does not depend on the high-fidelity model, then, conceptually, one would have to extend the signature of the correction map $\tilde{R}$ and the signature of the correction map $\tilde{P}$ in the sense that

$$
\begin{align*}
& \tilde{P}=(x, \mathrm{~K}, \tilde{\mathrm{~K}}) \mapsto \tilde{\mathrm{P}}(x, \mathrm{~K}, \tilde{\mathrm{~K}}): X_{0} \times Y_{0}^{X_{0}} \times Y_{1}^{X_{1}} \rightarrow X_{1},  \tag{3.125a}\\
& \tilde{R}=(\tilde{y}, \tilde{\mathrm{~K}}, \mathrm{~K}) \mapsto \tilde{\mathrm{R}}(\tilde{y}, \tilde{\mathrm{~K}}, \mathrm{~K}): Y_{1} \times Y_{1}^{X_{1}} \times Y_{0}^{X_{0}} \rightarrow Y_{0} . \tag{3.125b}
\end{align*}
$$

Though, in the common treatment of theoretical issues regarding the space-mapping paradigm, the high-fidelity model and the low-fidelity model are assumed to be fixed. Thus, they are implicitly incorporated into the assignment rules of the correction maps (see, e.g., [49, ch. 3] and references therein). This tactic resembles the definition of the empirical generalization error $\mathfrak{e}_{\mathrm{H}, s_{\S}}\left(\hat{Q}_{\xi}\right)$ in (3.21).

From a program's viewpoint (recall Figure 1.4), the low-fidelity model and the high-fidelity-model are defined within the scope of the programs associated with the correction maps.

Subsequently, I deem it advisable to move from an abstract function definition to a concrete function definition. More precisely, let us assume $X_{0} \subseteq \mathbb{R}^{d}, X_{1} \subseteq \mathbb{R}^{d}$, $Y_{0} \subseteq \mathbb{R}$, and $Y_{1} \subseteq \mathbb{R}$, that is, let us cover the multivariate scalar-valued use case w.r.t. the high-fidelity model and the low-fidelity model, then one can provide a possible definition of the assignment rule of the domain-oriented correction map P by means of

$$
\begin{equation*}
\tilde{\mathrm{P}}_{s}=x \mapsto \tilde{\mathrm{P}}_{s}(x):=\underset{\tilde{x} \in X_{1}}{\arg \min }\left(\frac{1}{2}(\tilde{\mathrm{~K}}(\tilde{x})-\mathrm{K}(x))^{2}+_{\mathbb{R}} \frac{\alpha}{2}\|\tilde{x}-x\|_{L_{2}}^{2}\right), \tag{3.126}
\end{equation*}
$$

where the index $s$ in $\tilde{\mathrm{P}}_{s}$ emphasizes the scalar-valued use case, $\alpha \in \mathbb{R}^{+}$is a userassigned smoothing parameter for the purpose of existence and uniqueness of a solution (compare with the parameter $\beta$ in (2.31)). For more details regarding the
definition in (3.126), see [95].
Let us redefine the high-fidelity model and the low-fidelity model in the sense that

$$
\begin{align*}
& \mathrm{K}=(w, x) \mapsto \mathrm{K}(w, x): W_{0} \times X_{0} \rightarrow Y_{0},  \tag{3.127a}\\
& \tilde{\mathrm{~K}}=(\tilde{w}, \tilde{x}) \mapsto \tilde{\mathrm{K}}(\tilde{w}, \tilde{x}): W_{1} \times X_{1} \rightarrow Y_{1}, \tag{3.127b}
\end{align*}
$$

then one can introduce the map $\mathrm{K}_{w}$ and the map $\tilde{\mathrm{K}}_{\tilde{w}}$ that read as

$$
\begin{align*}
& \mathrm{K}_{w}=w \mapsto(x \mapsto \mathrm{~K}(w, x)): W_{0} \rightarrow Y_{0}^{X_{0}},  \tag{3.128a}\\
& \tilde{\mathrm{~K}}_{\tilde{w}}=\tilde{w} \mapsto(\tilde{x} \mapsto \tilde{\mathrm{~K}}(\tilde{w}, \tilde{x})): W_{1} \rightarrow Y_{1}^{X_{1}}, \tag{3.128b}
\end{align*}
$$

where the operation currying is applied to the map K and $\tilde{\mathrm{K}}$. The operation currying is conceived as an operation that transforms a function with multiple arguments into a sequence of functions with single arguments.

If we suppose that $X_{0} \subseteq \mathbb{R}^{d}, X_{1} \subseteq \mathbb{R}^{d}, Y_{0} \subseteq \mathbb{R}, Y_{1} \subseteq \mathbb{R}, W_{0} \subseteq \mathbb{R}$, and $W_{1} \subseteq \mathbb{R}$, then one can consider $m_{w}$ points with $m_{w} \in \mathbb{N}$ such that one can construct lists of functions, i.e.,

$$
\begin{align*}
&\left(\mathrm{K}_{w}\left(w_{1}\right), \ldots, \mathrm{K}_{w}\left(w_{m_{w}}\right)\right) \equiv\left(x \mapsto \mathrm{~K}\left(w_{1}, x\right), \ldots, x \mapsto \mathrm{~K}\left(w_{m_{w}}, x\right)\right),  \tag{3.129a}\\
&\left({\left.\tilde{\mathrm{K}} \tilde{\tilde{w}}\left(\tilde{w}_{1}\right), \ldots, \tilde{\mathrm{K}}_{\tilde{w}}\left(\tilde{w}_{m_{w}}\right)\right)}\right) \equiv\left(\tilde{x} \mapsto \tilde{\mathrm{~K}}\left(\tilde{w}_{1}, \tilde{x}\right), \ldots, \tilde{x} \mapsto \tilde{\mathrm{~K}}\left(\tilde{w}_{m_{w}}, \tilde{x}\right)\right), \tag{3.129b}
\end{align*}
$$

where the operation $\equiv$ is conceived as a componentwise operation. Furthermore, if we evaluate the lists of functions at $x$ and at $\tilde{x}$, respectively, then one can construct lists of evaluated functions, i.e.,

$$
\begin{align*}
&\left(\mathrm{K}_{w}\left(w_{1}\right)(x), \ldots, \mathrm{K}_{w}\left(w_{m_{w}}\right)(x)\right) \equiv\left(\mathrm{K}\left(w_{1}, x\right), \ldots, \mathrm{K}\left(w_{m_{w}}, x\right)\right),  \tag{3.130a}\\
&\left(\tilde{\mathrm{K}}_{\tilde{w}}\left(\tilde{w}_{1}\right)(\tilde{x}), \ldots, \tilde{\mathrm{K}}_{\tilde{w}}\left(\tilde{w}_{m_{w}}\right)(\tilde{x})\right) \equiv\left(\tilde{\mathrm{K}}\left(\tilde{w}_{1}, \tilde{x}\right), \ldots, \tilde{\mathrm{K}}\left(\tilde{w}_{m_{w}}, \tilde{x}\right)\right) . \tag{3.130b}
\end{align*}
$$

Finally, assuming the redefinition in (3.127), one can cover the multivariate vectorvalued use case ${ }^{26}$ w.r.t. the high-fidelity model and the low-fidelity model, that is, if we assume that $X_{0} \subseteq \mathbb{R}^{d}, X_{1} \subseteq \mathbb{R}^{d}, Y_{0} \subseteq \mathbb{R}^{m_{w}}, Y_{1} \subseteq \mathbb{R}^{m_{w}}$, then one can overload the maps in (3.127) such that

$$
\begin{align*}
& \mathrm{K}=x \mapsto \mathrm{~K}(x):=\left(\mathrm{K}\left(w_{1}, x\right), \ldots, \mathrm{K}\left(w_{m_{w}}, x\right)\right): X_{0} \rightarrow Y_{0},  \tag{3.131a}\\
& \tilde{\mathrm{~K}}=\tilde{x} \mapsto \tilde{\mathrm{~K}}(\tilde{x}):=\left(\tilde{\mathrm{K}}\left(\tilde{w}_{1}, \tilde{x}\right), \ldots, \tilde{\mathrm{K}}\left(\tilde{w}_{m_{w}}, \tilde{x}\right)\right): X_{1} \rightarrow Y_{1} . \tag{3.131b}
\end{align*}
$$

Notice that the entity $w$ is often associated with the time variable $t$ such as, e.g., in (2.9); or it is associated with the frequency $\omega$ such as, e.g., in (2.10), that is, $w \equiv t$ or $w \equiv \omega$. In this context, it is usually assumed that $W_{1} \equiv W_{0}$ in (3.127). These associations constitute some common interpretations of the entity $w$ within the semantics of electromagnetics. Due to these interpretations, the case $\forall d . \forall m_{w} . d<m_{w}$ is often considered in practical applications (see, e.g., [194, p. 11] or [49, p. 91]).

In the vector-valued use case, technically, the dimensions of the domains and co-domains do not necessarily have to match, i.e., the cases $\operatorname{dim}\left(X_{0}\right) \neq \operatorname{dim}\left(X_{1}\right)$ and $\operatorname{dim}\left(Y_{0}\right) \neq \operatorname{dim}\left(Y_{1}\right)$ are conceivable. For instance, if the low-fidelity model admits an application of automatic differentiation (see $\S 2.3 .3$ ) such as in the case of data-fit low-fidelity models, then one can unleash the machinery of sensitivity computation in order to determine an importance ranking of input variables (see § 3.2.1). Hence,

[^36]by means of sensitivity computation, one could construct a low-fidelity model such that $\operatorname{dim}\left(X_{0}\right)>\operatorname{dim}\left(X_{1}\right)$. In the present work, though, this path is not pursued. For further elaborations regarding the dimension issue in the multivariate vector-valued use case, I refer to, e.g., [56, p. 60ff] or [95].

Observing (3.130), though, I conclude that, syntactically, there is another possible encoding of the vector-valued use case if it is set that

$$
\begin{align*}
&\left(\mathrm{K}_{w}\left(w_{1}\right)(x), \ldots, \mathrm{K}_{w}\left(w_{m_{w}}\right)(x)\right) \equiv\left(\mathrm{K}_{1}(x), \ldots, \mathrm{K}_{m_{w}}(x)\right),  \tag{3.132a}\\
&\left(\tilde{\mathrm{K}}_{\tilde{w}}\left(\tilde{w}_{1}\right)(\tilde{x}), \ldots, \tilde{\mathrm{K}}_{\tilde{w}}\left(\tilde{w}_{m_{w}}\right)(\tilde{x})\right) \equiv\left(\tilde{\mathrm{K}}_{1}(\tilde{x}), \ldots, \tilde{\mathrm{K}}_{m_{w}}(\tilde{x})\right), \tag{3.132b}
\end{align*}
$$

then one can rewrite (3.131) as

$$
\begin{align*}
& \mathrm{K}=x \mapsto \mathrm{~K}(x):=\left(\mathrm{K}_{1}(x), \ldots, \mathrm{K}_{m_{v}}(x)\right): X_{0} \rightarrow Y_{0},  \tag{3.133a}\\
& \tilde{\mathrm{~K}}=\tilde{x} \mapsto \tilde{\mathrm{~K}}(\tilde{x}):=\left(\tilde{\mathrm{K}}_{1}(\tilde{x}), \ldots, \tilde{\mathrm{K}}_{m_{v v}}(\tilde{x})\right): X_{1} \rightarrow Y_{1}, \tag{3.133b}
\end{align*}
$$

Mind that if we change the definitions in (3.128) such that it is set $W_{1} \equiv W_{0}$, and therefore, in (3.133), $\mathrm{K}_{w}$ and $\tilde{\mathrm{K}}_{w}$ with $w \in\left\{1, \ldots, m_{w}\right\}$ denote $m_{w}$ different component high-fidelity models and $m_{w}$ different component low-fidelity models, respectively, then the corresponding interpretation resembles partly multiobjective optimization (recall § 1.1). For some applications of this kind of interpretation of the vector-valued use case, I refer to, e.g., [56, ch. 5] or [49, ch. 6].

Given the definitions in (3.131), the assignment rule in (3.126) has to be adapted, more precisely, given $X_{0} \subseteq \mathbb{R}^{d}$ and $X_{1} \subseteq \mathbb{R}^{d}$, then one can implement the map $\tilde{\mathrm{P}}$ by

$$
\begin{equation*}
\tilde{\mathrm{P}}_{v}=x \mapsto \tilde{\mathrm{P}}_{v}(x):=\underset{\tilde{x} \in X_{1}}{\arg \min }\left(\frac{1}{2}\|\tilde{\mathrm{~K}}(\tilde{x})-\mathrm{K}(x)\|_{l_{2}}^{2}+\mathbb{R} \frac{\alpha}{2}\|\tilde{x}-x\|_{l_{2}}^{2}\right), \tag{3.134}
\end{equation*}
$$

where the index $v$ in $\tilde{P}_{v}$ emphasizes the vector-valued use case and the corresponding entities are conceived as column vectors (see also the commentary on the representation of vectors as column vectors in § 3.1.2). For other possible definitions of $\tilde{\mathrm{P}}_{v}(x)$, see, e.g., [49, p. 65] and references therein.

Depending on the choice of interpretation in (3.131) or in (3.133), one can recover the scalar-valued use case by setting $m_{w}:=1$ such that that $\tilde{\mathrm{P}}_{v} \equiv \tilde{\mathrm{P}}_{s}$.

Assuming some kind of differentiability structure regarding the map $\tilde{P}$ (cf. [49, p. 64]), one can invoke the notion of a first-order Taylor series expansion for multivariate vector-valued functions, thus, one can define an affine map as a representative of the map $\tilde{P}$ in (3.125a), that is,

$$
\begin{equation*}
\tilde{\mathrm{P}}=x \mapsto \tilde{\mathrm{P}}(x):=\tilde{\mathrm{P}}\left(x^{0}\right)+J_{\tilde{\mathrm{P}}}\left(x^{0}\right)\left(x-x^{0}\right), \tag{3.135}
\end{equation*}
$$

where, in the context of the space-mapping paradigm, the value of $\tilde{\mathrm{P}}$ at the expansion point $x^{0}$ is chosen as $\tilde{\mathrm{P}}\left(x^{0}\right):=\tilde{\mathrm{P}}_{v}\left(x^{0}\right)$; and $J_{\tilde{\mathrm{P}}}\left(x^{0}\right) \in \mathbb{R}^{d \times d}$ denotes the Jacobi matrix w.r.t. the domain-oriented correction map $\tilde{P}$ in (3.125a) and evaluated at a fixed argument $x^{0}$ of the high-fidelity model. ${ }^{27}$

[^37]By assuming some kind of sub-structure such as in (3.122), one can determine $J_{\tilde{\mathrm{P}}}\left(x^{0}\right)$ by the Jacobi matrix $J_{\tilde{\mathrm{K}}}\left(\tilde{\mathrm{P}}_{v}\left(x^{0}\right)\right) \in \mathbb{R}^{m_{w} \times d}$ w.r.t. the low-fidelity model evaluated at $\tilde{\mathrm{P}}_{v}\left(x^{0}\right)$ and by the Jacobi matrix $J_{\tilde{\mathrm{K}}_{\mathrm{s}}\left(\tau_{\tilde{R}}, \tilde{K}, \tilde{\mathrm{P}}\right)}\left(x^{0}\right) \in \mathbb{R}^{m_{w} \times d}$ w.r.t. the surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}\left(\iota_{\mathrm{R}}, \tilde{\mathrm{K}}, \tilde{\mathrm{P}}\right)$ and evaluated at $x^{0}$.

Utilizing a multivariate vector-valued version of the chain rule, the Jacobi matrix $J_{\tilde{p}}$ reads as

$$
\begin{equation*}
J_{\tilde{\mathrm{P}}}\left(x^{0}\right):=J_{\tilde{\mathrm{K}}}^{+}\left(\tilde{\mathrm{P}}_{v}\left(x^{0}\right)\right) J_{\tilde{\mathrm{K}}_{s}\left(\iota_{\mathrm{R}}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}\right)}\left(x^{0}\right), \tag{3.136}
\end{equation*}
$$

where $J_{\tilde{\mathrm{K}}}^{+}\left(\tilde{\mathrm{P}}_{v}\left(x^{0}\right)\right) \in \mathbb{R}^{d \times m_{w}}$ indicates the pseudoinverse of $J_{\tilde{\mathrm{K}}}\left(\tilde{\mathrm{P}}_{v}\left(x^{0}\right)\right)$ possessing at least a left inverse characteristic, i.e., $J_{\tilde{\mathrm{K}}}^{+}\left(\tilde{\mathrm{P}}_{v}\left(x^{0}\right)\right) J_{\tilde{\mathrm{K}}}\left(\tilde{\mathrm{P}}_{v}\left(x^{0}\right)\right) \equiv I$ with $I \in \mathbb{R}^{d \times d}$ being the identity matrix. The definition of the pseudoinverse follows the definition in (3.61).

Furthermore, by assuming some kind of sub-structure such as in (3.122) and by assuming $Y_{0} \subseteq \mathbb{R}^{m_{w}}, Y_{1} \subseteq \mathbb{R}^{m_{w}}$, one can define an ideal affine map as a representative of the map $\tilde{R}$, that is,

$$
\begin{equation*}
\tilde{\mathrm{R}}=\tilde{\mathrm{K}}(\tilde{x}) \mapsto \tilde{\mathrm{R}}(\tilde{\mathrm{~K}}(\tilde{x})):=\mathrm{K}\left(x^{*}\right)+_{\mathbb{R}^{m_{w}}} S\left(\tilde{\mathrm{~K}}(\tilde{x})-\tilde{\mathrm{K}}\left(\iota_{\tilde{\mathrm{p}}}\left(x^{*}\right)\right)\right), \tag{3.137}
\end{equation*}
$$

where $S \in \mathbb{R}^{m_{w} \times m_{w}}$ denotes the ideal rotation matrix that is defined by the Jacobian matrix $J_{K}\left(x^{*}\right) \in \mathbb{R}^{m_{w} \times d}$ w.r.t. the high-fidelity model K and evaluated at its optimal argument $x^{*}$ and by the Jacobian matrix $J_{\tilde{\mathrm{K}}}\left(\iota_{\tilde{\mathrm{P}}}\left(x^{*}\right)\right) \in \mathbb{R}^{m_{w} \times d}$ w.r.t. the low-fidelity model $\tilde{\mathrm{K}}$ and evaluated at $\iota_{\tilde{\mathrm{p}}}\left(x^{*}\right)$.

Assuming that (3.118) adapted to the point $x^{*}$ holds to be true for the surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}\left(\tilde{\mathrm{R}}, \tilde{\mathrm{K}}, \tau_{\tilde{\mathrm{P}}}\right)$ and utilizing a multivariate vector-valued version of the chain rule, the ideal rotation matrix $S \equiv J_{\tilde{\mathrm{R}}}\left(\tilde{\mathrm{K}}\left(\iota_{\tilde{\mathrm{p}}}\left(x^{*}\right)\right)\right)$ reads as

$$
\begin{equation*}
S:=J_{\kappa}\left(x^{*}\right) J_{\tilde{k}}^{+}\left(\iota_{\tilde{p}}\left(x^{*}\right)\right), \tag{3.138}
\end{equation*}
$$

where $J_{\tilde{\mathrm{K}}}^{+}\left(\iota_{\tilde{\mathrm{p}}}\left(x^{*}\right)\right) \in \mathbb{R}^{d \times m_{w}}$ indicates the pseudoinverse of $J_{\tilde{\mathrm{K}}}\left(\iota_{\tilde{\mathrm{P}}}\left(x^{*}\right)\right)$ possessing at least a right inverse characteristic, i.e., $J_{\tilde{\mathrm{K}}}\left(\iota_{\tilde{\rho}}\left(x^{*}\right)\right) J_{\hat{\mathrm{K}}}^{+}\left(\iota_{\tilde{\mathrm{p}}}\left(x^{*}\right)\right) \equiv I$ with $I \in \mathbb{R}^{m_{w w} \times m_{w w}}$ being the identity matrix. Hence, the definition of the pseudoinverse in (3.61) is adapted to the case

$$
\begin{equation*}
J_{\tilde{\mathrm{K}}}^{+}\left(l_{\tilde{\mathrm{P}}}\left(x^{*}\right)\right):=J_{\tilde{\mathrm{K}}}\left(l_{\tilde{\mathrm{P}}}\left(x^{*}\right)\right)^{\mathrm{T}}\left(J_{\tilde{\mathrm{K}}}\left(l_{\stackrel{\mathrm{L}}{ }}\left(x^{*}\right)\right) J_{\tilde{\mathrm{K}}}\left(\iota_{\tilde{\mathrm{P}}}\left(x^{*}\right)\right)^{\mathrm{T}}\right)^{-1} . \tag{3.139}
\end{equation*}
$$

Note that the attribute "ideal" reflects the fact that one cannot know apriori the optimal solution regarding the high-fidelity model optimization problem. For more details regarding the ideal map $\tilde{R}$ in (3.137), I refer to, e.g., [56, p. 44ff].

From an algorithmic viewpoint (recall Figure 1.4 and recall § 2.3.3), the statement in (3.124) constitutes the anchor point at the map level for any optimization algorithm that follows the space-mapping paradigm.

Thus, one can articulate the essential aim of the corresponding iteration procedures by

$$
\begin{equation*}
x^{(k+1)}:=\underset{x \in X_{0}}{\arg \min }\left(\hat{j} \circ\left(\tilde{\mathrm{R}} \circ \Upsilon_{10} \tilde{\mathrm{~K}} \circ \circ_{01} \tilde{\mathrm{P}}\right)\right)\left(x^{(k)}\right), \tag{3.140a}
\end{equation*}
$$

where $k \in \mathbb{N}$ and $x^{k}$ and $x^{k+1}$ denote the $k$-th iteration point and the $k+1$-th iteration point, respectively, such that, with regard to some appropriate norm,

$$
\begin{equation*}
x^{(k)} \rightarrow x^{*} \text { as } k \rightarrow \infty, \tag{3.140b}
\end{equation*}
$$

where $x^{*}$ refers to an existing optimal solution of the high-fidelity optimization problem in (3.116), more precisely,

$$
\begin{equation*}
x^{*} \in \underset{x \in X_{0}}{\arg \min }(\hat{j} \circ \mathrm{~K})(x) . \tag{3.140c}
\end{equation*}
$$

Analogous to (3.140c), one can refer to $\tilde{x}^{*}$ as an existing optimal solution of the lowfidelity optimization problem in (3.117), more precisely,

$$
\begin{equation*}
\tilde{x}^{*} \in \underset{\tilde{x} \in X_{1}}{\arg \min }(\hat{\hat{j}} \circ \tilde{K})(\tilde{x}) . \tag{3.141}
\end{equation*}
$$

Observe that, in the scalar-valued use case and the vector-valued use case, it is additionally supposed that $Z_{0} \subseteq \mathbb{R}^{+}$and $Z_{1} \subseteq \mathbb{R}^{+}$in (3.116) and in (3.117), respectively.

If we invoke an object $X_{1_{1}}$ such that there is some kind of substructure between $X_{1_{1}}$ and $X_{1}$, then one can utilize the inclusion map $\iota_{\tilde{p}}$ in (3.122a) in order to define a preimage of $X_{1_{1}}$ under $\iota_{\tilde{p}}$ in the sense that

$$
\begin{equation*}
\iota_{\stackrel{\mathrm{p}}{-1}}^{-1}\left(X_{1_{1}}\right):=\left\{x \in X_{0} \mid \iota_{\tilde{\mathrm{P}}}(x) \in X_{1_{1}}\right\} . \tag{3.142}
\end{equation*}
$$

If we suppose that $\tilde{x}^{*} \in X_{1_{1}} \subset X_{1}$, then one can define a proposal for an initial iteration point $x^{(0)}$ by

$$
\begin{equation*}
\iota_{\tilde{p}}\left(x^{(0)}\right):=\tilde{x}^{*} \tag{3.143}
\end{equation*}
$$

such that $x^{(0)} \in \iota_{\tilde{\mathrm{P}}}^{-1}\left(X_{1_{1}}\right)$. Bear in mind that other initial iteration points are plausible, too, since they are commonly problem dependent. However, the choice in (3.143) appears heuristically as a promising decision in order to, hopefully, keep the total number of iterations as low as possible.

In the course of the years, many different algorithms have been presented to achieve the essential aim in (3.140). For elaborations on a large portion of corresponding optimization algorithms, see, e.g., [49, ch. 3] or [125] and references therein.

In the present work, though, solely a small subset of the large class of optimization algorithms within the space mapping paradigm is considered. Notice that those algorithms are regarded as algorithms within the space-mapping paradigm that conceptually distinguish a low-fidelity model $\tilde{K}$ and a surrogate model $K_{s}^{\tilde{R}, K, \tilde{P}}$ in (3.121) at the function level (recall Figure 1.4).

From the small subset of algorithms under consideration, let us focus primarily on a Trust Region Aggressive Space Mapping (TRASM) algorithm which assumes the surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}\left(\iota_{\tilde{\mathrm{R}}}, \tilde{\mathrm{K}}, \tilde{\mathrm{P}}\right)$ that is described with regard to (3.136).

Mind that the present work's version of the TRASM algorithm, that is, algorithm 3.1, builds upon the discussion in [95] and in [49, p. 67ff]) and extends their considerations by the context of a more general set of admissible solutions and the context of the Julia PL. A main novel use case of the TRASM algorithm 3.1 of the present work is its combination with a co-kriging low-fidelity model that we encounter in the next subsection.

Similarly to the proposed procedures SBO-DFLF and SBO-SPLF, the TRASM algorithm 3.1 builds upon the common canon of optimization algorithms (recall § 2.3.3).

The TRASM algorithm's basic building blocks are well covered by the theory of trust-region methods within nonlinear optimization (see, e.g., [158, ch. 4]). An essential overriding motivation of invoking a trust-region scaffolding for an optimization
algorithm is to equip the algorithm with good global convergence guarantees, that is, to ensure that any remote starting point will eventually converge, in the unconstrained case, to a stationary accumulation point, and, in the constrained case, to a KKT point where a KKT point is conceived as a stationary accumulation point that satisfies the KKT conditions (see § 2.3.1). Regarding the experimental rate of convergence, the corresponding equipped optimization algorithms exhibit satisfactory practical performance.

Let us assume some additional structure w.r.t. the domains and codomains of the respective models; more precisely, let us suppose some kind of topological vector space structure equipped with some metric structure, some norm structure, and some inner product structure. Hence, for the sake of simplicity, it is presupposed that the above-mentioned vector-valued use case incorporates all the structure needed. Furthermore, it is supposed that $\tilde{\mathrm{K}}_{\mathrm{s}}\left(\iota_{\tilde{\mathrm{R}}}, \tilde{\mathrm{K}}, \tilde{\mathrm{P}}\right) \equiv \tilde{\mathrm{K}}_{\mathrm{s}}\left(\mathrm{id}_{Y}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}\right)$ w.r.t. (3.122).

Let us define the $k+1$-th iteration point $x^{(k+1)}$ as

$$
\begin{equation*}
x^{(k+1)}:=x^{(k)}+h^{(k)}, \tag{3.144}
\end{equation*}
$$

where $h^{(k)} \in X_{0} \subseteq \mathbb{R}^{d}$ denotes the $k$-th step from $x^{(k)}$ to $x^{(k+1)}$ in which the step's direction and the step's length are encoded.

The definition in (3.135) is adapted in the sense that the identification $\left(x-x^{0}\right):=$ $h^{(k)}$ is made, it is set that $\tilde{\mathrm{P}}\left(x^{0}\right):=\tilde{\mathrm{P}}\left(x^{k}\right)$ and $\tilde{\mathrm{P}}_{v}\left(x^{0}\right):=\tilde{\mathrm{P}}_{v}\left(x^{k}\right)$ such that $\tilde{\mathrm{P}}\left(x^{k}\right):=$ $\tilde{\mathrm{P}}_{v}\left(x^{k}\right)$; and, finally, the Jacobi matrix $J_{\tilde{\mathrm{P}}}\left(x^{0}\right)$ is approximated by means of the Broyden's method for solving nonlinear equations (see, e.g., [158, p. 279-283]), i.e.,

$$
\begin{equation*}
J_{\tilde{\mathrm{P}}}\left(x^{0}\right):=B^{(k)}, \quad B^{(0)}:=I, \quad B^{(k+1)}:=B^{(k)}+_{\mathbb{R}^{d \times d}}\left(y_{\tilde{\mathrm{P}}_{v}}^{(k)}+\mathbb{R}^{d} B^{(k)} h^{(k)}\right) \otimes \frac{h^{(k)}}{\left\|h^{(k)}\right\|_{l_{2}}^{2}}, \tag{3.145}
\end{equation*}
$$

where $B^{(k)} \in \mathbb{R}^{d \times d}$ is referred to as the $k$-th iteration Broyden's matrix, the map $\otimes$ is conceived as the outer product w.r.t. two column vectors and the map $\otimes$ is granted a higher precedence than the map $+_{\mathbb{R}^{d \times d}}, I \in \mathbb{R}^{d \times d}$ is the identity matrix as a representative of a non-singular matrix and $y_{\tilde{\mathrm{P}}_{v}}^{(k)} \in \mathbb{R}^{d}$ denotes the change in $\tilde{\mathrm{P}}_{v}$ in (3.134) w.r.t. the step $h^{(k)}$, that is,

$$
\begin{equation*}
y_{\tilde{\mathrm{P}}_{v}}^{(k)}:=\tilde{\mathrm{P}}_{v}\left(x^{(k+1)}\right)-\tilde{\mathrm{P}}_{v}\left(x^{(k)}\right) . \tag{3.146}
\end{equation*}
$$

Hence, the adaptation of the definition in (3.135) reads as

$$
\begin{equation*}
\tilde{\mathrm{P}}=x^{(k+1)} \mapsto \tilde{\mathrm{P}}\left(x^{(k+1)}\right):=\tilde{\mathrm{P}}_{v}\left(x^{(k)}\right)+B^{(k)} h^{(k)} . \tag{3.147}
\end{equation*}
$$

By moving from the definition in (3.135) to its adaptation in (3.147), one can obtain the $k$-th step $h^{(k)}$ by solving the trust-region optimization sub-problem regarding (3.140) in which the corresponding model function is set to be the surrogate $\operatorname{model} \tilde{\mathrm{K}}_{\mathrm{s}}\left(\operatorname{id}_{Y_{0}}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}\right) \equiv \operatorname{id}_{Y_{0}}{ }^{\circ}{ }_{Y_{10}} \tilde{\mathrm{~K}}{ }^{\circ}{ }_{X_{01}} \tilde{\mathrm{P}}$ w.r.t. $h^{(k)}$, that is,

$$
\begin{equation*}
\min _{h^{(k)} \in F_{0}^{(k)}}\left(\hat{\hat{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \Upsilon_{10} \tilde{\mathrm{~K}} \circ \mathrm{X}_{01} \tilde{\mathrm{P}}\right)\right)\left(x^{(k)}+h^{(k)}\right) \stackrel{(3.147)}{\equiv} \min _{h^{(k)} \in F_{0}^{(k)}}\left(\hat{\hat{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \Upsilon_{10} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{\mathrm{P}}_{v}\left(x^{(k)}\right)+B^{(k)} h^{(k)}\right), \tag{3.148}
\end{equation*}
$$

where $F_{0}^{(k)} \subseteq \mathbb{R}^{d}$ incorporates the inequality constraint function $c_{1}$ that reads as

$$
\begin{equation*}
c_{1}=h^{(k)} \mapsto c_{1}\left(h^{(k)}\right):=\left\|D h^{(k)}\right\|_{l_{2}}-\Delta^{(k)}: X_{0} \rightarrow \mathbb{R}^{-}, \tag{3.149}
\end{equation*}
$$

with $\forall h^{(k)} . c_{1}\left(h^{(k)}\right) \leq 0$ and $\left\|D h^{(k)}\right\|_{l_{2}}$ being the trust-region in the $l_{2}$-norm and $\Delta^{(k)} \in \mathbb{R}^{+}$ being the trust-region radius and $D \in \mathbb{R}^{d \times d}$ being a diagonal matrix with positive entries that enables scaling in order to potentially enhance the solving process (cf. [158, p. 95ff]).

Choosing adequately the trust-region radius $\Delta^{(k)}$ in each iteration is an important part of an optimization algorithm based on the trust-region theory (cf. [158, p. 68]). In this context, let us define the trust-region reduction quotient $\rho^{(k)} \in \mathbb{R}$, that is,

$$
\begin{equation*}
\rho^{(k)}:=\frac{\operatorname{ared}\left(x^{(k)}, h^{(k)}\right)}{\operatorname{pred}\left(x^{(k)}, h^{(k)}\right)} \tag{3.150}
\end{equation*}
$$

where ared : $Z_{0}^{X_{0} \times X_{0}}$ denotes the actual reduction function and pred : $Z_{0}^{X_{0} \times X_{0}}$ denotes the predicted reduction function whose assignment rules read as

$$
\begin{align*}
\operatorname{ared}\left(x^{(k)}, h^{(k)}\right):=\left(\hat{\tilde{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \bigcirc_{Y_{10}} \tilde{\mathrm{~K}} \circ_{X_{01}} \tilde{\mathrm{P}}\right)\right)\left(x^{(k)}\right)-\left(\hat{\mathrm{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \bigcirc_{Y_{10}} \tilde{\mathrm{~K}} \circ_{X_{01}} \tilde{\mathrm{P}}\right)\right)\left(x^{(k)}+h^{(k)}\right),  \tag{3.151a}\\
\operatorname{pred}\left(x^{(k)}, h^{(k)}\right):=\left(\hat{\mathrm{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \circ_{Y_{10}} \tilde{\mathrm{~K}} \circ{ }_{X_{01}} \tilde{\mathrm{P}}\right)\right)\left(x^{(k)}\right)-\left(\hat{\mathrm{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \circ_{Y_{10}} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{\mathrm{P}}_{v}\left(x^{(k)}\right)+B^{(k)} h^{(k)}\right) . \tag{3.151b}
\end{align*}
$$

For the sake of the implementation of the TRASM algorithm 3.1, let us overload straightforwardly the function signature of ared by $Z_{0}^{X_{1} \times X_{1}}$ and the function signature of pred by $Z_{0}^{X_{1} \times X_{0}}$ such that the assignment rules read as

$$
\begin{align*}
\operatorname{ared}\left(\tilde{x}^{(k)}, \tilde{x}^{(k+1)}\right) & :=\left(\hat{\dot{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \gamma_{10} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{x}^{(k)}\right)-\left(\hat{\mathrm{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \gamma_{10} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{x}^{(k+1)}\right),  \tag{3.152a}\\
\operatorname{pred}\left(\tilde{x}^{(k)}, h^{(k)}\right): & :\left(\hat{\dot{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \gamma_{10} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{x}^{(k)}\right)-\left(\hat{\mathrm{j}} \circ\left(\operatorname{id}_{Y_{0}} \circ \gamma_{10} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{x}^{(k)}+B^{(k)} h^{(k)}\right), \tag{3.152b}
\end{align*}
$$

where $\tilde{x}^{(k)}:=\tilde{\mathrm{P}}_{v}\left(x^{(k)}\right)$ and $\tilde{x}^{(k+1)}:=\tilde{\mathrm{P}}_{v}\left(x^{(k)}+h^{(k)}\right)$.
Note that $\rho^{(k)}$ in (3.150) quantifies the degree of justification for the identification in (3.148) since the genuine nature of the statement in (3.147) and the statement in (3.135), respectively, is approximate instead of exact.

Given a trust-region reduction threshold $\left.\eta_{1} \in\right] 0,1[$ and a trust-region reduction threshold $\left.\eta_{2} \in\right] 0,1\left[\right.$ with $\forall \eta_{1}, \eta_{2} . \eta_{1}<\eta_{2}$, a trust-region reduction factor $\left.\gamma \in\right] 0,1[$, and a trustregion augmentation factor $\zeta \epsilon] 1, \infty[$, let us distinguish three cases regarding the $k+1$ th iteration trust-region radius $\Delta^{(k+1)}$, i.e.,

$$
\Delta^{(k+1)}:= \begin{cases}\gamma \Delta^{(k)}, & \text { if } \rho^{(k)}<\eta_{1}  \tag{3.153}\\ \Delta^{(k)}, & \text { if } \eta_{1} \leq \rho^{(k)}<\eta_{2} \\ \zeta \Delta^{(k)}, & \text { if } \rho^{(k)} \geq \eta_{2}\end{cases}
$$

Thus, depending on the value of $\rho^{(k)}$, the trust-region radius is decreased or unchanged or increased. Furthermore, if the trust-region radius is decreased, then the $k+1$-th iteration point $x^{(k+1)}$ in (3.144) and the $k+1$-th iteration Broyden's matrix $B^{(k+1)}$ in (3.145) are not accepted, more precisely, $x^{(k+1)}:=x^{(k)}$ and $B^{(k+1)}:=B^{(k)}$; otherwise they are accepted. For other strategies of updating $\Delta^{(k+1)}$, see, e.g., [49, p. 68] or [158, p. 69].

Mind that $F_{0}^{(k)}$ in (3.148) incorporates contingently also some other constraints (recall § 2.3.2) such as, e.g., box constraints (see § 2.3.2), inherited from the highfidelity model's domain. Hence, the corresponding inequality constraint functions $c_{2}$
and $c_{3}$ read as

$$
\begin{align*}
& c_{2}=h^{(k)} \mapsto c_{2}\left(h^{(k)}\right):=x_{l}-\left(x^{(k)}+h^{(k)}\right): X_{0} \rightarrow X_{0}  \tag{3.154a}\\
& c_{3}=h^{(k)} \mapsto c_{3}\left(h^{(k)}\right):=\left(x^{(k)}+h^{(k)}\right)-x_{u}: X_{0} \rightarrow X_{0}, \tag{3.154b}
\end{align*}
$$

where $\forall h^{(k)} \cdot c_{2}\left(h^{(k)}\right) \leq 0, \forall h^{(k)} . c_{3}\left(h^{(k)}\right) \leq 0$, and $x_{l} \in X_{0} \subseteq \mathbb{R}^{d}$ includes the lower bouds w.r.t. $x$, and $x_{u} \in X_{0} \subseteq \mathbb{R}^{d}$ includes the upper bounds w.r.t. $x$. Note that, e.g., in the case of box constraints, it could be useful to change the norm from the $l_{2}$ - to the $l_{\infty}$ or to the $l_{1}$-norm (cf. [158, p. 97]).

Given an evaluated quantity of interest which is supposed to be computationally expensive, if we incorporate this quantity in the constraints (recall § 2.3.1 and $\S 2.3 .3$ ), then, from an application-driven viewpoint, it might be advisable to replace this quantity by a low-fidelity model as well. Depending on the computational costs of a simplified-physics low-fidelity model in terms of memory storage and evaluation time, it can be more favorable to invoke a data-fit low-fidelity model than a simplified-physics low-fidelity model.

Finally, the $k$-th step $h^{(k)}$ in (3.144) is computed by

$$
\begin{equation*}
h^{(k)}:=\underset{h \in F_{0}^{(k)}}{\arg \min }\left(\hat{j} \circ\left(\mathrm{id}_{Y_{0}} \circ \Upsilon_{10} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{\mathrm{P}}_{v}\left(x^{(k)}\right)+B^{(k)} h\right) .{ }^{28} \tag{3.155}
\end{equation*}
$$

In order to translate the basic building blocks of the TRASM algorithm into a practical implementation, i.e., into a program (recall Figure 1.4), one has to provide at least one termination criterion. Mind that applying proper termination criteria for practical purposes is an intricate endeavor.

Let us consider the maximal number $k_{\max } \in \mathbb{N}$ of iteration points in (3.144) as one termination criterion, more precisely, the condition

$$
\begin{equation*}
\forall k . k<k_{\max } \tag{3.156}
\end{equation*}
$$

has to be true, otherwise the algorithm terminates.
Furthermore, let us consider the norm of the $k$-th step $h^{(k)}$ relative to the norm of the $k$-th iteration point $x^{(k)}$ in (3.144) such that

$$
\begin{equation*}
\frac{\left\|x^{(k+1)}-x^{(k)}\right\|_{l_{2}}}{\left\|x^{(k)}\right\|_{l_{2}}} \equiv \frac{\left\|h^{(k)}\right\|_{l_{2}}}{\left\|x^{(k)}\right\|_{l_{2}}} \tag{3.157}
\end{equation*}
$$

Given an absolute threshold w.r.t. the norm of the step $h^{(k)}$, i.e., $\left.\epsilon_{\text {abs }} \in\right] 0,1[$, and an relative threshold w.r.t. the norm of the step $h^{(k)}$, i.e., $\left.\epsilon_{\text {rel }} \epsilon\right] 0,1[$, where it is demanded that $\forall \epsilon_{\text {abs }}, \epsilon_{\text {rel }} \cdot \epsilon_{\text {abs }}<\epsilon_{\text {rel }}$, then one can reformulate (3.157) such that

$$
\begin{equation*}
\forall x^{(k)}, h^{(k)} \cdot \frac{\left\|h^{(k)}\right\|_{l_{2}}}{\left\|x^{(k)}\right\|_{l_{2}}}>\epsilon_{\mathrm{rel}}+\frac{\epsilon_{\mathrm{abs}}}{\left\|x^{(k)}\right\|_{l^{2}}} \tag{3.158}
\end{equation*}
$$

[^38]```
Algorithm 3.1: Trust Region Aggressive Space Mapping (TRASM)
\# Input:
    \(x^{(0)} \# \in \mathbb{R}^{d}\)... initial solution \# (3.143)
    \(B^{(0)} \# \in \mathbb{R}^{d \times d}\)... initial Broyden's matrix \(\#(3.145)\)
    \(\Delta^{(0)} \# \in \mathbb{R}^{+}\)... initial trust-region radius \#(3.153)
    \(\left.\eta_{1}, \eta_{2}, \gamma, \zeta \# \in\right] 0,1[\times] 0,1[\times] 0,1[\times] 1, \infty[\quad \#(3.153)\)
    \(F_{0}^{(0)} \# \subseteq \mathbb{R}^{d}\)... initial set of admissible solutions \#(3.149) \# (3.154)
    \(\left.\left.k_{\max } \# \in\right] 0,100\right]\)... maximal number of iterations \# (3.156)
    \(\left.\epsilon_{\text {abs }} \#\right] 0,1\left[\ldots\right.\) absolute threshold w.r.t. the norm of the step \(h^{(k)}\) \# (3.158)
    \(\left.\epsilon_{\text {rel }} \# \in\right] 0,1\left[\right.\)... relative threshold w.r.r.t. the norm of the step \(h^{(k)}\) \# (3.158)
\# Output:
    \(x^{(k+1)} \# \in \mathbb{R}^{d}\)... optimal solution after \(k+1\) iterations
    for \(k\) in \(0: 1: k_{\max } \quad\) \#(3.156)
    \(\tilde{x}^{(k)}:=\tilde{\mathrm{P}}_{v}\left(x^{(k)}\right) \quad\) \#(3.134) \# a high-fidelity model evaluation
    \(h^{(k)}:=\arg \min _{h \in F_{0}^{(k)}}\left(\hat{\hat{j}} \circ\left(\mathrm{id}_{Y_{0}} \circ{ }_{Y_{10}} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{x}^{(k)}+B^{(k)} h\right) \quad \#(3.155)\)
    \(\tilde{x}^{(k+1)}:=\tilde{\mathrm{P}}_{v}\left(x^{(k)}+h^{(k)}\right) \quad\) \#(3.134) \# a high-fidelity model evaluation
    \(\rho^{(k)}:=\frac{\operatorname{ared}\left(\tilde{x}^{(k)}, \tilde{x}^{(k+1)}\right)}{\operatorname{pred}\left(\tilde{x}^{(k)}, h^{(k)}\right) \quad \#(3.152)}\)
    if \(\rho^{(k)}<\eta_{1}\) \#(3.153)
                \(x^{(k+1)}:=x^{(k)}\)
                \(B^{(k+1)}:=B^{(k)}\)
                \(\Delta^{(k+1)}:=\gamma \Delta^{(k)}\)
            elseif \(\eta_{1} \leq \rho^{(k)}\) and \(\rho^{(k)}<\eta_{2} \quad \#(3.153)\)
                \(x^{(k+1)}:=x^{(k)}+h^{(k)} \quad \#\) (3.144)
                \(y_{\tilde{\mathrm{P}}_{v}}^{(k)}:=\tilde{x}^{(k+1)}-\tilde{x}^{(k)} \quad \#(3.146)\)
                \(B^{(k+1)}:=B^{(k)}+\mathbb{R}^{d \times d}\left(y_{\hat{P}_{v}}^{(k)}+\mathbb{R}^{d} B^{(k)} h^{(k)}\right) \otimes \frac{h^{(k)}}{\left\|h^{(k)}\right\|_{1_{2}}^{2}} \quad \#(3.145)\)
                \(\Delta^{(k+1)}:=\Delta^{(k)}\)
            else \# (3.153)
                \(x^{(k+1)}:=x^{(k)}+h^{(k)} \quad \#\) (3.144)
                \(y_{\tilde{\mathrm{P}}_{v}}^{(k)}:=\tilde{x}^{(k+1)}-\tilde{x}^{(k)} \quad \#(3.146)\)
                \(B^{(k+1)}:=B^{(k)}+_{\mathbb{R}^{d \times d}}\left(y_{\stackrel{P}{v}^{2}}^{(k)}+\mathbb{R}^{d} B^{(k)} h^{(k)}\right) \otimes \frac{h^{(k)}}{\left\|h^{(k)}\right\|_{L_{2}}^{2}} \quad \#(3.145)\)
                \(\Delta^{(k+1)}:=\zeta \Delta^{(k)}\)
            end \#if
        if \(\frac{\left\|h^{(k)}\right\|_{I^{2}}}{\left\|x^{(k)}\right\|_{12}} \begin{aligned} & \text { break }\end{aligned} \epsilon_{\text {rel }}+\frac{\epsilon_{\text {abs }}}{\left\|x^{(k)}\right\|_{I^{2}}} \quad\) (3.158)
        end \#if
    end \# for
```

has to be true, otherwise the algorithm terminates. Technically, the termination criterion in (3.158) consists of a combination of an absolute termination criterion and a relative termination criterion.

Optionally, as an additional safeguard, one can incorporate the norm of the evaluated actual reduction function in (3.150) regarding an absolute threshold and a relative threshold. If, in some way, the gradient information concerning ( $\hat{\hat{j}} \circ\left(\mathrm{id}_{Y_{0}} \circ{ }_{Y_{10}}\right.$ $\left.\tilde{\mathrm{K}} \circ_{X_{01}} \tilde{\mathrm{P}}\right)\left(x^{(k)}\right)$ or $\left(\hat{j} \circ\left(\operatorname{id}_{Y_{0}} \circ_{Y_{10}} \tilde{\mathrm{~K}}\right)\right)\left(\tilde{x}^{(k)}\right)$ is available, then the gradient information
can be utilized for a termination criterion as well.
Finally, let us choose an intermediate level between a pseudocode and a code from a programming language in industry (cf. the Listing 3.1) in order to represent the TRASM algorithm 3.1.

For the sake of completeness, let us discuss briefly the basic building blocks of the Manifold Mapping (MM) algorithm which utilizes the surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}\left(\tilde{\mathrm{R}}, \tilde{\mathrm{K}}, \imath_{\tilde{\mathrm{P}}}\right)$. The corresponding family of algorithms is based upon theoretical considerations that, in the multivariate vector-valued use case, focus particularly on the situation in which $\operatorname{dim}\left(X_{0}\right)<\operatorname{dim}\left(Y_{0}\right)$ with $\operatorname{dim}\left(X_{0}\right):=d$ and $\operatorname{dim}\left(Y_{0}\right):=m_{w}$ (cf. [56, p. 44]). The elaborations are built upon the discussion in [56, p. 43-48] and in [49, p. 67ff]).

Adapting the constructions in (3.137), in (3.138), and in (3.139) and providing a desired high-fidelity function value $\mathrm{K}_{d} \in \mathbb{R}^{m_{w}}$ (cf. $\mathbf{y}_{d}$ in (2.31) and $Q_{d}$ in (2.33)) as well as an initial iteration point $x^{(0)}$ such as in (3.143) and set the initial iteration matrix $S^{(0)}:=I$ and the initial iteration matrix $T^{(0)}:=S^{(0)}$ with $I \in \mathbb{R}^{m_{w} \times m_{w}}$ being the identity matrix, then one can concretely define an update scheme for $x^{(k+1)}$ by

$$
\begin{align*}
& y^{(k)}:=\left(\iota_{\tilde{\mathrm{R}}}{ }^{\circ} \mathrm{Y}_{10} \tilde{\mathrm{~K}} \circ_{X_{01}} \iota_{\mathrm{P}}\right)\left(x^{(k)}\right)-T^{(k)}\left(\mathrm{K}\left(x^{(k)}\right)-\mathrm{K}_{d}\right),  \tag{3.159a}\\
& x^{(k+1)}:=\underset{x \in X_{0}}{\arg \min }\left\|\left(\iota_{\tilde{\mathrm{R}}}{ }^{\circ} Y_{10} \tilde{\mathrm{~K}} \circ_{X_{01}} \iota_{\tilde{\mathrm{P}}}\right)(x)-y^{(k)}\right\|_{l_{2}}^{2},  \tag{3.159b}\\
& D_{\mathrm{K}}^{(k+1)}:=\left[\mathrm{K}\left(x^{(k+1)}\right)-\mathrm{K}\left(x^{(k)}\right), \ldots, \mathrm{K}\left(x^{(k+1)}\right)-\mathrm{K}\left(x^{(\max (k+1-d, 0))}\right)\right], \\
& D_{\mathrm{K}}^{(k+1)}:=\left[\left(\iota_{\tilde{\mathrm{R}}} \circ{ }_{Y_{10}} \tilde{\mathrm{~K}} \circ_{X_{01}} \iota_{\tilde{\mathrm{P}}}\right)\left(x^{(k+1)}\right)-\left(\iota_{\tilde{\mathrm{R}}} \circ \mathrm{Y}_{10} \tilde{\mathrm{~K}} \circ \mathrm{X}_{01} \iota_{\tilde{\mathrm{P}}}\right)\left(x^{(k)}\right), \ldots,\right.  \tag{3.159c}\\
& \left.\left(\iota_{\tilde{\mathrm{R}}}{ }^{\circ} \mathrm{Y}_{10} \tilde{\mathrm{~K}}{ }_{o_{01}} \iota_{\tilde{\mathrm{P}}}\right)\left(x^{(k+1)}\right)-\left(\iota_{\tilde{\mathrm{R}}}{ }^{\circ}{\Upsilon_{10}}^{\tilde{\mathrm{K}}}{ }_{\mathrm{O}_{X_{01}}} \iota_{\tilde{\mathrm{P}}}\right)\left(x^{(\max (k+1-d, 0))}\right)\right], \tag{3.159d}
\end{align*}
$$

$$
\begin{align*}
& {\left[U_{\mathrm{K}}^{(k+1)}, \Sigma_{\mathrm{K}}^{(k+1)}, V_{\mathrm{K}}^{(k+1)}\right]:=\operatorname{svd}\left(D_{\mathrm{K}}^{(k+1)}\right) \text {, }}  \tag{3.159e}\\
& {\left[U_{\grave{\mathrm{K}}}^{(k+1)}, \Sigma_{\grave{\mathrm{K}}}^{(k+1)}, V_{\stackrel{\mathrm{K}}{2}}^{(k+1)}\right]:=\operatorname{svd}\left(D_{\grave{\mathrm{K}}}^{(k+1)}\right) \text {, }}  \tag{3.159f}\\
& A^{(k+1)}:=\left(I-U_{\mathrm{K}}^{(k+1)}\left(U_{\mathrm{k}}^{(k+1)}\right)^{\mathrm{T}}\right) \text {, }  \tag{3.159~g}\\
& S^{(k+1)}:=D_{\kappa}^{(k+1)}\left(D_{\grave{\mathrm{K}}}^{(k+1)}\right)^{+}+A^{(k+1)}\left(I-U_{\tilde{\mathrm{K}}}^{(k+1)}\left(U_{\tilde{\mathrm{K}}}^{(k+1)}\right)^{\mathrm{T}}\right) \text {, }  \tag{3.159h}\\
& T^{(k+1)}:=\left(S^{(k+1)}\right)^{+} \text {, } \tag{3.159i}
\end{align*}
$$

where the matrix $A^{(k+1)} \in \mathbb{R}^{m_{w v} \times m_{w v}}$ serves as a potential stabilizer for the MM algorithm (cf. [56, p. 46]), and the operation svd refers to the singular value decomposition method, i.e., the operation svd performs the corresponding factorization for a given matrix such that

$$
\begin{align*}
& D_{\grave{k}}^{(k+1)}:=U_{\mathrm{k}}^{(k+1)} \Sigma_{\mathrm{k}}^{(k+1)}\left(V_{\mathrm{k}}^{(k+1)}\right)^{\mathrm{T}}  \tag{3.160}\\
& D_{\grave{k}}^{(k+1)}:=U_{\mathrm{k}}^{(k+1)} \Sigma_{\stackrel{\mathrm{K}}{(k+1)}}^{\left(V_{\mathrm{k}}^{(k+1)}\right)^{\mathrm{T}},} \tag{3.161}
\end{align*}
$$

with $U_{\mathrm{K}}^{(k+1)}, U_{\tilde{\mathrm{K}}}^{(k+1)} \in \mathbb{R}^{m_{w} \times m_{w}}, \Sigma_{\mathrm{K}}^{(k+1)}, \Sigma_{\tilde{\mathrm{K}}}^{(k+1)} \in \mathbb{R}^{m_{w} \times d}$, and $V_{\mathrm{K}}^{(k+1)}, V_{\tilde{\mathrm{K}}}^{(k+1)} \in \mathbb{R}^{d \times d}$. Note that the matrix $D_{\mathbb{K}}^{(k+1)} \in \mathbb{R}^{m_{w v} \times d}$ in (3.159c) and the matrix $D_{\tilde{\mathrm{K}}}^{(k+1)} \in \mathbb{R}^{m_{w} \times d}$ in (3.159d) can be equivalently determined by

$$
\begin{align*}
& \forall j \in\left\{1, \ldots, \min \left(\operatorname{dim}\left(X_{0}\right), k+1\right)\right\} . \forall i \in\left\{1, \ldots, \operatorname{dim}\left(Y_{0}\right)\right\} . \\
& {\left[d_{i, j}\right]_{\mathrm{K}}^{(k+1)}=\mathrm{K}\left(x^{(k+1)}\right)-\mathrm{K}\left(x^{(k+1-j)}\right),}  \tag{3.162}\\
& {\left[d_{i, j}\right]_{\tilde{\mathrm{K}}}^{(k+1)}=\left(\iota_{\tilde{\mathrm{R}}}{ }^{\circ} Y_{10} \tilde{\mathrm{~K}} \circ_{X_{01}} \iota_{\mathrm{P}}\right)\left(x^{(k+1)}\right)-\left(\iota_{\tilde{\mathrm{R}}}{ }^{\circ}{Y_{10}}^{\left.\tilde{\mathrm{K}} \circ_{X_{01}} \iota_{\mathrm{P}}\right)\left(x^{(k+1-j)}\right) .} .\right.} \tag{3.163}
\end{align*}
$$

Observe that the definition of the matrix $D_{\mathrm{K}}^{(k+1)}$ and the definition of the matrix $D_{\hat{K}}^{(k+1)}$ reveal that, at the program level (recall Figure 1.4), there is some kind of allocationsensitive bookkeeping necessary for constructing $D_{K}^{(k+1)}$ and $D_{\mathrm{K}}^{(k+1)}$.

Proper termination criteria for the MM algorithm can be defined analogously to the TRASM algorithm 3.1. Furthermore, in [124], the authors propose some heuristics in order to extend the basic building blocks of the MM algorithm in (3.159) by a trust-region framework - sort of like in the TRASM algorithm 3.1.

At the algorithm level (recall Figure 1.4), though, there are two notable differences between the MM algorithm and the TRASM algorithm 3.1. Firstly, the MM algorithm does not rely on solving an additional optimization problem such as for determining $\tilde{\mathrm{P}}_{v}(x)$ in (3.134), but it relies on performing computational efficiently two singular value decompositions. And, secondly, the MM algorithm exploits one high-fidelity model evaluation in the $k+1$-th iteration instead of two high-fidelity model evaluations in the TRASM algorithm 3.1.

Building upon investigations of algorithms of Space Mapping (SM) kind - such as the TRASM algorithm 3.1 - which utilize the surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}\left(\iota_{\tilde{R}}, \tilde{\mathrm{~K}}, \tilde{\mathrm{P}}\right)$ and investigations of algorithms of Manifold Mapping (MM) kind which utilize the surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}\left(\tilde{\mathrm{R}}, \tilde{\mathrm{K}}, \imath_{\tilde{\mathrm{P}}}\right)$, the author in [49, p. 112ff] proposes the Response and Parameter Mapping (RPM) algorithm which utilizes the surrogate model $\tilde{\mathrm{K}}_{\mathrm{s}}(\tilde{\mathrm{R}}, \tilde{\mathrm{K}}, \tilde{\mathrm{P}})$ and in which basic building blocks from SM algorithms and MM algorithms are combined. The RPM algorithm's additional computational overhead compared to a SM algorithm or to a MM algorithm is justified by a hopefully higher accuracy.

In the RPM algorithm, the update scheme of the matrix $B^{(k+1)}$ in (3.145) is performed in a manner similar to the update scheme of the matrix $S^{(k+1)}$ in (3.159h) which could lead to an interesting adaptation of the TRASM algorithm 3.1. However, the data situation regarding a widely ramified overall assessment of the RPM algorithm in terms of, e.g., accuracy, speed or convergence is very limited. Hence, it is presumed that the scope of the benefit of a corresponding adaptation of the TRASM algorithm 3.1 may be limited as well.

Observing the different kinds of algorithms, then one can generally discern that the Achilles' heel - or the spot that requires the most attention - of all the optimization algorithms within the space-mapping paradigm is the need for a benign resemblance of the high-fidelity model and the low-fidelity model (cf. § 3.1.3).

However, due to their dependence on properties of the low-fidelity model $\tilde{K}$, the surrogate model $K_{s}^{\tilde{R}, \tilde{K}, \tilde{F}}$, and the high-fidelity model $K$, the convergence analysis of the corresponding iteration procedures for (3.140) is fairly intricate (see, e.g., [49, p. 76-84]).

If we invoke the NREGE $\mathfrak{e}_{\mathrm{H}, \boldsymbol{s}_{\tilde{S}}}^{\mathrm{NR}}\left(\hat{Q}_{\tilde{\mathcal{S}}}\right)$ in (3.24) and the SSPCC $r_{\hat{y} \hat{y}}^{2}$ in (3.31) w.r.t. the low-fidelity model and if we adapt the NREGE and the SSPCC w.r.t. the surrogate
 fidelity model in (3.24) and in (3.31) with the surrogate model, then, driven by heuristics, it is presumably reasonable to assert formally that, with regard to some appropriate norms,

$$
\begin{align*}
& \forall \mathrm{K}_{\mathrm{s}}^{\tilde{R}, \tilde{K}, \tilde{\mathrm{P}}} \cdot \forall \tilde{\mathrm{~K}}\left(\mathfrak{e}_{\tilde{H}, s_{g}}^{\mathrm{NR}}\left(\hat{Q}_{\tilde{\xi}}\right) \rightarrow 0 \wedge r_{\hat{y} \tilde{y}}^{2} \rightarrow 1\right) \wedge\left(\mathfrak{e}_{\mathrm{H}, s_{8}, K \mathrm{~K}}^{\mathrm{NR}}, \tilde{\mathrm{R}, \tilde{P}},\right. \\
& \Longrightarrow x^{(k)} \rightarrow x^{*} \text { as } k \rightarrow \infty \text { w.r.t }(3.140 a), \tag{3.164}
\end{align*}
$$

where $m$ denotes the number of sampling plan points such as in (3.35).

Similarly to the comments on (3.35), note that, from an application-driven viewpoint, we are rather drawn towards the pre-asymptotic behavior than towards the asymptotic behavior. Furthermore, note that the need for convergence w.r.t. both the low-fidelity model and the surrogate model is comprehensible if we plug the high-fidelity model as a low-fidelity model, i.e., $\mathrm{K} \equiv \tilde{\mathrm{K}}$, in (3.164). In this case, one can observe by definition convergence w.r.t. the low-fidelity model, but one could observe non-convergence w.r.t. the surrogate model due to an improper choice of correction maps.

Chiefly, the statement in (3.164) is a novel attempt to express formally the abovementioned intricateness of a convergence analysis of iteration procedures associated with (3.140).

A practical value of the statement in (3.164) resides in the issue concerning the quantitative assessment of the quality of a low-fidelity model and a surrogate model. Hence, it contributes to the discussion regarding this issue (see, for instance, in [121] or in [49, p. 82ff]) where the NREGE and the SSPCC can serve as quality measures.

Another practical value of the statement in (3.164) resides in the treatment of the NREGE and the SSPCC as potential safeguards for any optimization algorithm within the space-mapping paradigm. Hence, keeping the concrete implementations separated from the abstract specification in (3.140), I deem it beneficial to incorporate the NREGE and the SSPCC at the zeroth step, at an intermediate step or at the last step of an iteration procedure concerning (3.140). At the last step of a corresponding iteration procedure, the NREGE and the SSPCC could serve as an ultimate termination criterion.

Similarly to the sequential kriging optimization in § 3.3.1, the experimental rate of convergence of iteration procedures concerning (3.140) might be very low or the corresponding iteration procedures might not be convergent at all (see, e.g., [49, p. 84] or [49, p. 101f]). Thus, the condition in (3.107) for a given problem is either not applicable since its assumption is not satisfied or not necessarily true at all.

Therefore, from an application-driven viewpoint embedded in the context of validation and verification, it might be judicious to adapt the procedure SBO-DFLF and the procedure SBO-SPLF in order to construct the schematic procedure that I refer to as SGO-SPLF:

1) create a sample w.r.t. the high-fidelity model (\$3.1.1);
2) provide a simplified-physics low-fidelity model (§3.1.3) and compute $\left.\bar{r}_{\tilde{y} \tilde{y}}^{2}\right|_{k}$ w.r.t. to the sample from step 1) and if $\left.\bar{r}_{y \tilde{y}}^{2}\right|_{k}<0.75$, break off the current procedure and invoke the procedure SBO-DFLF, otherwise continue the current procedure;
3) invoke a global optimization algorithm (\$2.3.3) w.r.t. the provided simplifiedphysics low-fidelity model from step 2);
4) use the minimizer from step 3) as a starting point in order to construct in the minimizer's vicinity a data-fit low-fidelity model (\$3.1.2) w.r.t. the simplifiedphysics low-fidelity model from step 2);
5) invoke an optimization algorithm within the space-mapping paradigm where the low-fidelity model is identical with the data-fit low-fidelity model from step 4);
6) use the minimizer from step 3) as a starting point for a local optimization algorithm ( $\$ 2.3 .3$ ) w.r.t. the high-fidelity model from step 1 ).
7) compare the minimizer from step 6) with the minimizer from step 5) and if the discrepancy is much larger than a user-assigned threshold, improve the simplified-physics low-fidelity model from step 2) or improve the data-fit lowfidelity model from step 4 ) or improve both low-fidelity models.

Similarly to the TRASM algorithm 3.1 and the proposed procedures SBO-SPLF and SBO-DFLF, the proposed schematic procedure SGO-SPLF builds upon the common canon of optimization algorithms (recall § 2.3.3).

Note that, in the procedure SGO-SPLF, I extend an approach proposed in [124] which is incorporated by step 4) and step 5). More precisely, it is assumed that the simplified-physics low-fidelity model - e.g., a low-fidelity model based on a coarse-grid discretization or a weakened termination criteria of an iterative solver or a combination of both (recall §3.1.3) - is computationally too expensive such that there is a need to construct a data-fit low-fidelity model w.r.t. the simplified-physics low-fidelity model as well.

In order to mitigate the curse of dimensionality (recall § 3.1.2), the data-fit lowfidelity model is constructed in the vicinity of the simplified-physics low-fidelity model's minimizer, that is, the domain under consideration (see, e.g., Figure 2.2 and Figure 2.3) for the sampling plan $X_{s}$ is smaller.

Finally, observe that the $k+1$-th iteration point in (3.140a) of an optimization algorithm within the space-mapping paradigm can be theoretically interpreted as a new $m+1$ sampling plan point in (3.115) of the sequential kriging optimization. This contemplation furnishes us with some kind of semantics in order to contrast loosely the corresponding optimization procedures that are all conceived as subtypes of the model management strategy adaptation.

Furthermore, if we select a kriging low-fidelity model within the procedure SGOSPLF, then, technically, the afore-mentioned contemplation gives us the opportunity to suggest an extension of the procedure SGO-SPLF by an inner sequential kriging optimization in order to provide a mechanism for an adaptive improvement of the data-fit low-fidelity model in step 5). The intricateness of such an interweavement is captured by the statement in (3.164) as well.

### 3.3.3 Co-kriging optimization

Recalling our explanations regarding the mathematical machinery behind the kriging low-fidelity model in § 3.1.2 and regarding the simplified-physics low-fidelity models in § 3.1.3, especially, the generic statement in (3.103), one can articulate the essential ideas behind the co-kriging optimization.

Conceptually, a co-kriging low-fidelity model is a kind of a kriging low-fidelity model in which the prediction at an arbitrary point $x$, i.e., $\hat{y}(x)$ in (3.98), incorporates the information of a sample $\mathbf{s}$ in (3.13) with respect to a high-fidelity model K and the information of a sample $\mathbf{s}_{\tilde{\mathrm{K}}}$ with respect to a low-fidelity model $\tilde{\mathrm{K}}$ (cf. [70, p. 168]). Due to this construction principle, the optimization based on a co-kriging low-fidelity model is conceived as a subtype of the model management strategy fusion.

A very rough intuition behind the prediction of the co-kriging low-fidelity model reveals that it behaves as an interpolation problem with regard to the sample $s$ and, as long as there is no coincidence between $\mathbf{s}_{\tilde{\mathrm{K}}}$ and $\mathbf{s}$, it behaves as some kind of regression problem with regard to the sample $\mathbf{s}_{\tilde{K}}$ (cf. [70, p. 172]).

Mind that the following technical explanations concerning a co-kriging low-fidelity model are condensed and built upon the discussion in [70, p. 167-177]. Hence, for
more details regarding some derivations, I refer to the corresponding reference and the references therein.

Adapting the notation in § 3.3.2 regarding a high-fidelity model K and a lowfidelity model $\tilde{\mathrm{K}}$, respectively, let us assume a sampling plan $X_{\mathbf{s}} \subseteq X_{0}^{m}$ in (3.14) with respect to a high-fidelity model K and a sampling plan $X_{\mathrm{s}_{\mathrm{K}}} \subseteq X_{1}^{m_{\mathrm{K}}}$ with respect to a low-fidelity model $\tilde{\mathrm{K}}$ with $m$ and $m_{\tilde{\mathrm{K}}}$ being the respective sample sizes such that $\forall m, m_{\tilde{\mathrm{K}}} . m<m_{\tilde{\mathrm{K}}}$. Furthermore, supposing some kind of sub-structure between $X_{0}$ and $X_{1}$ as well as between $Y_{1}$ and $Y_{0}$ such as in (3.122), it is demanded that $X_{\mathbf{s}} \subset X_{\mathbf{s}_{\bar{k}}}$. However, for the sake of brevity, let us omit the explicit mentioning of the inclusion maps.

If we identify $X_{\mathbf{s}}$ with an $m \times d$ matrix and the corresponding sampling plan points $x_{i}$ with $1 \times d$ matrices and if we identify $X_{s_{\tilde{K}}}$ with an $m_{\tilde{K}} \times d$ matrix the corresponding sampling plan points $\tilde{x}_{i}$ with $1 \times d$ matrices where $d$ denotes the number of parameters, then one can define a $\left(m_{\tilde{\mathrm{K}}}+m\right) \times d$ matrix $X_{\mathbf{s}_{\bar{k}}, \mathbf{s}}$ by

$$
\begin{align*}
X_{\mathrm{s}_{\kappa_{,}, s}} & :=\left[\begin{array}{ll}
X_{\mathrm{s}_{\bar{K}}}^{\mathrm{T}} & X_{\mathbf{s}}^{\mathrm{T}}
\end{array}\right]^{\mathrm{T}},  \tag{3.165a}\\
& \equiv\left[\begin{array}{lllll}
\tilde{x}_{1}^{\mathrm{T}} & \ldots & \tilde{x}_{m_{\nwarrow}}^{\mathrm{T}} & x_{1}^{\mathrm{T}} & \ldots
\end{array} x_{m}^{\mathrm{T}}\right]^{\mathrm{T}},  \tag{3.165b}\\
& \equiv\left[\begin{array}{c}
X_{\mathrm{s}_{\overleftarrow{K}}} \\
X_{\mathbf{s}}
\end{array}\right], \tag{3.165c}
\end{align*}
$$

where $X_{\mathrm{s}_{\mathrm{K}}, \mathrm{s}}$ denotes the joint sampling plan.
With regard to a given joint sampling plan $X_{s_{\bar{K}}, s,}$, let us encode in matrix representation the joint output points as a column vector $y_{\mathrm{s}_{\bar{k}}, s} \in \mathbb{R}^{\left(m_{\overline{\mathrm{K}}}+m\right) \times 1}$ with

$$
\begin{align*}
y_{\mathbf{s}_{K}, s} & :=\left[\begin{array}{llllll}
\tilde{y}_{1} & \ldots & \tilde{y}_{m_{\bar{k}}} & y_{1} & \ldots & y_{m}
\end{array}\right]^{\mathrm{T}},  \tag{3.166a}\\
& :=\left[\begin{array}{lll}
\tilde{y}^{\mathrm{T}} & y^{\mathrm{T}}
\end{array}\right]^{\mathrm{T}}, \tag{3.166b}
\end{align*}
$$

where $\tilde{y}_{i}$ refer to the output points regarding the sampling plan $X_{\mathbf{s}_{\bar{K}}}$ with respect to a low-fidelity model $\tilde{K}$ and $y_{i}$ refer to the output points regarding the sampling plan $X_{\mathbf{s}}$ with respect to a high-fidelity model K. Similarly to (3.83), the output points $\tilde{y}_{i}$ are represented by the column vector $\tilde{y} \in \mathbb{R}^{m_{\bar{\kappa}} \times 1}$ and the output points $y_{i}$ are represented by the column vector $y \in \mathbb{R}^{m \times 1}$.

Technically, $y_{s_{\bar{K}}, s}$ is associated with a corresponding random field $\underline{Y}_{s_{反}, s}$, i.e., a vector of random variables. In this context, a crucial ingredient is the so-called autoregressive model assumption, which is a kind of a memoryless property or Markov property; it states that

$$
\begin{equation*}
\forall x \neq x_{i} \cdot \operatorname{cov}\left(Y_{\mathbf{s}}\left(x_{i}\right), Y_{\mathbf{s}_{\tilde{K}}}(x) \mid Y_{\mathbf{s}_{\overleftarrow{K}}}\left(x_{i}\right)\right)=0 \tag{3.167}
\end{equation*}
$$

The statement in (3.167) reflects the idea that one can conceive the high-fidelity model as an exact representation. More precisely, if one possesses all the information about the high-fidelity model at $x_{i}$, then the low-fidelity model will not provide any new information about $Y_{s}\left(x_{i}\right)$, but any potential errors are only due to the lowfidelity model. For more details regarding the statement in (3.167), see, e.g., (cf. [70, p. 168]) and references therein.

Let us concretize the entities $\mathrm{K}(x), \tilde{\mathrm{K}}(x)$, and $\mathrm{Z}_{\Delta}(x)$ in the generic statement in (3.103) by Gaussian processes and the entity $Z_{\rho}(x)$ by a constant scaling factor $\rho \in \mathbb{R}$, then one receives the concrete statement

$$
\begin{equation*}
\forall x \in X_{\mathrm{s}} \cdot \mathrm{Z}(x)=\rho \tilde{\mathrm{Z}}(x)+\mathrm{Z}_{\Delta}(x), \tag{3.168}
\end{equation*}
$$

where the high-fidelity model evaluation and the low-fidelity model evaluation are redefined as Gaussian process $Z(x)$ and $\tilde{Z}(x)$, respectively.

Similarly to the covariance matrix in (3.88), one can construct the corresponding covariance matrix $C \in \mathbb{R}^{\left(m_{\bar{\kappa}}+m\right) \times\left(m_{\bar{\kappa}}+m\right)}$ by the five individual correlation matrices

$$
\begin{align*}
& \Psi_{\tilde{\mathrm{K}}}\left(X_{\mathrm{s}_{\tilde{K}^{\prime}}}, X_{\mathrm{s}_{\tilde{\mathrm{K}}}}\right) \in \mathbb{R}^{m_{\bar{\kappa}} \times m_{\tilde{\kappa}}},  \tag{3.169a}\\
& \Psi_{\tilde{K}}\left(X_{\mathbf{s}_{\kappa}}, X_{\mathbf{s}}\right) \in \mathbb{R}^{m_{\bar{\kappa} \times m}},  \tag{3.169b}\\
& \Psi_{\tilde{\kappa}}\left(X_{\mathbf{s}}, X_{\mathbf{s}_{\tilde{\kappa}}}\right) \in \mathbb{R}^{m \times m_{\tilde{\kappa}}},  \tag{3.169c}\\
& \Psi_{\tilde{\mathrm{K}}}\left(X_{\mathbf{s}}, X_{\mathbf{s}}\right) \in \mathbb{R}^{m \times m} \text {, }  \tag{3.169d}\\
& \Psi_{\Delta}\left(X_{\mathbf{s}}, X_{\mathbf{s}}\right) \in \mathbb{R}^{m \times m} . \tag{3.169e}
\end{align*}
$$

Let us interpret the matrices in (3.169) as the evaluations of maps $\Psi$. with the signatures $\mathbb{R}^{a \times d} \times \mathbb{R}^{b \times d} \rightarrow \mathbb{R}^{a \times b}$ with $a, b, d \in \mathbb{N}$ and let us invoke (3.89) in order to determine the entries of the correlation matrices, then one can define the family of maps $\Psi$. as

$$
\begin{equation*}
\Psi_{\bullet}=\left({ }^{1} X,{ }^{2} X\right) \mapsto \Psi_{\bullet}\left({ }^{1} X,{ }^{2} X\right):=\left[\psi_{i, l}\right] . \equiv\left[\exp \left(-\sum_{j=1}^{d} \cdot \theta_{j}| |^{1} x_{i}^{j}-\left.x^{2} x_{l}^{j}\right|^{\bullet} p_{j}\right)\right] \tag{3.170}
\end{equation*}
$$

where $i \in\{1, \ldots, a\}$ and $l \in\{1, \ldots, b\}$ and ${ }^{1} x_{i}$ refers to sampling plan points of the sampling plan ${ }^{1} X$ and ${ }^{2} x_{l}$ refers to sampling plan points of the sampling plan ${ }^{2} X$. Finally, one can construct the covariance matrix $C$ as
where, similarly to (3.87), $\sigma_{\hat{\mathbb{K}}}^{2}, \sigma_{\Delta}^{2} \in \mathbb{R}$.
Given the definition in (3.171) and the definition in (3.170), it is observable that, similarly to (3.97), one has to determine maximum likelihood estimates of the parameters $\left(\theta_{\tilde{K}}, p_{\tilde{K}}, \theta_{\Delta}, p_{\Delta}, \rho\right) \in\left[0,+\infty\left[{ }^{d} \times[0,2]^{d} \times\left[0,+\infty\left[{ }^{d} \times[0,2]^{d} \times \mathbb{R}\right.\right.\right.\right.$. For the sake of notational ease - especially in (3.170), let us treat the 5-tuple ( $\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{K}}}, \theta_{\Delta}, p_{\Delta}, \rho$ ) and the 5-tuple ( $\left.{ }^{\tilde{\mathrm{F}}_{\theta}}, \tilde{\mathrm{K}}_{p},{ }^{\Delta} \theta,{ }^{\Delta} p, \rho\right)$ as definitional equal, i.e., let us demand that the expression $\left(\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{K}}}, \theta_{\Delta}, p_{\Delta}, \rho\right) \equiv\left({ }^{\tilde{\mathrm{K}}_{\theta}}, \tilde{\mathrm{K}}_{p},{ }_{\theta}, \Delta_{p}, \rho\right)$ holds componentwise.

It is assumed that the information associated with the low-fidelity model is independent of the information associated with the high-fidelity model. Hence, one can determine the maximum likelihood estimates for $\mu_{\tilde{\mathrm{K}}} \in \mathbb{R}$ and $\sigma_{\tilde{\mathrm{K}}}^{2}$ by

$$
\begin{align*}
& \hat{\mu}_{\tilde{K}}:=\frac{\underline{1}_{\tilde{K}}^{T} \Psi_{\tilde{K}}\left(X_{\tilde{s}_{\tilde{K}}}, X_{\mathrm{s}_{\tilde{K}}}\right)^{-1}}{\underline{-}_{\tilde{K}}^{T} \Psi_{\tilde{K}}\left(X_{\mathrm{s}_{\tilde{K}}}, X_{\mathrm{s}_{\tilde{K}}}\right)^{-1} \underline{1_{\tilde{K}}}} \tilde{y},  \tag{3.172a}\\
& \hat{\sigma}_{\tilde{K}}^{2}:=\frac{1}{m_{\tilde{K}}}\left(\tilde{y}-\hat{\mu}_{\tilde{K}}\right)^{\mathrm{T}} \Psi_{\tilde{\mathrm{K}}}\left(X_{\mathrm{s}_{\tilde{K}}}, X_{\mathrm{s}_{\tilde{K}}}\right)^{-1}\left(\tilde{y}-\underline{\hat{\mu}}_{\tilde{\mathrm{K}}}\right), \tag{3.172b}
\end{align*}
$$

where $\hat{\underline{\mu}}_{\widehat{K}}$ is defined similarly to (3.95), that is,

$$
\begin{equation*}
\hat{\underline{\hat{H}}}_{\tilde{\mathrm{K}}}:=\hat{\mu}_{\tilde{\mathrm{K}}} \cdot \underline{1}_{\tilde{\mathrm{K}}} \tag{3.173}
\end{equation*}
$$

where $\underline{1}_{\tilde{\mathrm{K}}}:=\left[\begin{array}{lllll}1 & 1 & \ldots & 1 & 1\end{array}\right]^{\mathrm{T}}$ with $\underline{1}_{\tilde{\mathrm{K}}} \in \mathbb{R}^{m_{\bar{\kappa}} \times 1}$.
Let us define the concentrated $\ln$-likelihood function associated with the information regarding the low-fidelity model $L_{c \ln , \tilde{\mathrm{~K}}}=\left(\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{K}}}\right) \mapsto L_{c \ln , \tilde{\mathrm{~K}}}\left(\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{K}}}\right)$ with the
signature $\left[0,+\infty\left[{ }^{d} \times[0,2]^{d} \rightarrow\right]-\infty, 0\right]$ such that the assignment $L_{c \ln , \tilde{\mathrm{~K}}}\left(\theta_{\tilde{K}}, p_{\tilde{\mathrm{K}}}\right)$ reads as

$$
\begin{equation*}
L_{c \ln , \tilde{\mathrm{~K}}}\left(\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{K}}}\right):=-\frac{m_{\tilde{\mathrm{K}}}}{2} \ln \left(\hat{\sigma}_{\tilde{\mathrm{K}}}^{2}\right)-\frac{1}{2} \ln \left(\left|\Psi_{\tilde{\mathrm{K}}}\left(X_{\mathrm{s}_{\tilde{K}}}, X_{\mathrm{s}_{\tilde{\mathrm{K}}}}\right)\right|\right) . \tag{3.174}
\end{equation*}
$$

By invoking a suitable optimization algorithm (recall § 2.3.3), one can compute the maximum likelihood estimates of $\left(\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{K}}}\right)$ by considering the expression

$$
\begin{equation*}
\left(\hat{\theta}_{\tilde{K}}, \hat{p}_{\tilde{K}}\right):=\underset{\left(\theta_{\theta_{\tilde{K}},}, \bar{p}_{\tilde{K}}\right) \in\left[0,+\infty\left[{ }^{d} \times[0,2]^{d}\right.\right.}{\arg \min }-L_{c \ln , \tilde{\mathrm{~K}}}\left(\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{K}}}\right) . \tag{3.175}
\end{equation*}
$$

In order to determine the maximum likelihood estimates of $\left(\theta_{\Delta}, p_{\Delta}, \rho\right)$, one has to provide a column vector $y_{\Delta} \in \mathbb{R}^{m \times 1}$ which encodes the difference between the column vector $y$ and the column vector $\tilde{y}$. Due to the auto-regressive model assumption in (3.167), one has to consider only those output points of $\tilde{y}$ that are associated with the sampling plan $X_{s}$. Hence, one has to construct a column vector $\left.\tilde{y}\right|_{X_{s}} \in \mathbb{R}^{m \times 1}$ that can be interpreted as a restriction of $\tilde{y}$ to the sampling plan $X_{\mathrm{S}} .{ }^{29}$

Conceptually, it is favorable to construct initially $X_{s_{\tilde{k}}}$ and $\tilde{y}$, and, subsequently, to construct $X_{\mathbf{s}}$ - such that $X_{\mathbf{s}} \subset X_{s_{\bar{K}}}$ - and $y$. Mind that, though, a sampling plan should possess desirable properties: It should be space-filling and non-collapsing (recall §3.1.1).

Hence, in the case of an Audze-Eglais LHC or a Maximin LHC (see Figure 3.1), one has to adopt an exchange algorithm (see [70, p. 28f]) in order to construct $X_{\mathbf{s}}$ from $X_{s_{\bar{K}}}$ such that $X_{\mathbf{s}} \subset X_{s_{\bar{K}}}$ and $X_{\mathrm{s}}$ possessing the desirable properties for a sampling plan. More precisely: Let us randomly select an initial $X_{\mathbf{s}}$ such that $X_{\mathbf{s}} \subset X_{\mathbf{s}_{K}}$ and $X_{s_{\bar{K}}} \backslash X_{\mathbf{s}} \in \mathbb{R}^{\left(m_{\overline{\mathrm{K}}}-m\right) \times d} \cdot{ }^{30}$

Furthermore, it is set that $X_{\mathrm{s}}^{(1)}:=X_{\mathrm{s}}$. Given the running index $k \in\{1, \ldots, m\}$, let us compute the corresponding space-filling (and non-collapsing) criterion for $X_{s}^{(k)}$, i.e., the Audze-Eglais criterion for an Audze-Eglais LHC and the Morris-Mitchell criterion for a Maximin LHC. Given the running index $j \in\left\{1, \ldots, m_{\tilde{\mathrm{K}}}-m\right\}$, one can exchange the sampling plan point $x_{k}$ of $X_{\mathbf{s}}^{(k)}$ with each $j$ sampling plan point of $X_{\mathbf{s}_{\mathrm{K}}} \backslash X_{\mathrm{s}}$; thus, constructing theoretically $j$ sampling plans $X_{s}^{(k, j)}$. For each $j$, one can compute the corresponding criterion of $X_{\mathbf{s}}^{(k, j)}$. If there is a $j^{*} \in\left\{1, \ldots, m_{\tilde{\mathrm{K}}}-m\right\}$ such that, for all $j$, the corresponding criterion for $X_{s}^{\left(k, j^{*}\right)}$ is optimal compared with the criterion for $X_{\mathrm{s}}^{(k)}$, then one can set $X_{\mathrm{s}}^{(k+1)}:=X_{\mathrm{s}}^{\left(k, j^{*}\right)}$, otherwise one can set $X_{\mathrm{s}}^{(k+1)}:=X_{\mathrm{s}}^{(k)}$; thus, completing an iteration of the exchange procedure. The exchange procedure is continued for all sampling plan points $x_{k}$ of $X_{s}$ until it terminates for $k=m$ and it returns $X_{\mathrm{s}}$ with $X_{\mathrm{s}}:=X_{\mathrm{s}}^{(m)}$.

In the case of a Sobol quasi-random sequence (see Figure 3.2), though, one can construct initially $X_{s_{\bar{K}}}$ and, then, we either pick the sampling plan points $\tilde{x}_{k}$ of $X_{s_{\bar{K}}}$ with $k \in\{1, \ldots, m\}$ as the sampling plan points $x_{i}$ of $X_{\mathrm{s}}$; or, alternatively, we construct $X_{\mathbf{s}}$ as a Sobol quasi-random sequence with $m$ sampling plan points from

[^39]scratch. In both approaches, the resulting sampling plan $X_{\mathbf{s}}$ satisfies $X_{\mathbf{s}} \subset X_{\mathbf{s}_{反}}$ and it is a Sobol quasi-random sequence itself, thus, it is a space-filling and non-collapsing sampling plan itself. Therefore, utilizing Sobol quasi-random sequences to construct the sampling plans $X_{\mathrm{s}}$ and $X_{\mathrm{s}_{\mathrm{K}}}$ can be seen as a computationally time-saving and cost-reducing alternative to the usage of an Audze-Eglais LHC or a Maximin LHC and the necessity of an exchange algorithm.

Imagine a use case in which $X_{\mathrm{s}_{\bar{K}}}$ is constructed as an Audze-Eglais LHC or a Maximin LHC and $X_{\mathrm{s}}$ is constructed as a Sobol quasi-random sequence. In such a use case, it is very likely that there are no output points within $\tilde{y}$ which can be associated with the sampling plan points of $X_{s}$. Hence, one has to invoke a fallback plan (cf. [70, p. 169]), i.e., given the maximum likelihood estimates ( $\hat{\theta}_{\tilde{K}}, \hat{\rho}_{\tilde{\mathrm{K}}}$ ), let us construct a kriging low-fidelity model (see (3.98)) of the low-fidelity model $\tilde{\mathrm{K}}$ as

$$
\begin{equation*}
\hat{y}(\tilde{x}):=\hat{\mu}_{\tilde{\mathrm{K}}}+\tilde{r}^{\mathrm{T}} \Psi_{\tilde{\mathrm{K}}}\left(X_{\mathrm{s}_{\tilde{\mathrm{K}}}}, X_{\mathrm{s}_{\tilde{\mathrm{K}}}}\right)^{-1}\left(\tilde{y}-\underline{\hat{\mu}}_{\tilde{\mathrm{K}}}\right), \tag{3.176}
\end{equation*}
$$

where $\hat{y}: X_{1} \rightarrow \mathbb{R}$ such that $\hat{y}(\tilde{x})$ indicates the prediction at an arbitrary point $\tilde{x}$ and $\tilde{r}:=\left[\tilde{r}_{i}\right] \in \mathbb{R}^{m_{\bar{\kappa}} \times 1}$ denotes the correlation column vector that reads as

$$
\tilde{r}:=\left[\begin{array}{lllll}
\tilde{r}_{1} & \tilde{r}_{2} & \ldots & \tilde{r}_{m_{\mathfrak{k}}-1} & \tilde{r}_{m_{\tilde{k}}} \tag{3.177}
\end{array}\right]^{\mathrm{T}},
$$

where, similarly to (3.81), the components $\tilde{r}_{i}$ are defined as

$$
\begin{equation*}
\tilde{r}_{i}:=\sum_{j=1}^{d} \tilde{\mathrm{~K}}_{\theta_{j}}\left|\tilde{x}^{j}-\tilde{x}_{i}^{j}\right|^{{ }^{\mathrm{K}}} p_{j} . \tag{3.178}
\end{equation*}
$$

By means of (3.176), let us forge a column vector $\left.\hat{y}\right|_{X_{s}} \in \mathbb{R}^{m \times 1}$.
Supposing $\left.\tilde{y}\right|_{X_{s}}$, one can ultimately create the column vector $y_{\Delta}$ that reads as

$$
\begin{equation*}
y_{\Delta}:=y-\left.\rho \tilde{y}\right|_{X_{\mathbf{s}}}, \tag{3.179}
\end{equation*}
$$

where the term $\left.\rho \tilde{y}\right|_{X_{s}}$ encodes a multiplication of the column vector $\left.\tilde{y}\right|_{X_{\mathrm{s}}}$ with the scalar $\rho$. At the programs level (recall Figure 1.4), the term $\left.\tilde{y}\right|_{X_{s}}$ forces us to ensure that we filter those components of the vector $\tilde{y}$ such that we establish a map $\left.\tilde{y} \mapsto \tilde{y}\right|_{X_{s}}$ in order to adequately compute $y_{\Delta}$ in (3.179).

Let us compute the maximum likelihood estimates for $\mu_{\Delta} \in \mathbb{R}$ and $\sigma_{\Delta}^{2}$ by

$$
\begin{align*}
& \hat{\mu}_{\Delta}:=\frac{\underline{1}_{\Delta}^{\mathrm{T}} \Psi_{\Delta}\left(X_{\mathbf{s}}, X_{\mathbf{s}}\right)^{-1}}{\underline{1}_{\Delta}^{\mathrm{T}} \Psi_{\Delta}\left(X_{\mathbf{s}}, X_{\mathbf{s}}\right)^{-1} \underline{1}_{\Delta}} y_{\Delta},  \tag{3.180a}\\
& \hat{\sigma}_{\Delta}^{2}:=\frac{1}{m}\left(y_{\Delta}-\hat{\hat{\mu}}_{\Delta}\right)^{\mathrm{T}} \Psi_{\Delta}\left(X_{\mathbf{s}}, X_{\mathbf{s}}\right)^{-1}\left(y_{\Delta}-\hat{\underline{\mu}}_{\Delta}\right), \tag{3.180b}
\end{align*}
$$

where $\underline{\hat{\mu}}_{\Delta}$ is defined similarly to (3.95), that is,

$$
\begin{equation*}
\hat{\underline{\mu}}_{\Delta}:=\hat{\mu}_{\Delta} \cdot \underline{1}_{\Delta}, \tag{3.181}
\end{equation*}
$$

where $\underline{1}_{\Delta}:=\left[\begin{array}{lllll}1 & 1 & \ldots & 1 & 1\end{array}\right]^{\mathrm{T}}$ with $\underline{1}_{\Delta} \in \mathbb{R}^{m \times 1}$.
Let us define the concentrated $\ln$-likelihood function associated with the information regarding the low-fidelity model and the high-fidelity model, i.e., let us define the map $L_{c \ln , \Delta}=\left(\theta_{\Delta}, p_{\Delta}, \rho\right) \mapsto L_{c \ln , \Delta}\left(\theta_{\Delta}, p_{\Delta}, \rho\right)$ such that the map's signature is
$\left[0,+\infty\left[{ }^{d} \times[0,2]^{d} \times \mathbb{R} \rightarrow\right]-\infty, 0\right]$ and the assignment $L_{c \ln , \Delta}\left(\theta_{\Delta}, p_{\Delta}, \rho\right)$ reads as

$$
\begin{equation*}
L_{c \ln , \Delta}\left(\theta_{\Delta}, p_{\Delta}, \rho\right):=-\frac{m}{2} \ln \left(\hat{\sigma}_{\Delta}^{2}\right)-\frac{1}{2} \ln \left(\left|\Psi_{\Delta}\left(X_{\mathbf{s}}, X_{\mathbf{s}}\right)\right|\right) . \tag{3.182}
\end{equation*}
$$

By invoking a suitable optimization algorithm (recall § 2.3.3), one can compute the maximum likelihood estimates of $\left(\theta_{\Delta}, p_{\Delta}, \rho\right)$ by considering the expression

$$
\begin{equation*}
\left(\hat{\theta}_{\Delta}, \hat{p}_{\Delta}, \hat{\rho}\right):=\underset{\left(\theta_{\Delta}, p_{\Delta}, p\right) \in\left[0,+\infty\left[{ }^{d} \times[0,2]^{d} \times \mathbb{R}\right.\right.}{\arg \min }-L_{c \ln , \Delta}\left(\theta_{\Delta}, p_{\Delta}, \rho\right) \cdot{ }^{31} \tag{3.183}
\end{equation*}
$$

In practical applications, it is advisable to associate $\rho$ in (3.183) with a bounded interval, e.g., with the closed interval $[-a, a]$ such that $\rho \in[-a, a]$ where $a \in \mathbb{R}^{+}$is a user-assigned and problem-dependent entity.

Additionally, mind that, in order for $\hat{\rho}$ to be a reliable estimate of the scaling in (3.168) and in (3.179), respectively, the sample size $m$ regarding $X_{s}$ has to be greater than or equal to a problem-dependent lower bound (cf. [70, p. 176]).

After computing the maximum likelihood estimates ( $\left.\hat{\theta}_{\check{\mathrm{K}}}, \hat{p}_{\widetilde{\mathrm{K}}}, \hat{\theta}_{\Delta}, \hat{p}_{\Delta}, \hat{\rho}\right)$, one can specify the co-kriging low-fidelity model as

$$
\begin{equation*}
\hat{y}(x):=\hat{\mu}_{\tilde{K}, K}+c^{\mathrm{T}} C^{-1}\left(y_{\mathrm{S}_{\tilde{K}}, s}-\underline{\hat{\mu}}_{\tilde{\mathrm{K}}, \mathrm{~K}}\right), \tag{3.184}
\end{equation*}
$$

where $\hat{y}: X_{1} \rightarrow \mathbb{R}$ such that $\hat{y}(x)$ indicates the prediction at an arbitrary point $x$. The maximum likelihood estimate for $\mu_{\tilde{\mathrm{K}}, \mathrm{K}} \in \mathbb{R}$ in (3.184) is given by

$$
\begin{equation*}
\hat{\mu}_{\tilde{K}, K}:=\frac{1_{\tilde{K}, K}^{\mathrm{T}} C^{-1}}{\underline{1}_{\tilde{K}, K}^{\mathrm{T}} \mathrm{C}^{-1} \underline{\underline{K}}_{\tilde{K}, K}} y_{\mathrm{s}_{\tilde{K}, \mathrm{~s}}}, \tag{3.185a}
\end{equation*}
$$

whereas $\underline{\hat{\mu}}_{\tilde{\mathrm{K}}, K}$ is defined similarly to (3.95), that is,

$$
\begin{equation*}
\hat{\underline{\mu}}_{\tilde{\mathrm{K}}, \mathrm{~K}}:=\hat{\mu}_{\widetilde{\mathrm{K}}, \mathrm{~K}} \cdot \underline{1}_{\tilde{\mathrm{K}}, \mathrm{~K}}, \tag{3.186}
\end{equation*}
$$

where $\underline{1}_{\tilde{K}, K}:=\left[\begin{array}{lllll}1 & 1 & \ldots & 1 & 1\end{array}\right]^{\mathrm{T}}$ with $\underline{1}_{\tilde{K}, K} \in \mathbb{R}^{\left(m_{\tilde{K}}+m\right) \times 1}$.
Recalling (3.88), the column vector $c \in \mathbb{R}^{\left(m_{\bar{K}}+m\right) \times 1}$ in (3.184) encodes the covariance between the sampling plan $X_{\mathrm{S}_{\bar{K}}}$ and an arbitrary point $x$ as well as the covariance between the sampling plan $X_{\mathbf{s}}$ and an arbitrary point $x$. The column vector $c$ can be written as

$$
c:=\left[\begin{array}{ll}
c_{1}^{\mathrm{T}} & c_{2}^{\mathrm{T}} \tag{3.187}
\end{array}\right]^{\mathrm{T}},
$$

where the column vector $c_{1} \in \mathbb{R}^{m_{\kappa} \times 1}$ and the column vector $c_{2} \in \mathbb{R}^{m \times 1}$ are defined as

$$
\begin{align*}
& c_{1}:=\hat{\rho} \hat{\sigma}_{\tilde{\mathrm{K}}}^{2} \Psi_{\tilde{\mathrm{K}}}\left(X_{\mathrm{s}_{\tilde{K}}}, x\right),  \tag{3.188a}\\
& c_{2}:=\hat{\rho}^{2} \hat{\sigma}_{\hat{\mathrm{K}}}^{2} \Psi_{\Delta}\left(X_{\mathrm{s}}, x\right)+\hat{\sigma}_{\Delta}^{2} \Psi_{\Delta}\left(X_{\mathrm{s}}, x\right), \tag{3.188b}
\end{align*}
$$

where $\Psi_{\tilde{\mathrm{K}}}\left(X_{\mathrm{S}_{\tilde{K}^{\prime}}} x\right) \in \mathbb{R}^{m_{\tilde{\kappa}} \times 1}$ and $\Psi_{\Delta}\left(X_{\mathrm{s}}, x\right) \in \mathbb{R}^{m \times 1}$. Since the column vector $c$ plays a similar role such as the column vector $r$ in (3.98), one can invoke the definition of the

[^40]family of maps $\Psi_{0}$ in (3.170) where, by setting $b:=1$, one can specify the signature by $\mathbb{R}^{a \times d} \times \mathbb{R}^{1 \times d} \rightarrow \mathbb{R}^{a \times 1}$ with $a, d \in \mathbb{N}$.

Technically, one can apply the interpretation that ${ }^{2} \mathrm{X}$ is a sampling plan with a single sampling plan point ${ }^{2} x_{1}$ which is defined as ${ }^{2} x_{1}:=x$. Hence, the expression $\Psi_{\tilde{\mathrm{K}}}\left(X_{s_{\bar{K}}}, x\right)$ in (3.188a) and the expression $\Psi_{\Delta}\left(X_{\mathrm{s}}, x\right)$ in (3.188b) can be understood by adapting the assignment in (3.170) such that

$$
\begin{equation*}
\Psi_{\bullet}=\left({ }^{1} X, x\right) \mapsto \Psi .\left({ }^{1} X, x\right):=\left[\psi_{i, 1}\right]_{\bullet} \equiv\left[\exp \left(-\left.\sum_{j=1}^{d} \bullet \theta_{j}\right|^{1} x_{i}^{j}-\left.x^{j}\right|^{p_{j}}\right)\right] \text {. } \tag{3.189}
\end{equation*}
$$

where $i \in\{1, \ldots, a\}$, and ${ }^{1} x_{i}$ refers to sampling plan points of the sampling plan ${ }^{1} X$ and $x$ refers to an arbitrary point in (3.184).

An important observation regarding the co-kriging low-fidelity model in (3.184) is that if we choose a point $x$ such that $x$ is a sampling plan point of $X_{\mathrm{s}}$ in (3.165), then $\hat{y}(x)$ is an output point within $y$ in (3.166).

Thus, regarding the information associated with the high-fidelity model K, the co-kriging low-fidelity model behaves the same way as the kriging low-fidelity model in (3.98). Recalling the very rough intuition at the beginning of the section § 3.3.3, the co-kriging low-fidelity model does not show such a behavior with regard to the information associated with the low-fidelity model $\tilde{\mathrm{K}}$.

Similarly to (3.111), one can provide a mean squared prediction error $\left(\hat{s}_{y}(x)\right)^{2}$ at an arbitrary point $x$ for the co-kriging low-fidelity model $\hat{y}(x)$ in (3.184). Hence, the error $\left(\hat{s}_{y}(x)\right)^{2}$ (cf. [70, p. 172]) can be defined as

$$
\begin{equation*}
\left(\hat{s}_{y}(x)\right)^{2}:=\hat{\rho} \hat{\sigma}_{\tilde{K}}^{2}+\hat{\sigma}_{\Delta}^{2}-c^{T} C^{-1} c+\frac{\left(1-\underline{1}_{\tilde{K}, \mathrm{~K}}^{T} C^{-1} c\right)^{2}}{\underline{1}_{\tilde{K}, \mathrm{~K}} C^{-1} \underline{1}_{\tilde{K}, K}}, \tag{3.190}
\end{equation*}
$$

where $\hat{s}_{y}: X \rightarrow \mathbb{R}$. Analogously to (3.111), the term in (3.190) involving the fraction is negligibly small, thus, let us reformulate the mean squared prediction error $\left(\hat{s}_{y}(x)\right)^{2}$ as

$$
\begin{equation*}
\left(\hat{s}_{y}(x)\right)^{2}:=\hat{\rho} \hat{\sigma}_{\tilde{\mathrm{K}}}^{2}+\hat{\sigma}_{\Delta}^{2}-c^{\mathrm{T}} C^{-1} c . \tag{3.191}
\end{equation*}
$$

Utilizing the mean squared prediction error $\left(\hat{s}_{y}(x)\right)^{2}$ in (3.191), one could formulate a sequential co-kriging optimization in the same fashion as the sequential kriging optimization in §3.3.1.

However, in the present work, we do not dwell on the sequential co-kriging optimization, but we rather dwell on the co-kriging optimization. More precisely: Given a high-fidelity model K , and a low-fidelity model $\tilde{\mathrm{K}}$ which, in the context at hand, is considered conceptually indistinguishable from a surrogate model, the high-fidelity optimization problem, for instance, in the formulation in (3.116), is replaced by a co-kriging low-fidelity optimization problem which one can state as, e.g.,

$$
\begin{equation*}
\min _{x \in X_{0}}(\hat{j} \circ \hat{y})(x), \tag{3.192}
\end{equation*}
$$

where $\hat{y}(x)$ refers to the co-kriging low-fidelity model in (3.184).
Notice well that, in a kriging low-fidelity optimization problem corresponding to (3.116), one is solely capable of obtaining information from the high-fidelity model in order to forge the kriging low-fidelity model.

Thus, let us conceive the corresponding optimization kind, i.e., kriging optimization, as a subkind of surrogate-based optimization.

In a co-kriging low-fidelity optimization problem such as in (3.192), though, a necessary minimum of information regarding the high-fidelity model is established, and, then, one can steer the amount of information regarding the low-fidelity model to, hopefully, improve the co-kriging low-fidelity model.

Thus, let us conceive the corresponding optimization kind, i.e., co-kriging optimization, as a subkind of surrogate-guided optimization.

Depending on the computational costs of the low-fidelity model and the surrogate model, respectively, in the co-kriging optimization, one can build a data-fit low-fidelity model of the surrogate model - analogously to step 3) and step 4) in the procedure SGO-SPLF - in order to utilize the data-fit low-fidelity model of the surrogate model as a proxy for the original surrogate model in the construction of the co-kriging low-fidelity model.

Finally, comparing the co-kriging optimization and the corresponding kriging optimization, one can check whether the relation in (3.107) holds to be true. Interestingly, it is conceivable that the relation in (3.107) holds to be true and that a measurable computation time regarding the co-kriging optimization is still larger than a measurable computation time regarding the kriging optimization. It appears like a seemingly paradoxical behavior.

The rationale for this seemingly paradoxical behavior is that the determination of the maximum likelihood estimates of the parameters ( $\left.\theta_{\tilde{\mathrm{K}}}, p_{\tilde{\mathrm{k}}}, \theta_{\Delta}, p_{\Delta}, \rho\right)$ in (3.175) and in (3.183) is much more involved than the determination of the maximum likelihood estimates of the parameters $(\theta, p)$ in (3.97).

Therefore, if the number of parameters and the dimensions of the matrices in the maximum likelihood estimations are too large (cf. [70, p. 171]), then a measurable computation time regarding the co-kriging optimization can be larger than a measurable computation time regarding the kriging optimization - albeit the relation in (3.107) holds.

Observe that the sequential co-kriging optimization appears as a hybrid of the model management strategies fusion and adaptation. Mind that, to my best knowledge, such hybrids are not extensively discussed within the classification scheme of model management strategies proposed in [166].

For instance, recalling the end of $\S 3.3 .2$, the suggested extension of the procedure SGO-SPLF by an inner sequential kriging optimization can be interpreted as a hybrid of one kind of adaptation model management strategy and another kind of adaptation model management strategy.

Furthermore, if we apply a formalization-oriented viewpoint in the sense that we discuss the co-kriging low-fidelity model in (3.184) embedded in the context of the optimization within the space-mapping paradigm (see § 3.3.2), then one can replace the low-fidelity optimization problem in (3.117) by the co-kriging low-fidelity optimization problem in (3.192). In the context of § 3.3.2, the co-kriging low-fidelity model can be understood as a second-level low-fidelity model w.r.t. the first-level low-fidelity model K. .

A first benefit of such a formalization-oriented viewpoint is that it grants us with a novel procedure that can be interpreted as a hybrid of the model management strategies fusion and adaptation. Such a viewpoint suggests, therefore, to conceive this hybrid (a co-kriging low-fidelity model within the space-mapping paradigm) as somehow comparable with the hybrid constituted by the sequential co-kriging optimization - at least at the function level (recall Figure 1.4).

A second benefit of such a formalization-oriented viewpoint is that it nurtures a modular construction principle such that, e.g., the extension of the procedure SGOSPLF could be executed by an inner sequential co-kriging optimization. Hence, this
novel suggested extension can be interpreted as a hybrid of one kind of adaptation model management strategy, and a fusion model management strategy, and another kind of adaptation model management strategy.

A third benefit of such a formalization-oriented viewpoint is that it furnishes us with a formal suspicion about the important role of the low-fidelity model within the co-kriging low-fidelity model. Note that the numerical experiments in $\$ 3.2$ within the context of surrogate-based optimization - for instance, in terms of $\overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{N}}, \overline{\mathrm{e}}_{\mathrm{cv}}^{\mathrm{NR}}, \bar{r}_{\hat{\mathrm{y}}, \mathrm{cv}}^{2}$ or $S_{\tilde{y}, i}^{\mathrm{N}}$ as shown in the Figures 3.12-3.15-furnish us with a numerical initial suspicion. To the best of my belief, the role of the low-fidelity model within the co-kriging low-fidelity model is not exhaustively formally elaborated in the literature (see, e.g., [70, p. 167 - 177] and references therein), though. Mind that, however, if we consider the hybrid constituted by a co-kriging low-fidelity model within the space-mapping paradigm, then one can invoke the statement in (3.164) which has to be adjusted regarding the sample size $m$ and the sample size $m_{\tilde{\mathrm{K}}}$ in (3.165). Hence, let us write the adjusted statement as, with regard to some appropriate norms,

$$
\begin{aligned}
& \forall K_{s}^{\tilde{R}, \tilde{K}_{2}, \tilde{\mathrm{P}}} \cdot \forall \tilde{\mathrm{~K}}_{2} \cdot \forall \tilde{\mathrm{~K}}_{1} \cdot\left(\mathfrak{e}_{H, s_{s}, \tilde{K}_{1}}^{\mathrm{NR}}\left(\hat{Q}_{\tilde{\xi}}\right) \rightarrow 0 \wedge r_{\hat{y} \tilde{\hat{y}}, \tilde{\mathrm{~K}}_{1}}^{2} \rightarrow 1\right) \wedge\left(\hat{e}_{\mathrm{H}, s_{8}, \tilde{K}_{2}}^{\mathrm{NR}}\left(\hat{Q}_{\tilde{\xi}}\right) \rightarrow 0 \wedge r_{\hat{y} \tilde{\tilde{y}}, \tilde{K}_{2}}^{2} \rightarrow 1\right) \wedge
\end{aligned}
$$

$$
\begin{align*}
& \Longrightarrow x^{(k)} \rightarrow x^{*} \text { as } k \rightarrow \infty \text { w.r.t (3.140a), } \tag{3.193}
\end{align*}
$$

where $\tilde{\mathrm{K}}_{1}$ refers to the first-level low-fidelity model $\tilde{\mathrm{K}}$ with $m_{\tilde{\mathrm{K}}_{1}} \equiv m_{\tilde{\mathrm{K}}}$, and $\tilde{\mathrm{K}}_{2}$ refers to the co-kriging low-fidelity model in (3.176). The remaining entities in (3.193) are defined according to our comments on the statement in (3.164).

Hence, by observing the statement in (3.193), one can utter the formal reasonable suspicion that the choice of the low-fidelity model within the co-kriging low-fidelity model obeys similar restrictions as the low-fidelity model within the space-mapping paradigm.

Or to put it differently: By embedding the convergence issues related to the cokriging low-fidelity model into the convergence issues related to the space-mapping paradigm, one can formally argue that the quality of the low-fidelity model within the co-kriging low-fidelity model has to satisfy some problem-dependent lower bounds.

Mind that the present work has primarily focused on working out the benefits of a purely formalization-oriented viewpoint which has led to fertile novel insights of theoretical value (such as hybrid model management strategies) and practical value (such as the quality of the low-fidelity model within the co-kriging low-fidelity model). These novel insights reveal novel research directions at the algorithm level, at the program level, and at the application level as well (recall Figure 1.4).

However, the scrutineering of these new research directions, i.e., their thorough and extensive examination, for instance, by comparing with the results in (3.2.1) and in (3.2.2) and by extending the corresponding database, has to be left for future work.

### 3.4 In closing

The chapter's primary purpose has been to provide us with an in-depth elaboration of this thesis's key notion surrogate optimization and an in-depth elaboration of the proposed partitioning of this notion in ch. 1.2 into the three sub-notions: (1) surrogate modeling \& simulation, (2) surrogate-based optimization, and (3) surrogateguided optimization.

Throughout the elaborations, we have anticipated algebraic tools from the category theoretical language in ch. 4 such that we have beneficially tagged the various notions of surrogate optimization with algebraic notes. Similarly to a lubricant, the algebraic tools enabled us to smoothly operate between the various layers in Figure 1.4.

Regarding the sub-notion (1) surrogate modeling \& simulation, we have initially discussed an abstract setting in order to introduce relevant notions from the common methodological and terminological toolbox. We have looked at different classes of mathematical problems and we have encountered various important terms such as high-fidelity model and low-fidelity model.

Next, we have defined the high-fidelity function approximation error. In the discussion about sampling plans, we have encountered three different kinds of sampling plans and their peculiarities. In the literature regarding surrogate optimization, to my best knowledge, some sampling plans, such as, e.g., those constructed by a Sobol quasi-random sequence, are not widely represented, yet. Afterwards, we have introduced the empirical surrogate modeling error where we have defined the empirical training error and the empirical generalization error as well.

The empirical generalization error, in particular, is an important indicator within surrogate optimization and we have presented this error in various guises, for instance, within the $k$-fold cross-validation method. Another important indicator is the squared sample Pearson correlation coefficient. We have carved out some not exhaustively discussed nuances in the case that this coefficient is being used together with the empirical generalization error within the $k$-fold cross validation method. These nuances gain in significance through the fact that, in the present work, the number of sampling plan points concerning the high-fidelity model is assumed to be sparse.

Furthermore, we have developed a potential link between the sample Pearson correlation coefficient and a low-fidelity models' normalized global first-order sensitivity measures which has culminated in a cautiously formulated conjecture about the trustworthiness of low-fidelity models' normalized global first-order sensitivity measures.

Subsequently, we have examined deterministic data-fit low-fidelity models, i.e., multivariate polynomials and radial basis functions, and probabilistic data-fit lowfidelity models, that is, kriging low-fidelity models. We have investigated diverse aspects of these models in order to gain an holistic understanding of these models and to spot possible pitfalls and a potential room for improvement. Some of the investigated aspects are: the underlying construction principles of the models, the computational significance of the dimension of the domain, the numerical relevance of the evaluation scheme, and a sampling plan's influence on the condition number of a problem-representing matrix.

We have closed the subpart (1) surrogate modeling \& simulation by an elaboration of simplified-physics low-fidelity models. We have examined the general concept of a user-prescribed hierarchy of problems with regard to the degree of fidelity
and the computational costs. We have concretized this general concept by presenting a user-prescribed hierarchy of magnetoquasistatic and magnetostatic problems which are associated with simplified-physics low-fidelity models. From this hierarchy, we have abstracted some diagrams in a loose category-theoretical style. At the end, we have paved the way for a purely formalization-oriented viewpoint on some surrogate-guided optimization approaches.

Regarding the sub-notion (2) surrogate-based optimization, we have examined the optimization with the test functions in $\S 2.3 .3$ by data-fit low-fidelity models and by emulated simplified-physics low-fidelity models. The essential idea is to solve an optimization problem associated with the low-fidelity model whose optimal solution is utilized as a starting point for the optimization problem associated with the high-fidelity model.

The proof of work in form of, e.g., visualizations of the above-mentioned indicators within surrogate optimization, appear valuable since, to my best knowledge, there is a lack of a comprehensive database of corresponding benchmarks. Hence, the proof of work equips us with a benchmark-focused classification of test functions (more generally, high-fidelity models) whose advantages and disadvantages we have briefly discussed.

An advantage is the opportunity to classify very roughly the behavior of a corresponding optimization problem within the magnetoquasistatic and magnetostatic context. A disadvantage is that it is not clear whether there exists a reliable complete list of indicators.

We have closed the subpart (2) surrogate-based optimization by an elaboration of the proposed procedures SBO-DFLF and SBO-SPLF and their potential combinations.

Regarding the sub-notion (3) surrogate-guided optimization, I have argued from an application-driven viewpoint that it is worthwhile to check whether the number of high-fidelity model evaluations regarding a surrogate-based optimization approach is higher than the number of high-fidelity model evaluations regarding a surrogate-guided optimization approach. The additional value of such a check is comprehensible in the context of validation and verification.

Subsequently, we have dwelled on the sequential kriging optimization as a subkind of the model management strategy adaptation, on optimization procedures within the space-mapping paradigm which are a subkind of the model management strategy adaptation, and on the co-kriging optimization which can be seen as a subkind of the model management strategy fusion.

Concerning the optimization within the space-mapping paradigm, we have utilized a formalization-oriented viewpoint to pin down properly, e.g., the conceptional distinction between a low-fidelity model and a surrogate model. Mind that there is a loss of conceptional information if we solely operate with various representations of the real numbers such as, e.g., $\mathbb{R}, \mathbb{R}^{n}$, and $\mathbb{R}^{n \times m}$ with $n, m \in \mathbb{N}$.

Nevertheless, we have concretized the formal concepts to investigate the syntax and the semantics of the multivariate scalar-valued use case and the multivariate vector-valued use case.

Subsequently, we have examined the basic building blocks of a representation of the Trust Region Aggressive Space Mapping (TRASM) algorithm, that is, the algorithm 3.1. Additionally, we have discussed the basic building blocks of some other proposed algorithms within the literature about the space-mapping paradigm.

Driven by heuristics, we have formulated a convergence statement that incorporates some of the above-mentioned indicators that, to the best of my belief, furnishes
us with a novel access to the delicate aspect of convergence-related issues within the space-mapping paradigm.

At the end, we have elaborated on the proposed procedure SGO-SPLF.
Concerning the co-kriging optimization, we have examined the basic building blocks for constructing the co-kriging low-fidelity model. A special challenge is the handling of a sampling plan associated with the high-fidelity model and a sampling plan associated with a low-fidelity model where the usage of algebraic notions facilitates the consideration.

An intriguing novel observation is that sampling plans constructed by a Sobol quasi-random sequence may help to reduce the overall computational costs of constructing a co-kriging low-fidelity model.

We have closed the subpart (3) surrogate-guided optimization by elaborating on the benefits of a purely formalization-oriented viewpoint that provides us with novel insights of theoretical value (such as potential hybrid model management strategies) and of practical value (such as convergence-related issues regarding the quality of the low-fidelity model within the co-kriging low-fidelity model).

## Chapter 4

## An algebraic modeling framework using the category theoretical language for applications in surrogate optimization

In $\S 1.3$, I have briefly adduced the formal language of category theory as a holisticstructural approach to mathematics which can serve as a promising mediator between the tool set from logical analysis and the tool set from numerical analysis. Furthermore, I have pointed at the potential new opportunity that opens up by employing the category theoretical language in order to complement the primarily numerical analytic perspective in the context of surrogate optimization.

In § 2.1.2, we have made a detour to a structural perspective on a structural property; in § 2.2.2, we have made a detour to a structural perspective on another structural property; and, in § 2.3.2, we have made a detour to a structural perspective on the objective functions. Hence, I have used these detours to show by examples that the formal language of category theory is lurking in the background of some established perspectives on optimization within the electromagnetics context.

In ch. 3, various notions of surrogate optimization have been tagged with algebraic notes. In $\S 3.1 .3$, especially, algebraic tools from the category theoretical language have been anticipated in the elaborations on simplified-physics low-fidelity models. In $\S 3.3 .2$ and in $\S 3.3 .3$, some benefits of a formalization-oriented viewpoint on surrogate-guided optimization have been shown in order to, e.g., recognize hybrid model management strategies or formulate heuristics-driven convergence statements.

In the present chapter, let us head further into the research direction of the formal-ization-oriented viewpoint and aim at strengthening its theoretical foundations.

Firstly, we recapitulate some conceptualities from the previous chapters. Moreover, in addition to the context of validation and verification for the category theoretical language (see, e.g., § 2.3.2), we briefly sketch the emerging research direction of full automation of surrogate-guided optimization (SGO) and how it can serve as another potential context for the category theoretical language. Afterwards, I concisely mention some relevant related works.

Secondly, let us introduce the category theory toolset where we focus on core tools. Initially, we foster some intuition about the toolset and, subsequently, we apply some rigor in the reasoning. Finally, I illuminate a couple of computational facets.

Thirdly, the category theory toolset is used for specifying a general optimization problem and for specifying surrogate-guided optimization methods where the focus
is on methods within the space-mapping paradigm.
Fourthly, I examine other use cases for the category theory toolset related to high-fidelity models and low-fidelity models relevant to applications in electrical engineering.

We close the chapter, and thus also the advance in the research direction of the formalization-oriented viewpoint, at a fork with three open roads for future use cases.

### 4.1 Recapitulating and enlarging the contextual landscape

By recollecting some landmarks, let us concisely recapitulate the contextual landscape so far (cf. § 1.2). Next, we succinctly illuminate the context of full automation (recall § 1.1) of surrogate-guided optimization (recall § 3.3). A vague and intuitive idea underlying full automation of SGO - or, in general, the full automation of surrogate optimization (recall ch. 3) - is that, given an optimization problem by a user, an ideal software system ascertains the "best" (in some sense) SGO approach for the optimization problem at hand. Admittedly, the in-depth investigation of this idea is out of the scope of the present work (cf. the third disclaimer in § 1.3). However, we at least sketch the potential practical contribution of the formal language of category theory to the discussion of this idea. We end the section by naming some relevant related work to our subsequent elaborations.

### 4.1.1 Recapitulating \& the context of full automation of SGO

In engineering applications, the class of surrogate-guided optimization (SGO) methods are gainful in accelerating the numerical search for optimal solutions (see, e.g., [70], [116]). A fundamental assumption regarding SGO schemes is that the overall computational costs of the numerical optimization are dominated by the costs of evaluating the objective function (aka high-fidelity function). Since the aim is to quickly find the high-fidelity function's optimal solution, the following basic ideas of SGO methods arise: (1) Approximate the objective function by one or more surrogate functions (aka low-fidelity functions) - which capture the high-fidelity function's structure and, by design, have much lower evaluation costs than those of the highfidelity function; and (2) draw sparingly on the high-fidelity function.

Understandably, a lot of research effort in the field of surrogate-guided optimization is circled around the numerical properties of the interplay between the highfidelity function and different kinds of surrogate functions. There are two common ways to classify surrogate functions: (\#1) data-fit, simplified-physics, or projection-based; and (\#2) intrusive or non-intrusive - where intrusive means that there is a need to modify the numerical software that is underlying the high-fidelity function. Note that, in general, computational models and non-computational models such as physical experiments are conceivable.

Despite all the advances in this field, there remains a need for investigating how to achieve full automation of surrogate-guided optimization methods - as, for instance, it has been commented in [119, p. 1513]: "Full automation of space mapping and other surrogate-assisted design methods is a necessary condition of widespread acceptance of such methods by the designers and industry." Furthermore, full automation is indirectly described by stressing aspects like "ensuring global convergence, immunity to coarse model inaccuracy, as well as robustness with respect to the surrogate model setup" (cf. [119, p. 1512]). All these aspects are undoubtedly
essential since they are hinting at important common language features underlying numerical analysis; but it seems unlikely that these common language features alone can express all the intuitive associations with the idea of full automation. Thus, there is an opportunity to point the research in this field at a new orthogonal direction, more precisely, at category theoretical language features.

Imagine the following conceivable realization of full automation: An ideal software system selects a surrogate model and chooses the appropriate algorithm for the given optimization problem - without a user's intervention. For this thought experiment, there is a need for a formal language in which one can define the optimization problem, the algorithms and the surrogate models in a suitable way for the software system because it can only deal with well-defined tasks.

I argue that thinking of full automation in terms of software systems shifts the analysis from point by point considerations of single concrete optimization problems to a holistic view of the modeling chain. This holistic perspective sheds some light on the hidden costs of the modeling chain. And it allows to reassess the role of the empirically undeniable savings in terms of high-fidelity function evaluations. Depending on a user's capabilities and experiences, the error-proneness of selfimplemented surrogate-guided optimization schemes will vary heavily. And even a lot of testing cannot prevent potential hidden bugs. Therefore, striving for reproducibility, a thorough cost-benefit analysis could lead a user to stick to classical optimization schemes provided by commercial or non-commercial software systems, and to put more trust in solutions found by methods that have stood the test of time (see, e.g., [35], [158], [142] or [96]).

Let us examine how to reduce the gap between mathematical modeling involved in surrogate guided-optimization and a software system by the formal language of category theory which is a holistic-structural approach to mathematics (see, e.g., [11], [177] or [180]). I admit that, obviously, the restriction to a formal language is a simplification since real-world software systems are vastly more complex and more deeply anchored in physical machines than any mathematical abstraction or formalism could ever capture. Nevertheless, this simplification is reasonable because a crucial part of a software system is the programming language - and a programming language is a formal language, too.

Note that programming languages are a vivid research field within the computer science community. In this context, type theory is an essential theoretical cornerstone that is related to a practical programming language's type system. In the elucidations of the present work, solely a working or rudimentary knowledge of type theory and type systems is supposed. For more details on type theory, see, e.g., [88] and references therein.

The design of a type system determines heavily to which extent it can express all the properties of a well-defined task: At one end of the spectrum, there are dynamically-checked languages (such as, e.g., the MATLAB ${ }^{\circledR}$ PL or the Python PL), where a trend can be seen towards richer types systems for performance reasons (see, e.g., the Julia PL in [26]); at the other end of the spectrum, there are staticallychecked languages (such as, e.g., the Haskell PL or the Agda ${ }^{1}$ PL), where verification reasons have been an impetus for very rich types systems.

Let us conceive a type as a set equipped with an equivalence relation - but without committing to one specific technical concept (see more details about various

[^41]technical concepts, e.g., in [214]); since, in the present chapter, the focus is on category theory, not on type theory. But the connection between the type theoretical language and the category theoretical language can be understood in a very narrow way by means of the Curry-Howard-Lambek correspondence (see, e.g., [167, p. 59ff]) or in a broader way by means of a pragmatic mental model in which a category serves as a model of a (functional) programming language (see, e.g., in [16, p. 20-24]).

Let us put the pragmatic mental model to use such that the category theoretical language can act as a mediating instance between the mathematical modeling involved in surrogate-guided optimization and the type-theoretical aspects of a programming language. From this point of view, other formalization approaches, e.g., Hilbert spaces, manifolds, and similar, can be thought of as domain-specific languages embedded in the host language (or "general-purpose language") of category theory (cf. [102]).

The category theoretical language can offer a language-focused comparison of different surrogate-guided optimization methods - which has a very practical benefit; because, at least in the field of computational electromagnetics, there is a lack of well-defined equivalence classes of benchmarks that could enable a standardized benchmark-focused comparison. For instance, the popular T.E.A.M. (Testing Electromagnetic Analysis Methods) problems do not seem to be used much in the research field of surrogate-guided optimization.

Aiming at strengthening the theoretical foundations of the formalization-oriented viewpoint on surrogate optimization, one main concern in this chapter is the development of a novel comparison toolset for surrogate-guided optimization methods by explicitly category theoretical (CT) language features.

I illustrate the CT approach by discussing the space mapping paradigm's basic building blocks (recall §3.3.2) within the frame of model management strategies (recall § 1.2).

Furthermore, I depict the usefulness of the CT approach with regard to formalization use cases within the electromagnetics context related to simplified-physics low-fidelity models and related to transformations - such as, e.g., coordinate transformations - of a high-fidelity model and a low-fidelity model.

### 4.1.2 Relevant related work

To my best knowledge, in the literature on surrogate-guided optimization in electrical engineering, formalization issues have not been addressed exhaustively. Therefore, the selection of the following articles aims at creating a context in order to make the added value of this chapter's contribution comprehensible.

In [166], the authors try to organize various numerical methods in the fields of uncertainty propagation, statistical inference, and optimization by means of the concept of multifidelity model management. Applying this concept to the field of optimization corresponds to the class of surrogate-guided optimization methods.

Originating as one correction methodology for surrogate functions, the space mapping notion led to a very ramified family of surrogate-guided optimization methods like, e.g., space mapping (see, e.g. [14]), manifold mapping (see, e.g. [56]), and many others (see a survey, e.g., in [126]).

In [166], the authors identify correction methodologies like space mapping or the first-order approximation and model management optimization (AMMO) paradigm (see, e.g., [4]) as members of one class of multifidelity model management strategies: adaptation; i.e., during the optimization process, the low-fidelity model is adapted by high-fidelity model's information.

In [123], the authors present an automated low-fidelity model selection based on correlation analysis between low- and high-fidelity models for surrogate-guided design optimization of antennas.

Various approaches to automated algorithm selection based on machine learning techniques are discussed in [113]. Category theoretical approaches to machine learning and its underlying mathematical concepts (like Bayesian probability) have been conducted in, e.g., [68] or [52]. In [58], automatic differentiation - a key concept in machine learning techniques - is discussed in a functional language environment.

Using the idea of object-oriented coding, the authors in [111] discuss how to implement finite element software systems that imitate the mathematical structures of Maxwell's equations. The emphasis on the mathematical structures appears very fruitful in the realm of computational electromagnetism (see, e.g., [205], [174], [8], [28], [32]).

In [65], the author discusses various category theoretical ideas in the realm of general software engineering that are beyond the scope of the present work that limits itself to ideas in programming languages (see, e.g., [219]).

In order to justify certain statistical modeling approaches, the author in [148] offers a precise mathematical definition of a statistical model by the language of category theory.

### 4.2 Category theory toolset

Since the formal language of category theory (CT) operates at an even higher level of abstraction than the languages such as functional analysis or differential geometry (recall ch. 2), the first step is to foster some intuition regarding the category theoretical language. From an application-driven viewpoint, this intuition is valuable in order to better comprehend the nature of the problems where the category theoretical language shines, and therefore to better anticipate its potential practical benefits.

The subsequent step is to apply some rigor to the intuitive reasoning about category theory. The corresponding elaborations rely only on elementary notions of category theory, more precisely, no deep theorems of category theory are applied. It is primarily used as a strong and stable notational scaffolding, especially, by diagrams of arrows (see, e.g., [144, p. 1ff]).

At the end, some computational facets regarding the category theoretical language are illuminated.

### 4.2.1 Fostering some intuition

In order to harness partly the high level of abstraction regarding the category theoretical language, it is useful to recall occasionally some common concrete perspectives on categories.

From one perspective, a category can be viewed as a kind of algebraic structure like a group or a vector space. A vector space models linearity, a group models symmetry, and a category models composition.

Another perspective is to consider a category as a mathematical context. In the domain of linear algebra, for instance, the language of matrix dimensions and matrices and the language of vector spaces and linear maps describe essentially the same underlying structure and properties of this domain that can be encoded by an equivalence of the corresponding categories.

Category theory emphasizes rather the structure-preserving maps between objects than the objects themselves; e.g., it emphasizes rather the linear maps between vector spaces than the vector spaces themselves. This observation is reflected in a third perspective in which a category $\mathcal{A}$ encodes a syntax and another category $\mathcal{B}$ encodes a semantics and the structure-preserving map F from $\mathcal{A}$ to $\mathcal{B}$ encodes an interpretation of $\mathcal{A}$ within $\mathcal{B}$ (see, e.g., [136, p. 166ff]). The map $\mathrm{F}: \mathcal{A} \rightarrow \mathcal{B}$ is called a functor; and, most abstractly, it is a tool for comparing categories (see, e.g., [36]).

Inferring from the second and the third perspective, a functor can also be regarded as an interpretation of one mathematical context within another. A different interpretation can be encoded in a different functor $G: \mathcal{A} \rightarrow \mathcal{B}$. The need for a comparison of the two interpretations F and B is covered by the notion of a natural transformation. ${ }^{2}$


The diagram of arrows in (4.1) unites the very basic tools of category theory: categories $(\mathcal{A}, \mathcal{B})$, functors ( $\mathrm{F}, \mathrm{G}$ ), and natural transformations $(\alpha)$.

### 4.2.2 Applying some rigor

All following definitions build upon first-order logic and a primordial concept of a set, i.e., no specific axiomatic system of set theory is used. In addition, if one wants to express a proposition, let us use the set membership symbol " $\epsilon$ "; and if one wants to express a judgment, let us use the type annotation symbol ":". For instance, "one is an element of the natural numbers" is a judgment, hence, one would write " $1: \mathbb{N}$ ". Similarly, in a statically-checked language (recall §4.1.1), "one is an element of the integer type" is a judgment, not a proposition; hence, one would write " 1 : Int". Regarding further logical technicalities (hierarchy of universes, Grothendieck universes, axiom of choice, and similar), I refer to, e.g., [214] and references therein.

Let us provide a definition of a category by emphasizing its three constituting parts (data, structure, laws) as a mathematical entity.
Definition 4.2.1 (Category). A category $\mathcal{C}$ is constituted by

- data:
- a collection $\operatorname{obj}(\mathcal{C})$ of objects $X, Y, Z, \ldots$
- $\forall X, Y: \operatorname{obj}(\mathcal{C}) \exists$ a set $\operatorname{hom}_{\mathcal{C}}(X, Y)$ of morphisms (or arrows or "structurepreserving" maps) $f, \tilde{g}, \tilde{\tilde{h}}, \ldots$
* notation (general) $f: \operatorname{hom}_{\mathcal{C}}(X, Y), f: Y^{X}$,
* notation (specific) $f: X \rightarrow Y, X \xrightarrow{f} Y$
* all $\operatorname{hom}_{\mathcal{C}}(X, Y)$ are pairwise disjoint;
- structure:
- $\operatorname{dom}(f)$ is the domain $X$ of morphism $f$, $\operatorname{cod}(f)$ is the codomain $Y$ of morphism $f$

[^42]- $\forall X: \operatorname{obj}(\mathcal{C}) \exists$ an identity morphism $\operatorname{id}_{X}: \operatorname{hom}_{\mathcal{C}}(X, X)$
- $\forall f: \operatorname{hom}_{\mathcal{C}}(X, Y) \forall g: \operatorname{hom}_{\mathcal{C}}(\operatorname{cod}(f), Z)$
$\exists$ a composite morphism $g \circ f: \operatorname{dom}(f) \rightarrow \operatorname{cod}(g)$;
- laws:

$$
\begin{aligned}
- & \forall f: \operatorname{hom}_{\mathcal{C}}(X, Y) . \operatorname{id}_{Y} \circ f \equiv f \equiv f \circ \operatorname{id}_{X} \text { (unity) } \\
- & \forall f: \operatorname{hom}_{\mathcal{C}}(X, Y) \forall g: \operatorname{hom}_{\mathcal{C}}(\operatorname{cod}(f), Z) \\
& \forall h: \operatorname{hom}_{\mathcal{C}}(\operatorname{cod}(g), W) \exists h \circ g \circ f: \operatorname{dom}(f) \rightarrow \operatorname{cod}(h) . \\
& h \circ(g \circ f) \equiv(h \circ g) \circ f(\text { associativity }) .
\end{aligned}
$$

Remark 4.2.1. The notion "structure-preserving" originates from considerations of structured sets and the structure-preserving functions between them. But morphisms do not necessarily have to be structure-preserving functions - or functions at all.

Remark 4.2.2. By simply turning around all the arrows of a category $\mathcal{C}$, one can define the dual or opposite category $\mathcal{C}^{\text {op }}$ that molds the duality principle.

Remark 4.2.3. Addressing the issue concerning the size of a category, a category is called small if both obj $(\mathcal{C})$ and hom $(X, Y)$ are sets. Moreover, let us suppose primarily locally small categories such that, at least, for all pairs of objects $X, Y$, the collection of morphisms between them, i.e., hom $\mathcal{C}_{\mathcal{C}}(X, Y)$, is a set-called a hom-set.

Some representations or implementations of categories are:

- Set (the category of (finite) sets and set functions),
- Top (the category of topological spaces and continuous maps),
- Man ${ }^{\infty}$ (the category of smooth manifolds and $\infty$-times continuously differentiable maps),
- $V^{f} t_{\mathrm{k}}^{\mathrm{fd}}$ (the category of finite dimensional vector spaces over the field $k$ and $k$-linear maps),
- TVect $_{k}$ (the category of topological vector spaces over the topological field $k$ and continuous $k$-linear maps),
- $V^{\text {ect }}{ }_{\mathrm{k}}^{\mathrm{basis}}$ (the category of finite dimensional vector spaces over the field $k$ with chosen basis and $k$-linear maps), and
- Mat $_{k}$ (the category of non-zero natural numbers [matrix dimensions $(m, n)$ ] and $m \times n$-matrices with values in the field $k$ ).

Bear in mind that the category's definition encompasses merely its specification, i.e., its minimal amount of essential characteristics; thus, various real encounters, e.g., inverse maps or products, lead to new features - like a category $\mathcal{C}$ with isomorphisms.

In the subsequent elaborations, it is supposed that the maps $f$ and $g$ are $a d h o c$ polymorphic or overloaded. Thus, let us invoke the maps $f$ and $g$ with different signatures.

Notice well that isomorphisms are special morphisms that allow to encode the idea of sameness of objects. A morphism $f: X \rightarrow Y$ is called an isomorphism if and only if there exits a morphism $g: Y \rightarrow X$ such that, in equational form, $f \circ g=\mathrm{id}_{Y}$
and $g \circ f=\mathrm{id}_{X}$. In diagrammatic form, one can choose exemplarily the subsequent representation:

where the symbol $\bigcirc$ is used to denote a commutative diagram. Furthermore, the objects $X$ and $Y$ are called isomorphic (symbolically: $X \cong_{\mathcal{C}} Y$ ). If the objects are isomorphic, then they are indistinguishable, hence, it possible to substitute one object with the other.

A morphism whose domain is identical with its codomain, that is, a morphism $f$ : $X \rightarrow X$ such that

$$
\begin{equation*}
\operatorname{dom}(f) \equiv \operatorname{cod}(f) \tag{4.3}
\end{equation*}
$$

is called an endomorphism. An endomorphism which is also an isomorphism (see the diagram in (4.2)) is called an automorphism.

One can construct a notable example for isomorphic objects by providing a category $\mathcal{C}$ with product objects $A \times B$. More precisely, given $A, B: \operatorname{obj}(\mathcal{C})$, then the object $A \times B: \operatorname{obj}(\mathcal{C})$ equipped with a pair of morphisms $\pi_{A}: A \times B \rightarrow A$ and $\pi_{B}: A \times$ $B \rightarrow B$ is a (binary) product object, if and only if it satisfies a universal mapping property (UMP), i.e.,

$$
\begin{equation*}
\forall P: \operatorname{obj}(\mathcal{C}) . \forall f: P \rightarrow A \cdot \forall g: P \rightarrow B \cdot \exists!h: P \rightarrow A \times B \tag{4.4a}
\end{equation*}
$$

such that the subsequent diagram commutes


Let us refer to the morphism $h$ as the (binary) product of the morphisms $f$ and $g$ (symbolically: $h:=\langle f, g\rangle)$. Finally, assuming ternary product objects in the category Set, a notable example for isomorphic objects is

$$
\begin{equation*}
\forall A, B, C: \operatorname{obj}(\text { Set }) .(A \times B) \times C \cong \operatorname{Set} A \times(B \times C) . \tag{4.5}
\end{equation*}
$$

In order to show that a parallel pair of morphisms $f, g: X \rightrightarrows Y$ is equal, i.e.,

$$
\begin{equation*}
f=\tilde{\mathcal{C}} g, \tag{4.6}
\end{equation*}
$$

one needs a category $\tilde{\mathcal{C}}$ with a terminal object $\mathbf{1}$. In a category $\tilde{\mathcal{C}}$, an object is terminal if and only if for all objects $X: \operatorname{obj}(\tilde{\mathcal{C}})$, there exists a unique morphism $X \rightarrow \mathbf{1}$. Invoking the duality principle (see Remark 4.2.2), there is an initial object $\mathbf{O}$ as well. In a category $\tilde{\mathcal{C}}$, an object is initial if and only if for all objects $X: \operatorname{obj}(\tilde{\mathcal{C}})$, there exists a unique morphism $\mathbf{O} \rightarrow X$.

Mind that, in category theory, there is no global set-membership relation; thus, the idea of an element " $x \in X$ " is encoded in a map $x$ such that $x: \mathbf{1} \rightarrow X$ or $\mathbf{1} \xrightarrow{x} X$. Hence, if for all morphisms $\mathbf{1} \xrightarrow{x} X$, the equation $f \circ x={ }_{1 \rightarrow Y} g \circ x$ holds to be true, then $f=\tilde{\mathcal{C}} g$. If the morphisms are equal, then they are indistinguishable, hence, it is
possible to substitute one morphism with the other.
This kind of test for equality w.r.t. (4.6) corresponds to the common extensional equality of functions (recall Remark 2.2.2) in which one treats functions as black-boxes such that one only considers their input-output behavior. It is important to note that this equality problem is algorithmically decidable only if the domain is finite and relatively small. If one is also interested in how the output is calculated, i.e., the particular "formulas" of the functions, one needs to consider intensional equality of functions. Finally, if one wants to know whether two functions point to the exact same instance in computer memory, one needs to invoke the notion of referential equality of functions.

Observe that the terminal object 1, if it exists, then it is only unique up to isomorphism. But, more importantly, it is part of objects (like the product object) that follow the principle of universality. Hence, the definition of objects such as the terminal object is based on a universal mapping property that is similar to (4.4) where the object's existence and uniqueness is related to all other objects of the category.

Given a category $\mathcal{C}$ and a category $\mathcal{D}$, one can define a map from $\mathcal{C}$ to $\mathcal{D}$ that preserves the structure of the category $\mathcal{C}$ (recall Definition 4.2.1) within the category $\mathcal{D}$. This structure-preserving map between two categories is called a functor.

Definition 4.2.2 (Functor). Given a category $\mathcal{C}$ and a category $\mathcal{D}$, a functor $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}$ is constituted by

- a map on the data of $\mathcal{C}$ and $\mathcal{D}$, i.e., a functor's assignment rule reads as
- $\forall X: \operatorname{obj}(\mathcal{C}) . \exists \mathrm{F}(X): \operatorname{obj}(\mathcal{D})$
- $\forall f: \operatorname{hom}_{\mathcal{C}}(X, Y) \cdot \exists \mathrm{F}(f): \operatorname{hom}_{\mathcal{D}}(\tilde{X}, \tilde{Y})$;
- that preserves the structure of $\mathcal{C}$ within $\mathcal{D}$, i.e., the functor laws read as

```
- \(\tilde{X} \equiv \mathrm{~F}(X), \tilde{Y} \equiv \mathrm{~F}(Y)\)
- \(\forall X: \operatorname{obj}(\mathcal{C}) . \mathrm{F}\left(\mathrm{id}_{X}\right) \equiv \mathrm{id}_{\mathrm{F}(X)}\)
\(-\forall g \circ f: \operatorname{hom}_{\mathcal{C}}(\operatorname{dom}(f), \operatorname{cod}(g)) . \mathrm{F}(g) \circ \mathrm{F}(f) \equiv \mathrm{F}(g \circ f)\).
```

Remark 4.2.4. The map F is ad hoc polymorphic or overloaded in the sense that its signature $\mathcal{C} \rightarrow \mathcal{D}$ is accompanied with the signature $\operatorname{obj}(\mathcal{C}) \rightarrow \operatorname{obj}(\mathcal{D})$ and the signature $\operatorname{hom}_{\mathcal{C}}(X, Y) \rightarrow \operatorname{hom}_{\mathcal{D}}(\tilde{X}, \tilde{Y})$.

Remark 4.2.5. Observe that the defined functors are called covariant in order to distinguish them from contravariant functors $\mathrm{F}: \mathcal{C}^{\mathrm{Op}} \rightarrow \mathcal{D}$ - with $\mathcal{C}^{\mathrm{op}}$ being the opposite category (see Remark 4.2.2) - where the direction of the arrows in the domain-category $\mathcal{C}^{\mathrm{Op}}$ is swapped in the codomain-category $\mathcal{D}$.

The aforementioned definitions of a category and a functor appear a bit cumbersome, but they unfold their power when we shift from this algebraic to a geometric description. If we consider a category as a directed multigraph equipped with an algebra of paths - the identity morphisms corresponding to 0-paths, the morphisms corresponding to 1-paths, and the composition corresponding to 2-paths -, then one can consider a functor as a graph-morphism that preserves paths. For the sake of brevity, let us not define categorically the concepts graph, and similar. For more category theoretical details regarding these concepts, I refer to, e.g., [144], [11] or [177] and references therein.

Let us assume two finite categories $\mathcal{A}$ and $\mathcal{B}$. Note that a finite category has only a finite number of objects, identity arrows, and non-identity arrows. Furthermore, let
us suppose a functor $\mathrm{F}: \mathcal{A} \rightarrow \mathcal{B}$. Hence, the diagram in (4.7) illustrates the geometric view regarding categories and functors.


Moreover, the diagram in (4.7) elucidates geometrically that all functors preserve isomorphisms, that is, if the commutative diagrams in (4.2) exist, then all functors preserve these commutative diagrams - in the sense that, if the objects $X, Y \in \operatorname{obj}(\mathcal{A})$ are isomorphic, then

$$
\begin{equation*}
X \cong \cong_{\mathcal{A}} Y \Longrightarrow \mathrm{~F}(X) \cong_{\mathcal{B}} \mathrm{F}(Y) \tag{4.8}
\end{equation*}
$$

Although all functors preserve isomorphisms, it is not necessarily true that they reflect isomorphisms, i.e., an isomorphism in the functor's codomain does not mean necessarily that the corresponding morphism in the functor's domain is an isomorphism.

Functors can have various attributes. Two very useful examples are forgetful functors that forget all or only some of the algebraic structure such as, e.g.,

$$
\begin{align*}
& \mathrm{F}: \operatorname{Vect}_{\mathrm{k}}^{\mathrm{fd}} \rightarrow \text { Set, }  \tag{4.9a}\\
& \mathrm{G}: \operatorname{Vect}_{\mathrm{k}}^{\text {basis }} \rightarrow \text { Vect }_{\mathrm{k}}^{\mathrm{fd}}, \tag{4.9b}
\end{align*}
$$

and faithful functors $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}$ who possess the defining property
for all $X, Y: \operatorname{obj}(\mathcal{C})$, the map $F: \operatorname{hom}_{\mathcal{C}}(X, Y) \rightarrow \operatorname{hom}_{\mathcal{D}}(F(X), F(Y))$ is injective. ${ }^{3}$
Forgetful and faithful functors are very useful since they help to pin down the concept of a structured set, i.e., a set equipped with extra structure; hence, a structured set is an object of a category $\mathcal{C}$ equipped with a faithful functor $\mathrm{F}: \mathcal{C} \rightarrow$ Set. Note that, in the setting of structured sets, the category $\mathcal{C}$ is called a concrete category. Some examples of concrete categories are the abovementioned categories Top, Man ${ }^{\infty}$, and Vect ${ }_{k}^{\mathrm{fd}}$. The category Mat $\mathrm{t}_{\mathrm{k}}$ is not a concrete category.

Recalling § 4.2.1 for the case of structured sets, one can observe that categories such as Top, Man ${ }^{\infty}$ or Vect ${ }_{k}^{\mathrm{fd}}$ encode a syntax and the category Set provides a semantics. A faithful functor between a syntax category and a semantics category encodes an interpretation of a syntax category within a semantics category.

If one wants to compare two different interpretations, then one needs a proper notion of comparing two different functors. Hence, let us define properly the notion of an arrow between two functors, i.e., let us define a natural transformation.

Definition 4.2.3 (Natural transformation). Given a category $\mathcal{C}$, a category $\mathcal{D}$, a functor $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}$, and a functor $\mathrm{G}: \mathcal{C} \rightarrow \mathcal{D}$, then a natural transformation $\alpha: F \Rightarrow \mathrm{G}$ comprises

- the components of $\alpha$ at $X$-, i.e., a family of morphisms $\alpha_{X}$ in $\mathcal{D}$ :

$$
-\forall X: \operatorname{obj}(\mathcal{C}) \exists \alpha_{X}: \mathrm{F}(X) \rightarrow \mathrm{G}(X)
$$

- such that $\forall f: \operatorname{hom}_{\mathcal{C}}(X, Y)$, i.e., $X \xrightarrow{f} Y$,

[^43]\[

$$
\begin{aligned}
& -\alpha_{Y} \circ \mathrm{~F}(f)=\mathrm{G}(f) \circ \alpha_{X} \text { in } \mathcal{D} \quad \text { or } \\
& \begin{array}{llll} 
& \mathrm{F}(X) \xrightarrow{\alpha_{X}} \mathrm{G}(X) \\
\text { - the diagram } & { }_{\mathrm{F}}(f) & 0 & \downarrow^{\mathrm{G}}(f) \text { commutes in } \mathcal{D} \text {. }
\end{array} \\
& \mathrm{F}(Y) \xrightarrow{\alpha_{Y}} \mathrm{G}(Y)
\end{aligned}
$$
\]

Remark 4.2.6. The lack of naturality (i.e., the lack of a natural transformation) is an even more interesting observation than the existence of naturality - because, in the former case, there is a need for an "unnatural choice of basis" in some sense that is motivated by the relation between a finite-dimensional vector space and its dual and its double dual.
Remark 4.2.7. If we apply the geometric view such as in (4.7), then one can conceive a natural transformation as a comparison of constructed paths.

Remark 4.2.8. A natural transformation $\alpha: \mathrm{F} \Rightarrow \mathrm{G}$ is a natural isomorphism (symbolically: $\alpha: \mathrm{F} \cong \mathrm{G}$ ) if and only if all components $\alpha_{X}$ are isomorphisms. Let us omit writing the corresponding category, that is, the so-called functor category $\mathcal{D}^{\mathcal{C}}$ where the objects are the functors $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}, \mathrm{G}: \mathcal{C} \rightarrow \mathcal{D}, \ldots$, and the morphisms are natural transformations $\alpha: \mathrm{F} \Rightarrow \mathrm{G}$. Hence, a natural isomorphism can symbolically be written as $\alpha: \mathrm{F} \cong_{\mathcal{D}^{c}} \mathrm{G}$.

Using the notion of natural isomorphisms in order to define the concept of equivalence of categories, we are able to tackle the question of when two categories are essentially the same.
Definition 4.2.4 (Equivalence of categories). An equivalence of categories $\mathcal{C}, \mathcal{D}$ (symbolically: $\mathcal{C} \simeq \mathcal{D}$ ) is constituted by

- functors $\mathrm{F}: \mathcal{C} \rightleftarrows \mathcal{D}: \mathrm{G}$ and
- natural isomorphisms $\eta: \mathrm{id}_{\mathcal{C}} \cong \mathrm{G} \circ \mathrm{F}$ and $\epsilon: \mathrm{F} \circ \mathrm{G} \cong \mathrm{id}_{\mathcal{D}}$.

Remark 4.2.9. Analogously to the identity morphism in Definition 4.2.1, the functors $\mathrm{id}_{\mathcal{C}}$ and $\mathrm{id}_{\mathcal{D}}$ denote the identity functors corresponding to the categories $\mathcal{C}$ and $\mathcal{D}$, respectively.
Remark 4.2.10. If and only if there is an equivalence of categories, then the functor G is called the inverse functor to the functor F .

Remark 4.2.11. If two categories are equivalent, then they are indistinguishable regarded as categories, hence, it is possible to substitute one category with the other.

One can state stronger forms of sameness, e.g., one can state equality $\mathcal{C}=\mathcal{C}$; or, one can replace the demand for natural isomorphisms in Definition 4.2.4 by the demand for F to be an isomorphism regarding the category Cat, i.e., the category of (locally) small categories (recall Remark 4.2.3) where the objects are (locally) small categories $\mathcal{C}, \mathcal{D}, \ldots$, and the morphisms are functors $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}$. Hence, a functor $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}$ is called an isomorphism w.r.t. Cat if and only if there exits a morphism $G: \mathcal{D} \rightarrow \mathcal{C}$ such that $\mathrm{id}_{\mathcal{C}}=\mathrm{G} \circ \mathrm{F}$ and $\mathrm{F} \circ \mathrm{G}=\mathrm{id}_{\mathcal{D}}$.

However, these two stronger forms of sameness are less helpful since they restrict unnecessarily the available expressive power. For instance, the second form tells us that, by going around through F and G , we will land at the exact same starting point. In most circumstances, though, we will land at a spot resembling our starting point; and this idea of resembling is encoded by demanding natural isomorphisms in Definition 4.2.4.

There is also a weaker form of sameness that replaces the natural isomorphisms in the definition above by natural transformations (id ${ }_{\mathcal{C}} \stackrel{\eta}{\Rightarrow} \mathrm{G} \circ \mathrm{F}$ and $\mathrm{F} \circ \mathrm{G} \stackrel{\epsilon}{\Rightarrow} \mathrm{id}_{\mathcal{D}}$ )
and some coherence conditions (triangle identities). This process leads to the notion of adjunction.

Regarding linear algebra (recall § 4.2.1), the important equivalence of categories is captured by

$$
\begin{equation*}
\text { Vect }_{k}^{\mathrm{fd}} \simeq \operatorname{Mat}_{\mathrm{k}_{k}} \text { (for any field } \mathbb{k} \text { ). } \tag{4.11}
\end{equation*}
$$

For more details regarding the proof of (4.11), I refer to [177, p. 33]. Remember that, in the setting of structured sets, the category Vect ${ }_{k}^{\text {fd }}$ is a concrete category and the category Mat ${ }_{k}$ is not a concrete category. However, the statement in (4.11) regards Vect ${ }_{k}^{\mathrm{fd}}$ and Mat ${ }_{k}$ solely through the lens of a category (recall Definition 4.2.1).

The diagram in (4.12) depicts a possible interplay of some aforementioned categories.


Notice well that most of the functors in (4.12) are forgetful. However, the functor D is forgetful and faithful and full. Full functors $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}$ possess the defining property
for all $X, Y: \operatorname{obj}(\mathcal{C})$, the map $F: \operatorname{hom}_{\mathcal{C}}(X, Y) \rightarrow \operatorname{hom}_{\mathcal{D}}(F(X), F(Y))$ is surjective.
In general, a diagram such as in (4.12) is not necessarily commutative.

### 4.2.3 Computational facets

The notion of computation in the category theoretical language depends on the chosen context, hereby reflecting the observation that the programming language's point of view on computation differs from the differential geometric's or functional analytic's.

In a programming language setting, (typed) $\lambda$-calculus is a model of computation. A typed $\lambda$-calculus (see, e.g., [15]) is, directly or indirectly, constituting the core of functional and imperative programming languages. ${ }^{4}$

In a differential geometric setting, a connection to computation is established by moving from the manifold level to the chart level - and hence ultimately, to the field of real numbers and matrices over the field of real numbers. Observe that, for instance, the authors in [117] give a category theory oriented exposition of classical differential geometry.

In a functional analytic setting, one has a concept of "a space approximating another space" and a notion of error, but ultimately, the connection to computation is established by moving to the field of real numbers and matrices over the field of real numbers.

Observe that approximating is related to iterating a function; and iterating a function is related to composing functions. Furthermore, observe that, in general, the notion of error is formalized in the context of complete normed vector spaces, more

[^44]precisely, Banach spaces. Notice that there is a well-designed connection between Banach space theory and category theory (see, e.g., [43]). Finally, observe that there is a common tension arising in theories using real numbers caused by the application to machines - which, inevitably, recourse to floating point arithmetic - and a leap of faith that the theory's high-level properties still hold to be true on the machine. Moreover, there is a tension between a theory's high-level data structures and their programming language counterparts (cf. [133]).

Relating the category theoretical language to programming languages, one observation is the correspondence between a cartesian closed category (CCC) and a typed $\lambda$-calculus. The defining properties specifying a CCC are the existence of a terminal object, a product object, and an exponential object. For a definition of an exponential object and a cartesian closed category, see, e.g., [167, p. 33f] and [167, p. 53-57], respectively.

Another observation is that a category can serve as a model for a functional programming language. Note that, in addition to exploiting categorical definitions for guidance on how to organize software, in recent times, the extension of functional programming languages with dependent types (see, e.g., the Agda PL or the Idris PL) enables implementing immediately categorical definitions by using software.

For technical details regarding dependent types, I refer to, e.g., [214]. However, recalling Definition 4.2.1, an instance of a dependent type is the hom-set type hom $_{\mathcal{C}}(X, X)$ since its definition depends on the value $X$ of the obj type obj( $\left.\mathcal{C}\right)$. Mostly, though, dependent types are used to encode the quantifiers " $\forall$ " and " $\exists$ ". Although implementing immediately categorical definitions by using software is a promising approach, it is still a very young approach (see, e.g., [72]); thus, its industrial application is not big yet.

Relating the category theoretical language to matrix-focused environments, the key observation is the equivalence in (4.11) which has to be adapted for a computational context in the sense that

$$
\begin{equation*}
\text { Vect }_{k}^{\mathrm{fd}} \simeq \text { Mat }_{k_{k}} \text { (for any computable field } \mathbb{k} \text { ). } \tag{4.14}
\end{equation*}
$$

Thus, the category Mat ${ }_{k}$ is used as a computational model for the category Vect ${ }^{\mathrm{fd}}$. Moreover, observe that the categories Vect $\mathrm{fd}_{\mathrm{k}}$ and Mat ${ }_{k}$ are examples of so-called Abelian categories (see, e.g., [144, ch. VIII Abelian Categories] or [71]) in which the notion of adding morphisms (and objects, respectively) exists.

For instance, the authors in [143] utilize the category Mat ${ }_{k}$ in order to develop a simple categorical type system for concrete linear algebra. Another example is the CAP project (cf. [168], [86]) which devotes itself to enable computing within a category, i.e., to calculate objects, morphisms etc. of a given representation of a category via a computational model such as, e.g., the category Mat ${ }_{k}$. The CAP project encompasses various software packages implemented in GAP (a system for computational discrete algebra). Note that GAP does not support dependent types.

In the zoo of dynamically-checked programming languages, there is, e.g., the new, still developing, Julia PL package Catlab. $j 1$ (see [163]) that is part of the larger open source software project called AlgebraicJulia which focuses on developing category theoretical approaches for technical computing. For more details on this software project, visit https://www.algebraicjulia.org/. However, the expressiveness of the Julia PL's type system is limited to encode fully the formal language of category theory.

Finally, let us address the general question of error in connection with category
theory - which is inevitably related to numerical analysis where, classically, a distinction is made between modeling error, approximation error, and numerical error.

In order to handle numerical errors due to, e.g., number representations as floating point numbers, there are not yet widespread methods such as interval arithmetics (see, e.g., [213], [104]) to produce reliably verified results.

In order to deal with approximation errors - that is, errors due to discrete representations of continuous representations, there is, as already mentioned above, the well-oiled machinery of functional analysis which is invaluable even if one would have computers with exact arithmetic.

Mostly, the flimsiest limb in the chain of errors is the modeling error - because it is most difficult to formally capture this error which is a result of representing a physical problem as a mathematical problem.

And if we consider all kinds of errors, then it is most likely to expect the greatest impact from a CT approach on the modeling error. This consideration is plausible since there is a take on category theory as a "mathematical model of mathematical modeling" (see, e.g., [197]); and, hence, it can provide in some sense a consistency test for the modeling. For instance, in differential geometry, both a linear map and a bilinear map can be represented by a matrix but the matrix associated with the linear map transforms differently under a change of coordinates than the matrix associated with the bilinear map. This important information would go unnoticed by merely focusing on the matrix representation.

Though, it is still too early to expect from a CT approach w.r.t. the modeling error something similar to the so-called fundamental theorem of numerical analysis (see, e.g., [7]).

However, in the upcoming section, driven by heuristics, I propose and discuss shortly a means of quantifying the modeling error by employing a problem-dependent degree of forgetfulness.

### 4.3 Using the CT toolset for SGO methods

In the area of optimization, it is intricate to provide a generally accepted taxonomy of the numerous solution methods. However, if we apply an engineering-driven pragmatism as a classifier, then surrogate-guided optimization methods have gained traction as an outstanding sub-area over the last decades (recall § 4.1).

I argue, though, that using unique category theoretical (CT) language features (recall § 4.2) can help to find a better match for pragmatic notions such as model hierarchies or fidelity - in order to capture more rigorously their intended meaning. By focusing on structure-related issues, CT offers formal methods that complement the usual comparison toolset for SGO methods (recall § 4.1). Due to its tight connection to (functional) programming, the CT approach highlights beneficially a blueprint of a software design where the focus is on the specification (or the interface) which is kept conceptually separated from the implementation.

Let us first devote ourselves to specifying a general optimization problem. Subsequently, we dedicate ourselves to specifying surrogate-guided optimization methods.

### 4.3.1 Specifying a general optimization problem

Recalling § 2.3, let us assume the objects $X, Y$, and $Z$ within a category $\mathcal{C}$, then we prescribe the $Z$-valued objective function as

$$
\begin{equation*}
J=(y, x) \mapsto z: Y \times X \rightarrow Z \tag{4.15}
\end{equation*}
$$

where $x: X$ denotes the control (or input) variable and $y: Y$ denotes the state (or intermediate) variable and $z: Z$ denotes the output variable. Mind that we use the barred arrow notation " $\mapsto$ " for the internal representation of a function and the straight arrow notation " $\rightarrow$ " for the external representation of a function. Notice well that the CT approach exploits primarily the external representation.

In (4.15), it is commonly assumed that there exists a unique control-to-state map $f$ such that

$$
\begin{equation*}
f=x \mapsto y: X \rightarrow Y \tag{4.16}
\end{equation*}
$$

Hence, one can write $J(f(x), x)$ or, assuming that $J$ is ad hoc polymorphic or overloaded, one can write shortly $J(f(x))$. One can generalize the statements in (4.15) and in (4.16) in the sense that one can make the assignment $Y:=B^{A}$ (see §4.2.2) in order to highlight the search for a function of type $A \rightarrow B$. Furthermore, one can make the assignment $X:=X_{1} \times X_{2} \times \ldots X_{n}$ and/or $Z:=Z_{1} \times Z_{2} \times \ldots Z_{n}$ to emphasize the arity of the input-object and the output-object.

Recalling the statements in (3.9) and in (3.10), let us prescribe the $X$-valued minimizer function argmin that reads as

$$
\begin{equation*}
\operatorname{argmin}=J_{f} \mapsto x^{*}:(X \rightarrow Z) \rightarrow X, \tag{4.17}
\end{equation*}
$$

where $x^{*}: X$ denotes the minimizer of the composite function $J_{f}$ with $J_{f} \equiv J \circ f$ and $J \circ f: \operatorname{hom}_{\mathcal{C}}(X, Z)$.

If we invoke set-builder notation and if we utilize an order structure on $Z$, one can define the set of optimal solutions $X^{*}$ as

$$
\begin{equation*}
X^{*}:=\left\{x^{*}: X \mid \forall x: X . J\left(x^{*}\right) \leq J(x)\right\} \tag{4.18}
\end{equation*}
$$

and one can define the set of corresponding optimal co-domain values $Y^{*}$ as

$$
\begin{equation*}
Y^{*}:=\left\{J(x): Z \mid x \in X^{*}\right\} \tag{4.19}
\end{equation*}
$$

which is encoded in the minimizer proposition

$$
\begin{equation*}
\operatorname{minimize} J(x) . \tag{4.20}
\end{equation*}
$$

If we use an order structure on $X$, one can define a set of admissible solutions $X_{\mathfrak{F}}$ due to, e.g., box constraints, as

$$
\begin{equation*}
X_{\mathfrak{F}}:=\left\{x: X \mid \forall x_{l}, x_{u}: X . x_{l} \leq x \leq x_{u}\right\}, \tag{4.21}
\end{equation*}
$$

Supposing a differentiability structure for the maps $J$ and $f$, gradient- or hessianexploiting solution methods can be utilized.

Finally, let us apply the diagrammatic notation from the CT toolset (see, e.g., the
diagram in (4.7)) as a graphical tool for the compact representation of the specification's basic building blocks (4.15), (4.16), and (4.17).


Bear in mind that the statements in (4.18), in (4.19), in (4.20), and in (4.21) are considered as implicitly encoded in the compact representation in (4.22). Furthermore, the diagrammatic notation in (4.22) is primarily only a graphical tool and not a graphical language. In order to have a graphical language, i.e., to enable diagrammatic reasoning (see, e.g. [1] or [45]), completeness theorems are necessary that match the diagrammatic representation with the algebraic representation. However, these kinds of considerations are left for future investigations.

Nevertheless, the abstract specification by means of the CT approach is indispensable since it keeps the objects and morphisms conceptually separated from a possible implementation, e.g., as sets and set functions in a category Set. And, generally, it is conceivable that there is no prevailing way of implementation in different categories. For more details on the method of categorical definition as a kind of abstract specification and the utility of keeping the specification separated from the implementation, see, e.g., [16].

Hereinafter, let us suppose the commonly defined sequential composition such that the corresponding laws of a category are satisfied (recall Definition 4.2.1). Hence, if we consider the three individual branches in (4.22) as depictions of three finite categories $\mathcal{A}, \mathcal{B}, \mathcal{C}$ (similarly to (4.7)), then an interaction between these three finite categories can be conducted by functors. Given exemplarily the functors $\mathrm{F}: \mathcal{A} \rightarrow \mathcal{B}$, $\mathrm{G}: \mathcal{B} \rightarrow \mathcal{C}$, and $\mathrm{H}: \mathcal{A} \rightarrow \mathcal{C}$, one can adapt in a simplistic manner the diagrams in (4.22) to a version that can be drawn as


Especially due to the observation of the differentiation operator as a functor (see, e.g., [177, p. 14f]), there are already some emerging applications of CT ideas in optimization (see, e.g., [191]). Hence, diagrams such as in (4.23) furnish us with a fresh and novel perspective on a general optimization problem.

Between the first branch $(\mathcal{A})$ and the second branch $(\mathcal{B})$ in (4.23), a forgetful functor F - such as, e.g., in (4.9) - can be deployed that forgets the product object. Interestingly, one cannot provide an appropriate functor in order to show an equivalence of categories (recall Definition 4.2.4). According to [177, p. 30f], a functor F:C $\rightarrow \mathcal{D}$
has to be faithful (such as in (4.10)), and full (such as in (4.13)), and essentially surjective on objects in order to define an equivalence of categories, that is, $\mathcal{C} \simeq \mathcal{D}$. Mind that a functor $\mathrm{F}: \mathcal{C} \rightarrow \mathcal{D}$ that is essentially surjective on objects possesses the defining property

$$
\begin{equation*}
\text { for all } X: \operatorname{obj}(\mathcal{D}), \text { there exists } \tilde{X}: \operatorname{obj}(\mathcal{C}) \text { such that } X \simeq_{\mathcal{D}} F(\tilde{X}) .^{5} \tag{4.24}
\end{equation*}
$$

In conclusion, one cannot treat the first branch $(\mathcal{A})$ and the second branch $(\mathcal{B})$ in (4.23) as indistinguishable regarded as categories and, therefore, one cannot substitute one with the other.

If we look at the arrow $Y \times X \xrightarrow{J} Z$ isolated within the corresponding category $(\mathcal{A})$ and if we provide appropriate functors F and G , one can observe a natural isomorphism $\alpha: F \cong G$ (recall Remark 4.2.8) that relates the arrow $Y \times X \xrightarrow{J} Z$ to another arrow $Y \xrightarrow{\hat{h}} Z^{X}$. In computer science, such a specific natural isomorphism $\alpha: F \cong G$ is called currying (see, e.g., [177, p. 54]). For more elaborations regarding the constructions behind currying, I refer to, e.g., [167, p. 33f] or [177, p. 129].

Exploiting the information regarding the first branch $(\mathcal{A})$ and the third branch $(\mathcal{C})$ in (4.23), let us suppose a category $\mathcal{D}$ in which one can draw a diagram such that


Observe that performance-related issues regarding a general optimization problem boil down to a matter of executing the composition in (4.25) as efficient as possible. Recalling Figure 1.4, the CT approach helps therefore with organizing concisely the needed knowledge at the level of generalized functions which precedes the level of associated algorithms which in turn precedes the level of programs that are corresponding implementations in a programming language. In informal parlance, one can say that programs are a subpart of algorithms and algorithms are a subpart of generalized functions. For more details, I refer to, e.g., [225].

Obviously, the CT approach is too abstract to utilize the standard analysis toolset; but it enables a complementary useful rigor in reasoning by highlighting an adequate specification.

Recalling for the sake of technicalities the category Cat which is the category of (locally) small categories as objects and functors as morphisms (see § 4.2.2), then, notably, the CT approach undertakes meaningfully a shift in perspective by lifting the classical set-oriented modeling paradigm "(set) functions as models" onto the category-oriented modeling paradigm: "Categories as models" and "functors as model transformations".

### 4.3.2 Specifying surrogate-guided optimization methods

At the level of generalized functions (see Figure 1.4), surrogate-guided optimization (SGO) methods try to address the aforementioned performance-related issue by introducing a fidelity notion. Inevitably, it adds complexity to the inherent complexity of the general optimization problem (see the diagram in (4.25)) by forcing the use of various models that exhibit some kind of fidelity, that is, some kind of indexing.

[^45]In a nutshell, I argue that there are two kinds of fidelity: An order-oriented fidelity and a hierarchy-oriented fidelity. Both kinds are based on a user-defined indexing driven by the idea of a "best fit" to physics' semantics; but the hierarchy-oriented fidelity demands additionally a meaningful limit definition. The hierarchy-oriented fidelity is at the core of multi-level methods, the order-oriented fidelity is at the core of multi-fidelity methods. In the remaining, I will exclusively dwell on the multi-fidelity methods.

To give an example of the "best fit" idea, let us use a semantics of electromagnetics to interpret the syntactical considerations in the previous subsection - especially in (4.15): the control variable $x$ as geometric parameters, the state variable $y$ as a current density vector field, the map $f$ as an assignment to a solution of a wellposed boundary value problem (WPBVP), and the map $J$ as a loss map. Solving numerically the WPBVP allows, e.g., to associated a "best fit" model with the highest degree of discretization; and a "low fit" model with a low degree of discretization hence, let us call the later a low-fidelity model and the former a high-fidelity model. Exploiting the category-oriented paradigm shift mentioned above in $\S 4.3 .1$, there are two options for representing a model. For the sake of conciseness, let us omit the identity arrows in the depictions:

$$
\begin{equation*}
X \xrightarrow{f} Y \xrightarrow{J} Z \quad \text { or } \quad X \xrightarrow{J_{f}} Z \tag{4.26}
\end{equation*}
$$

The option to choose depends on the information available. If we select, for example, the first model representation as our high-fidelity model $\mathcal{M}_{1}$, then one can ascribe a low-fidelity model $\mathcal{M}_{2}$ to the high-fidelity model by using a functor F .

$$
\begin{gather*}
X \xrightarrow{f} Y \xrightarrow{\text { F }} \mathrm{F}  \tag{4.27}\\
\mathrm{~F}(X) \xrightarrow{\mathrm{F}} \mathrm{~F}(Y) \xrightarrow{\mathrm{F}(J)} \mathrm{F}(Z)
\end{gather*}
$$

Given the functor $\mathrm{F}: \mathcal{M}_{1} \rightarrow \mathcal{M}_{2}$, one can interpret (recall § 4.2.1) the high-fidelity model $\mathcal{M}_{1}$ within the low-fidelity model $\mathcal{M}_{2}$. Hence, the CT approach gives a clear mathematical meaning to the colloquial idea in SGO methods that the models of variable fidelity should share some structural similarity.

Moreover, it establishes a good level of abstraction to encompass formally all kinds of models - and the models' interconnection - which are conceivable in SGO methods (recall §4.1).

In particular, the CT approach offers guidance to develop formalized guarantees regarding the models in SGO methods. For instance, if two models $\mathcal{K}$ and $\mathcal{L}$ are given, then one can analyze their relationship by studying various functors originating in $\operatorname{hom}_{\text {Cat }}(\mathcal{K}, \mathcal{L})$. To quantify the modeling error (see $\S 4.2 .3$ ), one can employ a problem-dependent degree of forgetfulness (DoFF). E.g., if two or three forgetful functors from TVect ${ }_{k}$ towards Set (cf. the diagram in (4.12)) are provided, then one can assign the value two or three to the degree of forgetfulness from TVect ${ }_{k}$ to Set. Thus, in this case, the modeling error associated with TVect ${ }_{k}$ would be two or three. It can be stipulated that a larger number reflects a greater knowledge of the model; and therefore a larger number indicates a more trustworthy model.

By using a new functor G , one can construct a new low-fidelity model $\mathcal{M}_{3}$ from the low-fidelity model $\mathcal{M}_{2}$. Likewise, this construction can be interpreted as an assignment of the new low-fidelity model $\mathcal{M}_{3}$ to the high-fidelity model $\mathcal{M}_{1}$ by using
a composite functor $\mathrm{G} \circ \mathrm{F}$.


One can also ascribe directly a new low-fidelity model $\mathcal{M}_{4}$ to the high-fidelity model $\mathcal{M}_{1}$ by using a functor H . Note that if, in addition, a functor $\tilde{\mathrm{G}}$ is provided such that $\tilde{\mathrm{G}} \not \equiv \mathrm{G}$ and $\tilde{\mathrm{G}}: \neq \mathrm{G}$, and if it is set that $\mathrm{H}:=\tilde{\mathrm{G}} \circ \mathrm{F}$, then we have a similar situation as in the previous diagram in (4.28) - mind that it is not the same situation, though. Thus, the low-fidelity models $\mathcal{M}_{4}$ and $\mathcal{M}_{3}$ are still distinguishable regarded as models.

To be seen as indistinguishable, it demands testing for equivalence $\mathcal{M}_{4} \simeq \mathcal{M}_{3}$ (recall Definition 4.2.4) by furnishing adequate functors $\mathrm{U}: \mathcal{M}_{4} \rightleftarrows \mathcal{M}_{3}: \mathrm{V}$. Generally, this equivalence test is a good initial tool to use if an arbitrarily new model is stated.

Further practical considerations require, for instance, to cover the issue of how to check for equivalence by normalization. Furthermore, admittedly, it could be that sometimes an even weaker form of equivalence is more appropriate (recall § 4.2.2).


It is perceivable that one can arbitrarily scale such diagrams in (4.30) - while maintaining the interpretability which facilitates a correct reasoning about the corresponding algorithms and programs. Because note that the inherent complexity at the level of generalized functions propagates through the level of algorithms to the level of programs (see Figure 1.4); but it is increasing at each level since each level adds to it some level-specific complexity such as programming language-dependent features.

Observe that, as instances for the presented diagrams such as in (4.27), one can
apply the category Set with appropriate additional properties and one can make use of the corresponding Set-valued functors, that is, these kinds of functors whose codomain is constituted by the category Set. However, the level of abstraction of the CT approach also allows us to apply, for example, a category Graph of graphs and graph homomorphisms. For more details on such graphs-related categories, see, e.g., the references regarding the diagrams in (4.7).

Such a category Graph is a reasonable choice in practical terms since a programming language's compiler operates with abstract syntax trees. Moreover, regarding the intrusiveness (see § 4.1.1) property assigned to the low-fidelity models in a traditional setting, this new context suggests to assign this property rather to the involved functors.

If we choose the second model representation in (4.26) as our high-fidelity model $\mathcal{M}_{1}$, then one can proceed analogously to the other choice. For the sake of clarity, let us omit the parentheses.


These previous considerations provide a useful supplement to the concept of multifidelity model management (see § 4.1.2). Using the CT approach, one can interpret this concept's basic building block as having "a finite category $\mathcal{C}_{m m m}$ with morphisms as models". Again, one can sketch two cases depending on the available information. To illustrate the two cases, let us choose one high-fidelity model (index 0 ) and two low-fidelity models (index 1 and index 2 , respectively).

$X_{0} \xrightarrow{J_{f_{0}}} Z_{0}$
$X_{1} \xrightarrow{f_{1}} Y_{1} \xrightarrow{J_{1}} Z_{1}$

$$
\begin{equation*}
X_{1} \xrightarrow{I_{f_{1}}} Z_{1} \tag{4.32}
\end{equation*}
$$

$$
X_{2} \xrightarrow{I_{f_{2}}} Z_{2}
$$

This concept requires mostly that there are isomorphisms (recall the diagram in (4.2)) between the input objects ( $X_{0} \cong_{\mathcal{C}} X_{1}, X_{0} \cong_{\mathcal{C}} X_{2}$, and $X_{1} \cong_{\mathcal{C}} X_{2}$ ), intermediate objects ( $Y_{0} \cong_{\mathcal{C}} Y_{1}, Y_{0} \cong_{\mathcal{C}} Y_{2}$, and $Y_{1} \cong_{\mathcal{C}} Y_{2}$ ), and output objects ( $Z_{0} \cong_{\mathcal{C}} Z_{1}, Z_{0} \cong_{\mathcal{C}} Z_{2}$, and $Z_{1} \cong \mathcal{C} Z_{2}$ ) of the models. One can substitute one object with the other, if they are isomorphic.


The notion of sameness of models is encoded by the equality of a parallel pair of morphisms (such as in (4.6)), for instance, $J_{f_{0}}, J_{f_{1}}: X_{0} \rightrightarrows Z_{0}, J_{f_{0}} J_{f_{2}}: X_{0} \rightrightarrows Z_{0}$.
or $J_{f_{1}} J_{f_{2}}: X_{0} \rightrightarrows Z_{0}$. For this purpose, one needs to presume the existence of the terminal object $\mathbf{1}$ within the given category.

Discern that the semantic link between the models in (4.32) is carried out by conceptions such as model evaluation costs and correlation coefficients - whereas in diagrams such as (4.30), the semantic link between the models is primarily carried out by functors.

To examine the space mapping notion (see § 4.1.2), let us also interpret its basic building blocks as having "a finite category $\mathcal{C}_{s m}$ with morphisms as models". Then, one can observe that it makes use of the intermediate objects. ${ }^{6}$
$X_{0} \xrightarrow{f_{0}} Y_{0} \xrightarrow{J_{0}} Z_{0}$


$$
\begin{equation*}
X_{1} \xrightarrow{f_{1}} Y_{2} \xrightarrow{J_{1}} Z_{1} \tag{4.34}
\end{equation*}
$$



In (4.34), there is no need to require any isomorphisms. More precisely, there is no necessity that, for any (not-ordered) $i, j:\{0,1,2\}^{2}$, the morphisms $p_{i j}: X_{i} \not \longrightarrow X_{j}: \tilde{p}_{i j}$ have to satisfy the conditions

$$
\begin{align*}
& \tilde{p}_{i j} \circ p_{i j}=\operatorname{id}_{X_{i}}  \tag{4.35a}\\
& p_{i j} \circ \tilde{p}_{i j}=\operatorname{id}_{X_{j}} \tag{4.35b}
\end{align*}
$$

or the morphisms $r_{i j}: Y_{i} \rightleftarrows Y_{j}: \tilde{r}_{i j}$ have to satisfy the conditions

$$
\begin{align*}
\tilde{r}_{i j} \circ r_{i j} & =\operatorname{id}_{Y_{i}}  \tag{4.36a}\\
r_{i j} \circ \tilde{r}_{i j}= & \operatorname{id}_{Y_{j}} \tag{4.36b}
\end{align*}
$$

or the morphisms $o_{i j}: Z_{i} \nLeftarrow Z_{j}: \tilde{o}_{i j}$ have to satisfy the conditions

$$
\begin{align*}
\tilde{o}_{i j} \circ o_{i j} & =\mathrm{id}_{Z_{i}}  \tag{4.37a}\\
o_{i j} \circ \tilde{o}_{i j} & =\mathrm{id}_{Z_{j}} . \tag{4.37b}
\end{align*}
$$

However, the corresponding morphisms in (4.34) can satisfy the conditions in (4.35), in (4.36), and in (4.37).

Similarly to (4.33), the notion of sameness of models in (4.34) is encoded by the equality of a parallel pair of morphisms with the signature $X_{0} \rightarrow Z_{0}$. Observe that various combinations of path compositions in (4.34) can exhibit the signature $X_{0} \rightarrow Z_{0}$.

To concretize the semantic link between the models in (4.34), various representations have been associated with some of the maps, e.g., affine maps with $\tilde{r}_{i j}$ (see, e.g., [56]) or argmin maps with $p_{i j}$ (see, e.g., [14], [95], or [145]).

Concerning the semantic link, bear in mind that different representations, and methods, respectively, suppose different isomorphisms regarding the objects in (4.34).

[^46]Hence, taking into account the individual isomorphisms - reflected by the conditions in (4.35), in (4.36), and in (4.37) - resembles a normalization process towards the state where there is essentially only one input object, one intermediate object, and one output object (see the diagrams in (4.33)).


The normalization process argument elucidates that the modeling decisions concerning the choice of isomorphisms can separate the concept of multifidelity model management encoded in $\mathcal{C}_{m m m}$ (see the diagrams in (4.32)) and the space mapping notion encoded in $\mathcal{C}_{s m}$ (see the diagrams in (4.34)), on the one hand.

On the other hand, the modeling decisions concerning the choice of isomorphisms reveals diagrammatically requirements for indistinguishability of the concept of multifidelity model management (see the diagrams in (4.32)) and the space mapping notion (see the diagrams in (4.34)).

In essence, the normalization process argument delivers a classification tool at the level of generalized functions (see Figure 1.4) for the concept of multifidelity model management and the space mapping notion.

An allied argument - which is algebraic-geometric in nature - utilizes functors. Recalling the commentary on the diagrams in (4.7), we possess the certainty that all functors preserve isomorphisms. Therefore, given the finite category $\mathcal{C}_{m m m}$ associated with (4.32) and the finite category $\mathcal{C}_{s m}$ associated with (4.34), one cannot provide a functor Q between these two categories, i.e., $\mathrm{Q}: \mathcal{C}_{m m m} \rightarrow \mathcal{C}_{s m}$, that maps the objects and the morphisms in the most obvious way such that it maps isomorphisms in $\mathcal{C}_{m m m}$ to non-isomorphisms in $\mathcal{C}_{s m}$.

Furthermore, a potential indistinguishability of the concept of multifidelity model management (see the diagrams in (4.32)) and the space mapping notion (see the diagrams in (4.34)) can be expressed by an equivalence of categories (recall Definition 4.2.4) $\mathcal{C}_{m m m} \simeq \mathcal{C}_{s m}$.

Hence, comparing the category $\mathcal{C}_{m m m}$ and the category $\mathcal{C}_{s m}$ provides another classification tool at the level of generalized functions (see Figure 1.4) for the concept of multifidelity model management and the space mapping notion.

Finally, recalling the diagrams in (4.22), let us consider the last part of the optimization. We continue our prior thread and ascribe a low-fidelity model optimization $\mathcal{O}_{1}$ to the high-fidelity model optimization $\mathcal{O}_{0}$ by using a functor P .

$$
\begin{equation*}
Z^{X} \xrightarrow{\operatorname{argmin}} X{\underset{\Gamma}{\tilde{P}}}_{\stackrel{\mathrm{P}}{\longrightarrow}} \mathrm{P}\left(Z^{X}\right) \xrightarrow{\mathrm{P}(\operatorname{argmin})} \mathrm{P}(X) \tag{4.39}
\end{equation*}
$$

Given the functor P : $\mathcal{O}_{0} \rightarrow \mathcal{O}_{1}$, one can interpret (recall § 4.2.1) the high-fidelity model optimization $\mathcal{O}_{0}$ within the low-fidelity model optimization $\mathcal{O}_{1}$. Thus, the

CT approach pins down the intuitive idea that there should be some structural similarity between the high-fidelity model optimization and the low-fidelity model optimization.

Moreover, the diagram in (4.39) signifies structurally the distinction between a surrogate-based optimization and a surrogate-guided optimization: In the former case, after establishing the functor P , there is no further interaction between $\mathcal{O}_{0}$ and $\mathcal{O}_{1}$. In the later case, after establishing the functor P , there is further interaction between $\mathcal{O}_{0}$ and $\mathcal{O}_{1}$.

Ideally, if we furnish adequate functors P : $\mathcal{O}_{0} \rightleftarrows \mathcal{O}_{1}$ : $\tilde{\mathrm{P}}$, then one can establish an equivalence of categories $\mathcal{O}_{0} \simeq \mathcal{O}_{1}$; hence, one can consider them as indistinguishable regarded as model optimizations.

However, if we take up the position of the finite category $\mathcal{C}_{m m m}$ associated with (4.32) and the finite category $\mathcal{C}_{s m}$ associated with (4.34), then the expressive power of the diagram in (4.39) reduces certainly. In the case of $\mathcal{C}_{m m m}$, for instance, I conclude from the diagrams in (4.33) that the high-fidelity model optimization and the lowfidelity model optimizations are all together encoded within the subsequent representation:

$$
\begin{equation*}
Z_{0}^{X_{0}} \xrightarrow{\operatorname{argmin}} X_{0} . \tag{4.40}
\end{equation*}
$$

Due to various combinations of path compositions in (4.34) that can exhibit the signature $X_{0} \rightarrow Z_{0}$, the high-fidelity model optimization and the low-fidelity model optimizations concerning $\mathcal{C}_{s m}$ can be all together encoded similarly to (4.40).

At the level of generalized function (see Figure 1.4), the encoding in (4.39) and the encoding in (4.40) do not contain any value judgment in the sense that one encoding is preferred more than the other encoding; they simply hint at possible logical mismatches between intuitive ideas and their mathematical encodings due to the choice of the modeling paradigm.

Let us continue with the encoding in (4.39). Notice well that the $X$-valued minimizer function argmin in (4.17) contains an internal representation (indicated by " $\mapsto$ ") and an external representation (indicated by " $\rightarrow$ ").

Since the CT approach exploits primarily the external representation (recollect § 4.3.1), though, we have to assume additionally a terminal object in order to encode properly the minimizer $x^{*}$ for (4.39).

Hence, let us introduce $\mathbf{1} \xrightarrow{x^{*}} X$ and $x^{*}: X^{1}$, respectively. Supposing that argmin is polymorphic, one can extend its signatures such that

$$
\begin{equation*}
\operatorname{argmin}: Z^{X} \rightarrow X^{1} \tag{4.41}
\end{equation*}
$$

Finally, let us adapt the diagram in (4.39) according to the statement in (4.41).

$$
\begin{equation*}
Z^{X} \xrightarrow{\operatorname{argmin}} X^{1} \underset{\sim}{\mathrm{P}} \stackrel{\mathrm{P}}{\longrightarrow} \mathrm{P}\left(Z^{X}\right) \xrightarrow{\mathrm{P}(\text { argmin })} \mathrm{P}\left(X^{1}\right) \tag{4.42}
\end{equation*}
$$

The adapted diagram in (4.42) formalizes suitably the idea of preserving the corresponding structure and minimizer. For the sake of brevity, I keep, however, the adapted signature implicit in the remaining exposition.

Similar to the abovementioned encounters, when we include multiple low-fidelity model optimizations, the CT ansatz unfolds its strength of formally correct bookkeeping of the involved interactions. It is set definitionally that am := argmin.

The functors $\mathrm{P}_{12}: \mathcal{O}_{1} \not \overrightarrow{\mathcal{O}_{2}}: \tilde{\mathrm{P}}_{12}$ indicate a possible test for equivalence (recall Definition 4.2.4) of the corresponding low-fidelity optimization problems.

Similarly to the insight about the scalability from the diagrams in (4.30), one can observe that one can arbitrarily scale diagrams such as in (4.43) while maintaining the interpretability (see § 4.2.1) - a vital component for the correct reasoning about associated algorithms and programs (see Figure 1.4).

The following depiction highlights the observation regarding the scalability. For the sake of clarity, let us omit the potential functors between the low-fidelity model optimizations whose existence is implicitly supposed.


I remark that, by default, a possible implementation of the discussed categories such as $\mathcal{C}_{m m m}$ that is associated with (4.32) and $\mathcal{C}_{s m}$ that is associated with (4.34) is the category Set, i.e., the category of (finite) sets and set functions (see § 4.2.2). Regarding the category Set, a singleton set can serve as a terminal object and isomorphisms are common bijective set functions.

An implementation of the discussed categories by the category Set seizes the difference between the classical set-oriented modeling paradigm "(set) functions as models" and the category-oriented modeling paradigm "categories as models" and "functors as model transformations" (see § 4.3.1). However, as shown in this section, one can discuss uniformly both modeling paradigms within the CT approach.

However, the set-oriented modeling paradigm benefits from a conglomeration of diverse insights at the level of programs and at the level of algorithms (see Figure 1.4).

Nevertheless, the category-oriented modeling paradigm flourishes at the level of generalized functions (see Figure 1.4), that is, it helps to encode more rigorously a lot of colloquial or intuitive ideas; and it provides new tools to compare and to classify methods, thus, it complements beneficially the tools from the set-oriented modeling paradigm.

A final remark regarding the two modeling paradigms is concerned with the direction of the arrows involved in the corresponding diagrams. In (4.26), the choice of the direction of the arrows is dictated by the semantics commonly provided by a general optimization problem (see § 4.3.1).

From a purely syntactical viewpoint, though, if we simply turn around the arrows (cf. Remark 4.2.2) in (4.26), then combinatorial deliberations reveal that there are four distinguishable diagrams conceivable regarding the three objects and two arrows in (4.26). The corresponding list of distinguishable diagrams reads as

$$
\begin{align*}
& X \xrightarrow{f} Y \xrightarrow{J} Z  \tag{4.45a}\\
& X \xrightarrow{f} Y \stackrel{J}{\longleftrightarrow} Z  \tag{4.45b}\\
& X \stackrel{f}{\longleftarrow} Y \stackrel{J}{\longleftarrow} Z  \tag{4.45c}\\
& X \stackrel{f}{\longleftrightarrow} Y \xrightarrow{J} Z . \tag{4.45d}
\end{align*}
$$

Notice well that the possibility in (4.45b), in (4.45c), and in (4.45d) are ruled out by semantics-based deliberations.

If we adapt (4.45) to categories $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and functors $\mathrm{F}, \mathrm{G}$, then we receive the subsequent list of distinguishable combinations:

$$
\begin{align*}
& \mathcal{A} \xrightarrow{\mathrm{F}} \mathcal{B} \xrightarrow{\mathrm{G}} \mathcal{C}  \tag{4.46a}\\
& \mathcal{A} \xrightarrow[\mathrm{~F}]{\longrightarrow} \mathcal{B} \stackrel{\mathrm{G}}{\longleftrightarrow} \mathcal{C}  \tag{4.46b}\\
& \mathcal{A} \stackrel{\mathrm{~F}}{\longleftrightarrow} \mathcal{B} \longleftrightarrow \mathrm{G}  \tag{4.46c}\\
& \mathcal{A} \stackrel{\mathrm{~F}}{\longleftrightarrow} \mathcal{B} \xrightarrow[\mathrm{G}]{\longrightarrow} \mathcal{C} \tag{4.46d}
\end{align*}
$$

Let us invoke the category-oriented modeling paradigm in a verbose mode such that one can spell out the statements in (4.46), for instance, in the following manner:

1. "model $\mathcal{C}$ follows from model $\mathcal{B}$, and model $\mathcal{B}$ follows from model $\mathcal{A}$ ",
2. "model $\mathcal{B}$ follows from model $\mathcal{A}$, and model $\mathcal{B}$ follows from model $\mathcal{C}$ ",
3. "model $\mathcal{A}$ follows from model $\mathcal{B}$, and model $\mathcal{B}$ follows from model $\mathcal{C}$ ",
4. "model $\mathcal{A}$ follows from model $\mathcal{B}$, and model $\mathcal{C}$ follows from model $\mathcal{B}$ ".

Thus, the category-oriented modeling paradigm offers implicitly a causal viewpoint in some sense. Such a viewpoint hints at potential additional facets regarding the notion of fidelity made at the beginning of the section. However, especially in the light of the research on causal modeling (see, e.g., [196]), the critical examination of these facets is left for future investigations. A popular adage in statistical parlance is: "Correlation does not imply causation." This adage could serve as a starting point for a critical examination of statements such as in (4.46).

### 4.4 Use cases of the CT toolset within the electromagnetics context

### 4.4.1 Use case \#1: Simplified-physics low-fidelity models

Recalling §3.1.3, one can observe that there is a difficulty regarding the formalization of all intuitive ideas w.r.t. figures such as the Figure 3.8. To put it in other words: To my best knowledge, there is a lack of a comprehensive theory to express formally all
conceivable relationships between different problems associated with a high-fidelity model and corresponding low-fidelity models.

In (Diagrams of Fig. 3.8), some possible diagrams are abstractly presented that can be associated with the Figure 3.8. In (3.102), some statements concerning (Diagrams of Fig. 3.8) are formulated in equational form. Mind that these statements and diagrams can be understood a bit more rigorously by applying the category theory toolset (see § 4.2).

From the viewpoint of the category theory toolset, one can make different modeling decisions in (Diagrams of Fig. 3.8). I illustrate some modeling decisions by the diagrams in (4.47).

(4.47)

Observe that, in (4.47), a modeling decision is exhibited that takes into account the different algebraic characters of, e.g., $\mathcal{T}_{h_{1}}^{3 D}, \mathcal{T}_{h_{1}}^{2 D}$, and $R_{1}$, by lifting them up to the level of categories. More precisely, entities such as, e.g., $\mathcal{T}_{h_{1}}^{3 D}, \mathcal{T}_{h_{1}}^{2 D}$, and $R_{1}$ are conceived as categories and their interaction is mediated by functors such as, e.g., the functor $\mathrm{I}_{1}: \mathcal{T}_{h_{1}}^{3 D} \rightarrow \mathcal{T}_{h_{1}}^{2 D}$ and the functor $\mathrm{I}_{2}: \mathcal{T}_{h_{1}}^{2 D} \rightarrow R_{1}$.

These categories can be understood as specifications (or interfaces) (recall § 4.3) of the corresponding numerical entities (recall § 3.1.3). However, I do not dwell on this modeling decision.

Regarding the other modeling decision in (4.47), it is supposed that, e.g., three finite categories $\mathcal{A}, \mathcal{B}, \mathcal{C}$ exist (similarly to (4.7)). In addition, it is presupposed that the category $\mathcal{A}$ (recall Definition 4.2.1) is constituted by the six morphisms $f_{1}, f_{2}, f_{3}$, $f_{4}, f_{5}$, and $f_{6}$ where

$$
\begin{array}{lll}
f_{1}: \operatorname{hom}_{\mathcal{A}}\left(\mathcal{T}_{h_{1}}^{3 D}, \mathcal{T}_{h_{2}}^{3 D}\right) & f_{2}: \operatorname{hom}_{\mathcal{A}}\left(\mathcal{T}_{h_{2}}^{3 D},{ }^{1} \Delta_{A x=b}^{3 D}\right) & f_{3}: \operatorname{hom}_{\mathcal{A}}\left({ }^{1} \Delta_{A x=b}^{3 D}{ }^{2} \Delta_{A x=b}^{3 D}\right) \\
f_{4}: \operatorname{hom}_{\mathcal{A}}\left(\mathcal{T}_{h_{1}}^{3 D},{ }^{2} \Delta_{A x=b}^{3 D}\right) & f_{5}: \operatorname{hom}_{\mathcal{A}}\left(\mathcal{T}_{h_{1}}^{3 D},{ }^{1} \Delta_{A x=b}^{3 D}\right) & f_{6}: \operatorname{hom}_{\mathcal{A}}\left(\mathcal{T}_{h_{2}}^{3 D},{ }^{2} \Delta_{A x=b}^{3 D}\right) . \tag{4.48b}
\end{array}
$$

Analogously to (4.48), one can constitute the category $\mathcal{B}$, and the category $\mathcal{C}$. Thus, it is supposed that the category $\mathcal{B}$ is constituted by the six morphisms $g_{1}, g_{2}$,
$g_{3}, g_{4}, g_{5}$, and $g_{6}$ where

$$
\begin{array}{lll}
g_{1}: \operatorname{hom}_{\mathcal{B}}\left(\mathcal{T}_{h_{1}}^{2 D}, \mathcal{T}_{h_{2}}^{2 D}\right) & g_{2}: \operatorname{hom}_{\mathcal{B}}\left(\mathcal{T}_{h_{2}}^{2 D},{ }^{1} \Delta_{A x=b}^{2 D}\right) & g_{3}: \operatorname{hom}_{\mathcal{B}}\left({ }^{1} \Delta_{A x=b}^{2 D}{ }^{2} \Delta_{A x=b}^{2 D}\right) \\
g_{4}: \operatorname{hom}_{\mathcal{B}}\left(\mathcal{T}_{h_{1}}^{3 D},{ }^{2} \Delta_{A x=b}^{2 D}\right) & g_{5}: \operatorname{hom}_{\mathcal{B}}\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{1} \Delta_{A x=b}^{2 D}\right) & g_{6}: \operatorname{hom}_{\mathcal{B}}\left(\mathcal{T}_{h_{2}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right) . \tag{4.49b}
\end{array}
$$

Finally, it is presupposed that the category $\mathcal{C}$ is constituted by the six morphisms $h_{1}$, $h_{2}, h_{3}, h_{4}, h_{5}$, and $h_{6}$ where

$$
\begin{array}{lll}
h_{1}: \operatorname{hom}_{\mathcal{C}}\left(R_{1}, R_{2}\right) & h_{2}: \operatorname{hom}_{\mathcal{C}}\left(R_{2}, R_{3}\right) & h_{3}: \operatorname{hom}_{\mathcal{C}}\left(R_{3}, R_{4}\right) \\
h_{4}: \operatorname{hom}_{\mathcal{C}}\left(R_{1}, R_{4}\right) & h_{5}: \operatorname{hom}_{\mathcal{C}}\left(R_{1}, R_{3}\right) & h_{6}: \operatorname{hom}_{\mathcal{C}}\left(R_{2}, R_{4}\right) . \tag{4.50b}
\end{array}
$$

For the sake of conciseness, let us assume implicitly that, regarding $\mathcal{A}, \mathcal{B}, \mathcal{C}$, all category laws are satisfied.

Notice that given a morphism $f_{3} \circ f_{2} \circ f_{1}$ as a decomposition of the morphism $f_{4}$ in (4.48b) where

$$
\begin{equation*}
f_{3} \circ f_{2} \circ f_{1}: \operatorname{hom}_{\mathcal{A}}\left(\mathcal{T}_{h_{1}}^{3 D},{ }^{2} \Delta_{A x=b}^{3 D}\right) \tag{4.51}
\end{equation*}
$$

one can conceive the morphism $f_{3} \circ f_{2} \circ f_{1}$, for instance, as a representation of an operation that encodes the change from a fine-grid discretization regarding a threedimensional space and a high threshold for a termination criterion of an iterative solver to a coarse-grid discretization regarding a three-dimensional space and a low threshold for a termination criterion of an iterative solver (cf. § 3.1.3).

Hence, given a morphism $g_{3} \circ g_{2} \circ g_{1}$ as a decomposition of the morphism $g_{4}$ in (4.49b) where

$$
\begin{equation*}
g_{3} \circ h_{2} \circ g_{1}: \operatorname{hom}_{\mathcal{B}}\left(\mathcal{T}_{g_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right), \tag{4.52}
\end{equation*}
$$

one can conceive the morphism $g_{3} \circ g_{2} \circ g_{1}$ as an analogue of the morphism $f_{3} \circ f_{2} \circ f_{1}$ in (4.51) with respect to a two-dimensional space.

Given a morphism $h_{3} \circ h_{2} \circ h_{1}$ as a decomposition of the morphism $h_{4}$ in (4.50b) where

$$
\begin{equation*}
h_{3} \circ h_{2} \circ h_{1}: \operatorname{hom}_{C}\left(R_{1}, R_{4}\right), \tag{4.53}
\end{equation*}
$$

one can conceive the morphism $h_{3} \circ h_{2} \circ h_{1}$, for instance, as a representation of an operation that encodes the change from a multivariate rational polynomial function of type $(m, 3)$ to a multivariate rational polynomial function of type $(m, 0)$ and leading coefficient of one (cf. § 3.1.3).

Additionally to the categories $\mathcal{A}, \mathcal{B}, \mathcal{C}$, it is supposed that the functors $\mathrm{F}_{1}: \mathcal{A} \rightarrow \mathcal{B}$, $\mathrm{F}_{2}: \mathcal{B} \rightarrow \mathcal{C}$, the functors $\mathrm{G}_{1}: \mathcal{A} \rightarrow \mathcal{B}, \mathrm{G}_{2}: \mathcal{B} \rightarrow \mathcal{C}$, and the functors $\mathrm{H}_{1}: \mathcal{A} \rightarrow \mathcal{B}$, $\mathrm{H}_{2}: \mathcal{B} \rightarrow \mathcal{C}$ exist.

Hence, one can reformulate and extend the statements in (3.102) by means of the functor $\mathrm{F}_{1}$ and the functor $\mathrm{F}_{2}$ (recall Definition 4.2.2), and with $\mathrm{F}_{2} \circ \mathrm{~F}_{1}: \mathcal{A} \rightarrow \mathcal{C}$, i.e.,

$$
\begin{array}{rlrl}
\mathrm{F}_{1}\left(\mathcal{T}_{h_{1}}^{3 D}\right):=\mathcal{T}_{h_{1}}^{2 D}\left(\mathcal{T}_{h_{2}}^{3 D}\right) & :=\mathcal{T}_{h_{2}}^{2 D} & \mathrm{~F}_{1}\left(f_{1}\right) & :=g_{1} \\
\left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(\mathcal{T}_{h_{1}}^{3 D}\right):=R_{1} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(\mathcal{T}_{h_{2}}^{3 D}\right) & :=R_{2} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{1}\right) \\
\mathrm{F}_{1}\left({ }^{1} \Delta_{A x=b}^{3 D}\right) & :=h_{1} \\
\left(\mathrm{~F}_{2} \circ \Delta_{A x=b}^{2 D}\right) & \mathrm{F}_{1}\left({ }^{1} \Delta_{A x}^{3 D}\right) & :=g_{2} \\
\mathrm{~F}_{1}\left({ }^{2} \Delta_{A x=b}^{3 D}\right) & :=R_{3} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{2}\right) & :=h_{2} \\
{ }^{3 D}{ }_{A x=b}^{3 D} & \mathrm{~F}_{1}\left(f_{3}\right) & :=g_{3} \\
\left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left({ }^{2} \Delta_{A x=b}^{3 D}\right) & :=R_{4} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{3}\right) & :=h_{3}
\end{array}
$$

$$
\begin{array}{rrrr}
\mathrm{F}_{1}\left(f_{4}\right):=g_{4} & \mathrm{~F}_{1}\left(f_{5}\right):=g_{5} & \mathrm{~F}_{1}\left(f_{6}\right):=g_{6} \\
\left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{4}\right):=h_{4} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{5}\right):=h_{5} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{6}\right):=h_{6} . \tag{4.54h}
\end{array}
$$

For the sake of conciseness, let us assume implicitly that, regarding $\mathrm{F}_{1}, \mathrm{~F}_{2}$, and $\mathrm{F}_{2} \circ \mathrm{~F}_{1}$, all functor laws are satisfied.

If we suppose that, similarly to (4.9), the functor $F_{1}$ and the functor $F_{2}$ are forgetful functors, then one can receive a precise encoding for the intuitive idea of considering the problems associated with the low-fidelity models as forgetful interpretations of the problem associated with the high-fidelity model (cf. the commentary on (3.102)).

Moreover, recalling § 4.3.2, one can invoke the heuristics-driven notion of a pro-blem-dependent degree of forgetfulness (DoFF) in order to quantify the modeling error. Bear in mind that my proposed notion of a DoFF is not cast in stone. The DoFF serves primarily as an auxiliary means regarding the attempt to grasp more formally the fidelity notion attached to models. I argue that the DoFF as an auxiliary means is especially useful when using many different models.

If we fix the category $\mathcal{C}$ as the model with the lowest fidelity, then one can associate the category $\mathcal{B}$ with the DoFF of 1, i.e., using one forgetful functor $F_{2}$ such that $\operatorname{cod}\left(\mathrm{F}_{2}\right) \equiv \mathcal{C}$, and one can associate the category $\mathcal{A}$ with the DoFF of 2, i.e., using two forgetful functors $\mathrm{F}_{1}$ and $\mathrm{F}_{2}$ such that $\operatorname{cod}\left(\mathrm{F}_{2} \circ \mathrm{~F}_{1}\right) \equiv \mathcal{C}$.

If we invoke the functor $G_{1}$ and the functor $G_{2}$ and the functor $G_{2} \circ G_{1}: \mathcal{A} \rightarrow \mathcal{C}$, then one can encode a different interpretation (see § 4.2.1) than the functor $F_{1}$ and the functor $\mathrm{F}_{2}$. For instance, the functor $\mathrm{G}_{1}$ can encode a "partial shift" w.r.t. the functor $F_{1}$ and the functor $G_{2}$ can encode a similar behavior w.r.t. the functor $F_{2}$ in the sense that the assignments in (4.54) can be adapted as

$$
\begin{align*}
& \mathrm{G}_{1}\left(\mathcal{T}_{h_{1}}^{3 D}\right):=\mathcal{T}_{h_{2}}^{2 D} \tag{4.55a}
\end{align*}
$$

$$
\begin{align*}
& \left(\mathrm{G}_{2} \circ \mathrm{G}_{1}\right)\left(\mathcal{T}_{h_{1}}^{3 \mathrm{D}}\right):=R_{1} \tag{4.55b}
\end{align*}
$$

$$
\begin{align*}
& \mathrm{F}_{1}\left(f_{4}\right):=g_{4}  \tag{4.55f}\\
& \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{4}\right):=h_{4} \tag{4.55g}
\end{align*}
$$

If we replace the assignments in (4.55e) by the assignments

$$
\begin{equation*}
\mathrm{G}_{1}\left({ }^{2} \Delta_{A x=b}^{3 D}\right):=\mathcal{T}_{h_{1}}^{2 D} \quad \mathrm{G}_{1}\left(f_{3}\right):=g_{1} \tag{4.56}
\end{equation*}
$$

then one would violate the functor laws in the sense that

$$
\begin{equation*}
\mathbf{G}_{1}\left(f_{3} \circ f_{2} \circ f_{1}\right) \equiv g_{1} \circ g_{3} \circ g_{2}, \tag{4.57}
\end{equation*}
$$

where $g_{1} \circ g_{3} \circ g_{2}$ is not a legitimate composite morphism within $\mathcal{B}$. From an appli-cation-driven viewpoint, one can conceive this violation, e.g., as a restriction w.r.t. the construction of an operation within another context that is grounded in, e.g., the representation of the morphism $f_{3} \circ f_{2} \circ f_{1}$ in (4.51).

If we invoke the functor $\mathrm{H}_{1}$ and the functor $\mathrm{H}_{2}$, and the functor $\mathrm{H}_{2} \circ \mathrm{H}_{1}: \mathcal{A} \rightarrow \mathcal{C}$,
then one can encode a third interpretation. For instance, the functor $\mathrm{H}_{1}$ can encode a similar behavior w.r.t. the functor $\mathrm{F}_{1}$ and the functor $\mathrm{H}_{2}$ can encode a "partial shift" w.r.t. the functor $\mathrm{F}_{2}$ in the sense that the assignments in (4.54) can be adapted as

$$
\begin{array}{rlrl}
\mathrm{F}_{1}\left(\mathcal{T}_{h_{1}}^{3 D}\right):=\mathcal{T}_{h_{1}}^{2 D} & \mathrm{~F}_{1}\left(\mathcal{T}_{h_{2}}^{3 D}\right) & :=\mathcal{T}_{h_{2}}^{2 D} & \mathrm{~F}_{1}\left(f_{1}\right):=g_{1} \\
\left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(\mathcal{T}_{h_{1}}^{3 D}\right):=R_{2} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(\mathcal{T}_{h_{2}}^{3 D}\right) & :=R_{3} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{1}\right):=h_{2} \\
\mathrm{~F}_{1}\left(\Delta_{A x}^{1} \Delta_{A x b}^{3 D}\right):={ }^{1} \Delta_{A x=b}^{2 D} & \mathrm{~F}_{1}\left(f_{2}\right):=g_{2} \\
\left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left({ }^{1} \Delta_{A x=b}^{3 D}\right):=R_{4} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{2}\right):=h_{3} \\
\mathrm{~F}_{1}\left({ }^{2} \Delta_{A x=b}^{3 D}\right) & :={ }^{2} \Delta_{A x=b}^{3 D} & \mathrm{~F}_{1}\left(f_{3}\right):=g_{3} \\
\left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left({ }^{2} \Delta_{A x=b}^{3 D}\right): & :=R_{1} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{3}\right):=h_{2} \\
\mathrm{~F}_{1}\left(f_{4}\right):=g_{4} & \left.\mathrm{~F}_{5}\right):=g_{5} & \mathrm{~F}_{1}\left(f_{6}\right):=g_{6} \\
\left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{4}\right):=h_{4} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{5}\right): & : h_{5} & \left(\mathrm{~F}_{2} \circ \mathrm{~F}_{1}\right)\left(f_{6}\right):=h_{6} .
\end{array}
$$

For the sake of brevity, let us omit the consideration of natural transformations (recall Definition 4.2.3) regarding the functors $F_{1}, F_{2}$, the functors $G_{1}, G_{2}$, and the functors $\mathrm{H}_{1}, \mathrm{H}_{2}$, that is, let us omit the comparison of the corresponding constructed paths (recall Remark 4.2.7).

However, mind that, if we suppose that the functors in (4.54) in (4.55), and in (4.58) are forgetful functors, then one can at least exclude that the categories $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ satisfy a test for equivalence (recall Definition 4.2.4).

### 4.4.2 Use case \#2: Coordinate transformations ${ }^{7}$

The second use case of the CT approach within the electromagnetics context is related to coordinate transformations. I construe the term "coordinate transformations" as an umbrella term for different applications within the electromagnetics context that are associated to coordinates in a narrower sense (e.g., w.r.t. a physical space) or to coordinates in a broader sense (e.g., w.r.t. an abstract vector space).

An application that is associated to coordinates in a narrower sense is depicted in Figure 4.1 which is inspired by [138]. In Figure 4.1, a three-dimensional helical coil of 5 turns (cf. (i) in Figure 1.3) is represented by two different coordinate systems. One coordinate system is constituted by xyz-coordinates and another coordinate system is constituted by uvw-coordinates.

In Figure 4.1, it is supposed that an initial coordinate system is provided by a user. Furthermore, I claim that the helical coil in xyz-coordinates possesses this kind of shape that is intuitively expected from common experiences in physics.

However, the helical coil in uvw-coordinates, i.e., a three-dimensional solid identified as a cylinder, possesses this kind of curvilinear geometric shape that is rather counterintuitive compared to common experiences in physics.

I do not dwell on the involved subtleties regarding such coordinate transformations such as, e.g., the proper transformation of the corresponding field quantities w.r.t. the well-posed boundary value problems (recall (2.16)). For more details concerning these subtleties, I refer to, e.g., [174] or [138].

[^47]

Figure 4.1: A schematic depiction of a three-dimensional helical coil of 5 turns in xyz-coordinates (cf. (i) in Figure 1.3) and in uvwcoordinates (inspired by [138]). For more elaborated considerations of such coordinate transformations within the electromagnetics context, I refer to, e.g., [174] or [138].

For the considerations in the present work, though, the significant aspect is that the evaluated quantities of interest (recall § 2.2.1) in both coordinate systems in Figure 4.1 should be identical. This aspect reflects the premise that (derived) measurable physical entities such as, e.g., the magnetic energy or the power loss, should be coordinate-invariant. More precisely, the (derived) measurable physical entities, and therefore the evaluated quantities of interest, should be independent of the choice of coordinates.

An application that is associated to coordinates in a broader sense is encoded by the subsequent matrix equation, that is, given a matrix $S \in \mathbb{R}^{4 \times 4}$, then

$$
\begin{equation*}
\forall S_{\mathrm{MM}} \in \mathbb{R}^{4 \times 4} \cdot \exists M \in \mathbb{R}^{4 \times 4} \cdot S_{\mathrm{MM}}=M S M^{-1} . \tag{4.59}
\end{equation*}
$$

If we assume an endomorphism $s: V \rightarrow V$ within the category Vect ${ }_{k}^{\text {basis }}$ (recall $\S 4.2 .2$ ) regarding an ordered basis $V_{b_{1}}$, then one can associate the endomorphism $s$ regarding an ordered basis $V_{b_{1}}$ with the matrix $S$ in (4.59) (cf. (4.11)).

Furthermore, one can associate the endomorphism $s$ regarding another ordered basis $V_{b_{2}}$ with the matrix $S_{\mathrm{MM}}$ such that one can conceive the matrix $M$ as the change-of-basis matrix from the ordered basis $V_{b_{1}}$ to another ordered basis $V_{b_{2}}$.

Hence, given $s$ regarding $V_{b_{2}}$ and given an input element $a: V$ regarding $V_{b_{2}}$, then we receive an output element $b: V$ regarding $V_{b_{2}}$ by setting

$$
\begin{equation*}
b:=s \circ a . \tag{4.60}
\end{equation*}
$$

If we instantiate the assignment in (4.60) by the matrix $S_{\mathrm{MM}}$, the column vector $a \in \mathbb{R}^{4 \times 1}$, and the column vector $b \in \mathbb{R}^{4 \times 1}$, then the statement in (4.60) can be written as

$$
\begin{align*}
& b:=S_{\mathrm{MM}} a  \tag{4.61a}\\
& b:=\operatorname{MSM}^{-1} a, \tag{4.61b}
\end{align*}
$$

where the term $M^{-1} a$ refers to the representation of $a: V$ regarding $V_{b_{1}}$ and the term $S M^{-1} a$ refers to the representation of $b: V$ regarding $V_{b_{1}}$.

A semantics for the statements in (4.61) within the electromagnetics context is provided by, for instance, the notion of 4-port S-parameters and the notion of 4port mixed-mode S-parameters (see, e.g., [29]) within the context of electromagnetic
compatibility (recall, e.g., the EMC filter in Figure 1.1).
Thus, in (4.61), the matrix $S$ corresponds to the 4-port S-parameters and the matrix $S_{\mathrm{MM}}$ corresponds to the 4-port mixed-mode S-parameters that incorporates the idea of differential-mode and common-mode scattering parameters. The entries of the change-of-basis matrix $M$ depend on the ordering of the entries of the matrix $S_{\mathrm{MM}}$. Let us omit an in-depth elaboration on these two kinds of S-parameters. For more details, I refer to, e.g., [29] and references therein.

Given the S-parameters semantics, the key insight from (4.61) is that, by providing the entries of the matrix $S$ which are (derived) measurable physical entities, one can determine the matrix $S_{\mathrm{MM}}$.

In the considerations of the present work, let us focus on how to embed applications such as, e.g., the application in Figure 4.1 or the application in (4.59) into the body of ideas of surrogate optimization.

Recalling § 3.1, one can intuitively state that, regarding Figure 4.1, if we associate a kriging low-fidelity model with a high-fidelity model that is linked to the problem corresponding to the helical coil in xyz-coordinates and if we associate a kriging low-fidelity model with a high-fidelity model that is linked to the helical coil in uvwcoordinates, then these two kriging low-fidelity models should behave equally. Let us refer to this situation as formalization issue $A$ (FI-A).

Regarding (4.59), given the $S_{i j}$-parameter w.r.t. the matrix $S:=\left[s_{i, j}\right] \in \mathbb{R}^{4 \times 4}$ which one can define as

$$
\begin{equation*}
S_{i j}:=\left[s_{i, j}\right] \tag{4.62}
\end{equation*}
$$

then, for instance, one can associate with every $S_{i j}$-Parameter a high-fidelity model and a low-fidelity model similarly to the multivariate vector-valued use case in (3.131). ${ }^{8}$

Therefore, we receive a matrix $S_{K} \in\left(Y_{0}^{X_{0}}\right)^{4 \times 4}$ that corresponds to the high-fidelity models and we receive a matrix $S_{\tilde{\mathrm{K}}} \in\left(Y_{1}^{X_{1}}\right)^{4 \times 4}$ that corresponds to the low-fidelity models. ${ }^{9}$

If we employ the matrix $S_{\mathrm{K}}$ and the matrix $S_{\tilde{\mathrm{K}}}$ in (4.59), we obtain the matrix $S_{\mathrm{MM}_{\mathrm{K}}}$ and the matrix $S_{\mathrm{MM}_{\tilde{k}}}$, respectively. Mind that the change-of-basis matrix $M$ operates componentwise w.r.t. $Y_{0} \subseteq \mathbb{R}^{m_{w v}}$ on the evaluated matrix $S_{\mathrm{K}}(x):\left(Y_{0}\right)^{4 \times 4}$ and it operates componentwise w.r.t. $Y_{1} \subseteq \mathbb{R}^{m_{w w}}$ on the evaluated matrix $S_{\tilde{\mathrm{K}}}(\tilde{x}):\left(Y_{1}\right)^{4 \times 4}$, respectively.

One can also construct a matrix $S_{\mathrm{MM}_{K, \tilde{K}}}$ whose entries are constituted by lowfidelity models regarding the entries of $S_{\mathrm{MM}_{k}}$.

Observe that, however, one cannot necessarily expect that the matrix $S_{\mathrm{MM}_{\mathrm{K}, \overline{\mathrm{K}}}}$ and the matrix $S_{\mathrm{MM}_{\overline{\mathrm{K}}}}$ behave equally. Let us refer to this situation as formalization issue $B$ (FI-B).

Even if we limit our considerations to the surrogate modeling \& simulation subpart of surrogate optimization, one can already recognize that a substantial amount of bookkeeping is needed to examine situations regarding coordinate transformations.

[^48]Formalizing adequately situations regarding coordinate transformations in surrogate optimization can be beneficial from a theory-driven viewpoint as well as from an application-driven viewpoint. From a theory-driven viewpoint such an adequate formalizing hints at the possibilities of, for lack of a better word, a coordinateinvariant surrogate modeling $\mathcal{E}$ simulation. From an application-driven viewpoint, such an adequate formalizing illuminates the path for the question of, e.g., to what extent post-processing entities of a high-fidelity model based on a numerical simulation (cf. § 2.2) can be captured by post-processing entities of a corresponding low-fidelity model.

I argue that the CT approach can at least assist in formalizing such intricate situations regarding coordinate transformations in surrogate optimization that, to my best knowledge, are not exhaustively studied.

By recalling the diagrams from (4.27) to (4.31), one can abstract a common pattern from the above-mentioned situations regarding coordinate transformations in the surrogate modeling \& simulation sub-notion of surrogate optimization. Hence, I propose the subsequent diagram as a means of formalizing this common pattern.

Observe that, in (4.63), there are five categories (recall Definition 4.2.1) $\mathcal{M}_{1}, \mathcal{M}_{2}$, $\mathcal{M}_{3}, \mathcal{M}_{4}, \mathcal{M}_{5}$ accompanied by the six functors (recall Definition 4.2.2) F: $\mathcal{M}_{1} \rightarrow \mathcal{M}_{2}$, $\mathrm{T}: \mathcal{M}_{1} \rightarrow \mathcal{M}_{3}, \mathrm{G}: \mathcal{M}_{3} \rightarrow \mathcal{M}_{4}, \mathrm{~S}: \mathcal{M}_{2} \rightarrow \mathcal{M}_{5}, \mathrm{U}: \mathcal{M}_{4} \rightarrow \mathcal{M}_{5}$, and $\mathrm{V}: \mathcal{M}_{5} \rightarrow \mathcal{M}_{4}$. Notice that, by providing the functors $\mathrm{F}, \mathrm{T}, \mathrm{G}, \mathrm{S}, \mathrm{U}, \mathrm{V}$, the identification of the categories $\mathcal{M}_{1}, \mathcal{M}_{2}, \mathcal{M}_{3}, \mathcal{M}_{4}, \mathcal{M}_{5}$ with their diagrammatic representation in (4.63) should be clear.

Let us associate $\mathcal{M}_{1}$ with a high-fidelity model, and link $\mathcal{M}_{3}$ with the corresponding coordinate-transformed high-fidelity model, and relate $\mathcal{M}_{4}$ to the lowfidelity model of the coordinate-transformed high-fidelity model. Furthermore, let us associate $\mathcal{M}_{2}$ with the corresponding low-fidelity model, and relate $\mathcal{M}_{5}$ to the corresponding coordinate-transformed low-fidelity model.

The inverse functor (recall Remark 4.2.10) $\mathrm{T}^{-1}: \mathcal{M}_{3} \rightarrow \mathcal{M}_{1}$ and the inverse functor $\mathrm{S}^{-1}: \mathcal{M}_{5} \rightarrow \mathcal{M}_{2}$ indicate a coordinate transformation at the level of the high-fidelity model and a coordinate transformation at the level of the low-fidelity model, respectively. Notice that the existence of the inverse functors implies the equivalence of categories (recall Definition 4.2.4) $\mathcal{M}_{1} \simeq \mathcal{M}_{3}$ and $\mathcal{M}_{2} \simeq \mathcal{M}_{5}$, respectively. From an application-driven viewpoint, a potential reading of these equivalences is that, roughly speaking, the high-fidelity model $\mathcal{M}_{1}$ and the corresponding coordinate-transformed high-fidelity model $\mathcal{M}_{3}$ are essentially the same as well as the low-fidelity model $\mathcal{M}_{2}$ and the corresponding coordinate-transformed lowfidelity model $\mathcal{M}_{5}$ are essentially the same.

The formalization issue FI-A can be expressed within the diagram in (4.63) by
setting $\mathcal{M}_{2} \equiv \mathcal{M}_{5}$, and $\mathrm{S}:=\mathrm{id}_{\mathcal{M}_{2}}$, and by demanding that the functor V is an inverse functor to the functor U , i.e., $\mathrm{V} \equiv \mathrm{U}^{-1}$, such that there is an equivalence of categories $\mathcal{M}_{4} \simeq \mathcal{M}_{2}$. Moreover, by assuming $\mathcal{M}_{1} \simeq \mathcal{M}_{3}$, one can ultimately regard $\mathrm{F} \equiv \mathrm{G}$. Or to put it in other words: The low-fidelity model $\mathcal{M}_{2}$ and the low-fidelity model of the coordinate-transformed high-fidelity model $\mathcal{M}_{4}$ are essentially the same.

Using the diagram in (4.63), the essence of the formalization issue FI-B is that there is not necessarily a functor V such that $\mathrm{V} \equiv \mathrm{U}^{-1}$, and therefore $\mathcal{M}_{4} \neq \mathcal{M}_{5}$. One can roughly construe this equivalence as follows: The low-fidelity model of the coordinate-transformed high-fidelity model $\mathcal{M}_{4}$ and the coordinate-transformed low-fidelity model $\mathcal{M}_{5}$ are not essentially the same.

### 4.5 Future use cases for the CT toolset

I point out that, compared to, e.g., the usage of the language of functional analysis and the language of differential geometry (recall ch. 2), the usage of the language of category theory is still in an early stage regarding surrogate optimization within the electromagnetics context.

However, by emphasizing the category theoretical language's ability as a strong notational scaffolding by diagrams of arrows, we have discussed in particular its potential usefulness concerning some formal aspects of surrogate optimization within the electromagnetics context.

Despite the fact that we have harnessed a lot of the CT toolset's rigor, there is, admittedly, more research needed to put the applications of the CT toolset on an even more rigorous footing. I remark that the language of category theory offers many more tools to be explored.

Besides the need for an even more rigorous footing of the applications of the CT toolset, the findings of the present chapter can serve as a good starting point for three interesting paths to pursue regarding future use cases for the CT toolset.

Concerning implementations of the CT toolset for surrogate optimization within the electromagnetics context, the first path points at theory-driven implementations in programming languages that follow different design principles: dynamicallychecked, statically-checked, imperative, functional, and similar. The expressiveness of the chosen programming language will determine how many work-around artifacts in the actually programs are needed and how the incorporation of, e.g., different numerical solvers are concretized.

The second path is inspired by [114] where the authors promote category theory as a multidisciplinary language in order to discuss a magneto-elasticity problem. Thus, the CT toolset could act as a guiding principle in designing surrogate-guided optimization methods for multidisciplinary design optimization (recall § 1.1). In constructing the corresponding numerical algorithms, one strength of the CT toolset could be particularly useful: the coherent change of perspective from an algebraic to a geometric presentation.

The third path is concerned with a thorough development of a modeling and reasoning environment for parallel SGO methods. Recap that the definition of a category (see Definition 4.2.1) deals solely with sequential composition. However, monoidal categories (see, e.g., [46, p. 72f]) seem promising for this path since they incorporate sequential composition and parallel composition as well. Recalling the
diagrams in (4.32), monoidal categories can assist in encoding meaningfully expressions such as ${ }^{10}$

$$
\begin{align*}
& \forall J_{0}, J_{1}, J_{2} \cdot \forall f_{0}, f_{1}, f_{2} \cdot\left(J_{0} \circ f_{0}\right) \otimes\left(J_{1} \circ f_{1}\right) \otimes\left(J_{2} \circ f_{2}\right)=\left(\left(J_{0} \otimes J_{1}\right) \circ\left(f_{0} \otimes f_{1}\right)\right) \otimes\left(J_{2} \circ f_{2}\right)  \tag{4.64a}\\
& \forall J_{0}, J_{1}, J_{2} \cdot \forall f_{0}, f_{1}, f_{2} \cdot\left(J_{0} \circ f_{0}\right) \otimes\left(J_{1} \circ f_{1}\right) \otimes\left(J_{2} \circ f_{2}\right)=\left(J_{0} \otimes J_{1} \otimes J_{2}\right) \circ\left(f_{0} \otimes f_{1} \otimes f_{2}\right) . \tag{4.64b}
\end{align*}
$$

Moreover, monoidal categories help to formalize, e.g., the subsequent diagrams.

$$
X_{0} \xrightarrow{f_{0}} Y_{0} \xrightarrow{J_{0}} Z_{0}
$$

$$
X_{0} \xrightarrow{J_{f_{0}}} Z_{0}
$$

[^49]\[

$$
\begin{align*}
& \otimes \\
& X_{2} \xrightarrow{f_{2}} Y_{2} \xrightarrow{J_{2}} Z_{2} \\
& X_{2} \xrightarrow{J_{f_{2}}} Z_{2} \\
& X_{0} \otimes X_{1} \otimes X_{2} \xrightarrow{f_{0} \otimes f_{1} \otimes f_{2}} Y_{0} \otimes Y_{1} \otimes Y_{2} \xrightarrow{J_{0} \otimes J_{1} \otimes J_{2}} Z_{0} \otimes Z_{1} \otimes Z_{2}  \tag{4.66}\\
& X_{0} \otimes X_{1} \otimes X_{2} \xrightarrow{J_{f_{0}} \otimes J_{f_{1}} \otimes J_{f_{2}}} Z_{0} \otimes Z_{1} \otimes Z_{2}  \tag{4.67}\\
& \otimes
\end{align*}
$$
\]

### 4.6 In closing

The major aims of the chapter have been (1) to furnish us with a critical examination of the algebraic tools from the language of category theory (CT) that have been anticipated in the preceding chapters; and (2) to investigate to which extent the category theoretical language is beneficial in the development of an algebraic modeling framework for applications in surrogate optimization.

In order to illustrate the principal application-oriented advantages of an algebraic modeling framework, we have recapitulated the contextual landscape of surrogate optimization in the previous chapters and we have partly enlarged this landscape by the emerging context of full automation of surrogate-guided optimization (SGO).

I have argued that to sufficiently comply with this emerging context, there is a need for dealing with issues regarding software systems, too. Currently, however, there is a gap between issues regarding software systems and the emerging context of full automation of SGO.

Observing this gap, I have proposed the use of the category theoretical language which can serve as a mediating instance between certain language-focused aspects of software systems and the mathematical modeling involved in SGO methods.

Furthermore, I have mentioned some relevant related work regarding the development of an algebraic modeling framework using the category theoretical language for applications in surrogate optimization.

Subsequently, we have elaborated on the category theory toolset used in the present work. This CT toolset, though, represents solely a subset of the large amount of tools available within category theory. By illuminating the strengths of the CT toolset as a strong notational scaffolding by diagrams of arrows, the power of the coherent change of perspective from an algebraic presentation to a geometric presentation and vice versa has been shown.

There has been an attempt to balance the need for rigor and the need for applicability as good as possible. Admittedly, more research is needed to put the applications of the CT toolset on an even more rigorous footing. However, an important concern has been to foster some intuition regarding the CT toolset. From an application-oriented viewpoint, this intuition is in particular important to promote the CT toolset's wider acceptance.

Besides the intuition, for the CT toolset's wider acceptance from an applicationoriented viewpoint, computational facets are another important point. Although, it is not the primary ambition in the present work, computational facets of the CT toolset have been signposted and the relationship to more established toolsets (recall ch. 2) has been illuminated. Finally, I have suggested a heuristics-driven notion of a problem-dependent degree of forgetfulness (DoFF) in order to quantify the so-called modeling error.

Afterwards, one way of specifying a general optimization problem by using the CT toolset has been shown. For instance, we have carved out a diagram of arrows that enables a discussion of performance-related issues regarding a general optimization problem.

In specifying SGO methods by using the CT toolset, we have examined the classical set-oriented modeling paradigm "(set) functions as models" and the categoryoriented modeling paradigm: "Categories as models" and "functors as model transformations".

By assuming an order-oriented fidelity notion regarding multi-fidelity methods, we have in particular investigated the comparability of the concept of multifidelity
model management and the space mapping notion. Thus, I have propounded some classification tools at the level of generalized functions (recall Figure 1.4) for the concept of multifidelity model management and the space mapping notion.

By exploring the direction of the arrows in the category-oriented modeling paradigm, some potential additional facets regarding the fidelity notion have been indicated.

Furthermore, we have discussed the scalability of diagrams of arrows - while maintaining the interpretability - concerning the handling of a high-fidelity optimization problem and multiple low-fidelity optimization problems.

Ultimately, we have examined two use cases of the CT toolset that are relevant to surrogate optimization for applications within the electromagnetics context: (1) simplified-physics low-fidelity models, and (2) coordinate transformations.

Regarding the first use case, it has been exemplified how the machinery of category theory can be applied to the construction process and the interaction of simpli-fied-physics low-fidelity models.

It has been shown, for instance, how certain modeling decisions can lead to restrictions w.r.t. certain constructions. Furthermore, we have invoked the heuristicsdriven notion of a problem-dependent DoFF as a useful auxiliary means to quantify simplified-physics low-fidelity models as forgetful interpretations of the highfidelity model.

Regarding the second use case, it has been exemplified how the machinery of category theory can be applied to handle certain formalization issues concerning the surrogate modeling \& simulation sub-notion of surrogate optimization.

One formalization issue is related to coordinates w.r.t. a physical space; another formalization issue is related to coordinates w.r.t. an abstract vector space. A context for the first formalization issue is provided by the coordinate transformation of a helical coil w.r.t. a well-posed boundary value problem. A context for the second formalization issue is provided by the transformation of 4-port S-parameters.

By using the CT toolset, I have proposed a diagram of arrows as a common generic interface of both formalization issues. It has been demonstrated that, by setting adequately the entities within the diagram, the essential statement of the respective formalization issue can be instantiated.

We have ended the investigation by exhibiting three potential future use cases for the CT toolset based on the findings of the present chapter: (1) implementations in programming languages, (2) SGO methods for multidisciplinary design optimization, and (3) a modeling and reasoning environment for parallel SGO methods.

## Chapter 5

## Surrogate optimization with the magnetoquasistatic model

Invoking the insights from the previous chapters, let us discuss how to apply a subset of these insights to four optimization problems relevant within an electrical engineering design context. Therefore, I deem it beneficial to utilize the image of a rapid prototyping part of a product development cycle as a means to contextualize the subsequent sections. Hence, I present a strategy of using surrogate optimization in order to support preexisting electrical engineering design workflows.

In the first section, I elaborate on two optimization problems with regard to a single inductive component, more precisely, to a solenoid with a core within the setting of a 2D-LBVP. In the preliminary considerations, I present all the basic building blocks to formulate the two optimization problems. Given an operating frequency, the first optimization problem uses the semantics of the time-averaged ohmic loss to encode the objective functional and it uses the semantics of the inductance to encode the constraints besides some box constraints. The second optimization problem extends the discussion of the first optimization problem to the case of multiple operating frequencies under test.

In the second section, let us discuss briefly one optimization problem with regard to a common-mode choke within a prototypical version of a simplistic EMC filter that is embedded in the setting of a 3D-LBVP. Furthermore, we discuss another optimization problem with regard to two common-mode chokes within the setting of a 2D-LBVP.

Concerning the first optimization problem, the aim is to minimize the magnitude of one scattering parameter for a very narrow frequency range. From a theoretical viewpoint, the notion of scattering parameters is not encoded within the magnetoquasistatic model. From an application-driven viewpoint, though, there is usually a fuzzy bounded overlapping modeling region of the magnetoquasistatic subsystem of Maxwell's equations and the complete system of Maxwell's equations to gain some knowledge about a given application for low frequency ranges such as, e.g., from 0 Hz to 200 MHz . The focus is mainly on the surrogate optimization, thus, we consider the corresponding high-fidelity model primarily as a black-box model.

Regarding the second optimization problem, I examine the optimal positioning of two common-mode chokes in order to lower their inductive coupling. I present an approximate encoding of the high-fidelity optimization problem and, for this use case, let us discuss concisely a surrogate optimization strategy as well.

### 5.1 Solenoid with a core

### 5.1.1 Preliminary consideration

The device under test is a solenoid with a core as a representative of the class of inductive components.

In order to formulate a meaningful optimization problem within the syntactical framework discussed in $\S 2.3$, let us briefly illustrate two physical behaviors of inductive components that are relevant in practical applications: (1) losses and (2) electromagnetic compatibility (EMC). These two main aspects are organized within a short list. Albeit for a comprehensive treatment of the specific sub-aspects in this short list, I refer to, e.g., [154] or [164] and references therein.

The short list reads as:
(1) losses
(1.1) losses in the winding
(1.1.1) losses due to a direct current
(1.1.2) losses due to an alternating current
(1.1.2.1) losses due to the skin effect
(1.1.2.2) losses due to the proximity effect
(1.2) losses in the core
(1.2.1) losses due to hysteresis
(1.2.2) losses due to eddy currents
(1.2.3) losses due to relaxation processes
(2) electromagnetic compatibility
(2.1) galvanic coupling or ohmic coupling
(2.2) electric coupling or capacitive coupling
(2.3) magnetic coupling or inductive coupling
(2.4) radiative coupling.

Invoking the EMC theory's so-called "Emitter-Coupler-Receiver" scaffolding, one can conceive the electromagnetic compatibility of an inductive component as the component's capability to interact in a mostly unwanted way with another electric device. In the chapter's exposition, let us be mainly concerned with the magnetic coupling. Recalling Figure 1.1, an interesting use case is the magnetic coupling of inductive components within an electromagnetic system such as, e.g., an EMC filter. Therefore, the positioning of the inductive components within such an EMC filter can influence the behavior of the filter.

Notice that I comprehend all the losses of an inductive component roughly as power losses converted into heat. With regard to the BVPs under consideration, though, it is assumed that the core is lossless which is a reasonable approximation of real inductive components if their core material is ferrite.

In order to encode the solenoid with a core within a 2D-LBVP, one can break, metaphorically speaking, the helicoidal winding with $N \in \mathbb{N}$ turns of an actual solenoid into $N$ toroids (recall Figure 1.3). Finally, one can exploit the rotational symmetry of the corresponding 3D-LBVP that is geometrically composed of $N$ toroids and a cylindrical core such that it turns the 3D-LBVP into a 2D-LBVP (cf. Figure 3.7 and Figure 3.8). Hence, for the purpose of a numerical simulation within FEMM4.2,
let us encode the solenoid with a core according to the schematic description in (i) of Figure 5.1. ${ }^{1}$


FIGURE 5.1: (i) A schematic illustration of a solenoid with core with an axially symmetric domain. (ii) A simplicial triangulation $\mathcal{T}_{h}$ of the space region $\Omega^{2 D}$ via FEMM4.2.

In (i) of Figure 5.1, the entity $d_{c} \in \mathbb{R}^{+}$refers to a fixed distance between the $N$ toroids. It is supposed that the radii of the cross-section areas of the $N$ toroids are all the same length which is encoded by the parameter $x_{1}$. The parameter $x_{2}$ denotes the radius of the $N$ toroids themselves. The height of the core $h_{\Omega_{c, 2}^{2 D}}$ is an affine function with the signature $\mathbb{R}^{+} \rightarrow \mathbb{R}^{+}$that depends on the parameter $x_{1}$, more precisely,

$$
\begin{equation*}
h_{\Omega_{c, 2}^{2 D}}=x_{1} \mapsto N\left(d_{c}+x_{1}\right)+a_{0}: \mathbb{R}^{+} \rightarrow \mathbb{R}^{+} \tag{5.1}
\end{equation*}
$$

where $a_{0} \in \mathbb{R}^{+}$is a fixed parameter to set up the core's vertical expansion below the lowest circle and above the highest circle in (ii) of Figure 5.1. Furthermore, note that the width of the core $w_{\Omega_{c, 2}^{2 D}}$ is an affine function with the signature $\mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}$that depends on the parameters $x_{1}$ and $x_{2}$, more precisely,

$$
\begin{equation*}
w_{\Omega_{c, 2}^{2 D}}=\left(x_{1}, x_{2}\right) \mapsto 2\left(x_{2}-x_{1}\right)-2 a_{1}: \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}, \tag{5.2}
\end{equation*}
$$

[^50]where $a_{1} \in \mathbb{R}^{+}$is a fixed parameter to set up a minimal distance with respect to the core's horizontal expansion towards the series of circles in (ii) of Figure 5.1. In the test cases, it is set that $N:=10, d_{c}:=0.1 \mathrm{~mm}, a_{0}:=2.0 \mathrm{~mm}$, and $a_{1}:=0.5 \mathrm{~mm}$.

With regard to the constitutive equations in (2.2), it is set that the material characteristics of the subdomain $\Omega_{n c, 1}^{2 D}$ equal to the material characteristics of air. The material properties of the subdomain $\Omega_{n c, 2}^{2 D}$ are identified with the material properties of the subdomain $\Omega_{n c, 1}^{2 D}$ - except for the relative magnetic permeability that is set to $5 \times 10^{3}$. Finally, for the subdomains from $\Omega_{c, 1}^{2 d}$ to $\Omega_{c, N}^{2 D}$, it is assumed that the material characteristics of plain copper (recall § 3.1.3).

With regard to the source term $\mathbf{J}_{\text {src }}$ in (2.16), a fixed current intensity (or in short, a fixed current) $I_{0} \in \mathbb{R}^{+}$where $I_{0}:=1 \mathrm{~A}$ is associated with the subdomains from $\Omega_{c, 1}^{2 D}$ to $\Omega_{c, N}^{2 D}$ such that, in the case of direct current and in the case of alternating current of sinusoidal waveform, the value of the root mean squared current $I_{\mathrm{rms}}$ is identical to the value of $I_{0}$, i.e.,

$$
\begin{equation*}
I_{\mathrm{rms}} \equiv I_{0} . \tag{5.3}
\end{equation*}
$$

Observe that, in the case of alternating current of sinusoidal waveform, the requirement in (5.3) implies that, due to the fact that the relationship

$$
\begin{equation*}
\forall I_{\mathrm{rms}} \in \mathbb{R}^{+} . \exists I_{\text {peak }} \in \mathbb{R}^{+} . I_{\mathrm{rms}}=\frac{I_{\text {peak }}}{\sqrt{2}} \tag{5.4}
\end{equation*}
$$

holds, the value of the peak current $I_{\text {peak }}$ has to be set equal to the value of $\sqrt{2} I_{0}$, i.e.,

$$
\begin{equation*}
I_{\text {peak }} \equiv \sqrt{2} I_{0} . \tag{5.5}
\end{equation*}
$$

Furthermore, it is demanded that the source current density's orientation is in the same direction for each subdomain $\Omega_{c, i}^{2 D}$ with $i \in\{1,2, \ldots, N\}$. Mind that, from an electrical network viewpoint on which we elaborate briefly below, one can consider the conducting subdomains as galvanically connected in series such that the following conditions hold:

$$
\begin{align*}
& \forall i \in\{1, \ldots, N\} \cdot \int_{\Omega_{c, i}^{2 d}} \operatorname{Re}\left(\mathbf{J}_{\text {cond }}\right) \cdot \mathrm{d} \mathbf{A}:= \begin{cases}I_{0}, & \text { if } \omega \equiv 2 \pi \cdot f \text { with } f:=0 \mathrm{~Hz} \\
I_{0} \sqrt{2}, & \text { if } \omega \equiv 2 \pi \cdot f \text { with } f>0 \mathrm{~Hz},\end{cases}  \tag{5.6a}\\
& \exists I=\{1, \ldots, N\} \cdot \int_{\mathcal{U}_{i \epsilon I} \Omega_{c, i}^{2 d}} \operatorname{Re}\left(\mathbf{J}_{\text {cond }}\right) \cdot \mathrm{d} \mathbf{A}:= \begin{cases}I_{0}, & \text { if } \omega \equiv 2 \pi \cdot f \text { with } f:=0 \mathrm{~Hz} \\
I_{0} \sqrt{2}, & \text { if } \omega \equiv 2 \pi \cdot f \text { with } f>0 \mathrm{~Hz},\end{cases} \tag{5.6b}
\end{align*}
$$

where the common map $\operatorname{Re}: C \rightarrow \mathbb{R}$ that maps a complex number to its real part is overloaded.

Considering the boundary conditions in (2.16), Dirichlet boundary conditions on $\partial \Omega^{2 D}$ are solely imposed. More precisely, it is prescribed that the magnetic vector potential $\mathbf{A}$ is equal to the zero vector field along $\partial \Omega^{2 D}$.

In the context of the numerical simulation of the magnetoquasistatic model (recall $\S 2.2$ ), one has to pay attention to two points regarding the boundary $\partial \Omega^{2 D}$.

First, the boundary $\partial \Omega^{2 D}$ is topologically isomorphic (or homeomorphic) to the boundary of a disk, i.e., to a one-dimensional sphere. If one examines other kinds of isomorphisms as well, then, in general, the shape of the boundary $\partial \Omega^{2 D}$ and, in particular, the distance from the subdomains $\Omega_{n c, 2}^{2 D}$ and $\Omega_{c, i}^{2 D}$ for all $i \in\{1,2, \ldots, N\}$
to the boundary $\partial \Omega^{2 D}$ may be relevant to the 2D-LBVP. However, let us invoke a pragmatic approach such that we set heuristically the distance $d_{\Omega_{n c, 2}^{2 D}, \partial \Omega^{2 D}} \in \mathbb{R}^{+}$, that is, the distance from the vertical half of the core to the end of the boundary $\partial \Omega^{2 D}$, to at least four times $h_{\Omega_{c, 2}^{22}}\left(x_{1}\right)$ in (5.1), more precisely,

$$
\begin{equation*}
\forall h_{\Omega_{c, 2}^{22}}\left(x_{1}\right) \in \mathbb{R}^{+} . d_{\Omega_{n c, 2}^{2}, \partial \Omega^{2 D}} \geq 4 h_{\Omega_{c, 2}^{2 D}}\left(x_{1}\right) . \tag{5.7}
\end{equation*}
$$

Second, the distance $d_{\Omega_{n c, 2}^{2 D}, \partial \Omega^{2 D}}$ is also relevant regarding the change of the parameters $x_{1}$ and $x_{2}$ such as within a numerical optimization. The parameters $x_{1}$ and $x_{2}$ are associated with the geometry of the space region $\Omega^{2 D}$, i.e., $\Omega^{2 D}\left(x_{1}, x_{2}\right)$ (recall $\S$ 2.2.3), thus, in a non-intrusive environment (recall § 3.1.3), a change of these parameters leads inevitably to a re-meshing. More precisely, a map $\varphi_{a}$ is provided that reads as

$$
\begin{equation*}
\varphi_{a}=\left(x_{1_{a}}, x_{2_{a}}\right) \mapsto \varphi_{a}\left(x_{1_{a}}, x_{2_{a}}\right):=\left(x_{1_{b}}, x_{2_{b}}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+} \times \mathbb{R}^{+}, \tag{5.8}
\end{equation*}
$$

where $\forall a, b \in \mathbb{N} . a \neq b \Longrightarrow\left(x_{1_{a}}, x_{2_{a}}\right) \neq\left(x_{1_{b}}, x_{2_{b}}\right)$, otherwise $\varphi_{a}$ is identical to the corresponding identity map $\mathrm{id}_{\varphi_{a}}$. If we associate a 2 -tuple ( $x_{1_{a}}, x_{2_{a}}$ ) with a simplicial triangulation $\mathcal{T}_{h_{a}}^{2 D}$ and if we associate a 2-tuple ( $x_{1_{b}}, x_{2_{b}}$ ) with a simplicial triangulation $\mathcal{T}_{h_{b}}^{2 D}$, then, by invoking the map $\varphi_{a}$ in (5.8), one can encode re-meshing figuratively by the map $\Phi_{a}$ that reads as

$$
\begin{equation*}
\Phi_{a}=\mathcal{T}_{h_{a}}^{2 D} \mapsto \Phi_{a}\left(\mathcal{T}_{h_{a}}^{2 D}\right):=\mathcal{T}_{h_{b}}^{2 D}: \Omega^{2 D}\left(x_{1_{a}}, x_{2_{a}}\right) \rightarrow \Omega^{2 D}\left(\varphi_{a}\left(x_{1_{a}}, x_{2_{a}}\right)\right), \tag{5.9}
\end{equation*}
$$

where if and only if $\Phi_{a}$ is identical to the corresponding identity map id $\Phi_{\Phi_{a}}$, the number of nodes, edges, and surfaces is preserved.

In order to mitigate the potential spurious influence of re-meshing on entities depending on the parameters $x_{1}$ and $x_{2}$ such as quantities of interest, let us apply a heuristic approach in the sense that

- the meshing is fixed to balance a highest possible resolution and a reasonable computation time (for a representative of the corresponding simplicial triangulation, see (ii) in Figure 5.1);
- any kind of adaptive meshing is disabled, and
- the distance $d_{\Omega_{n, 2}^{2 D}, \partial \Omega^{2 D}}$ in (5.7) is fixed to at least four times $h_{\Omega_{c, 2}^{2 D}}\left(x_{1, \text { max }}\right)$ where $x_{1, \max }$ is conceived as the maximal number in a bounded interval $\left[x_{1, \min }, x_{1, \max }\right]$.

For more details on topics such as re-meshing and the like, I refer to the research, for instance, within the field of shape optimization (see, e.g., [54] and references therein).

In technical applications (recall Figure 1.1), it is common to adopt an electrical network viewpoint regarding an inductive component within an electromagnetic system such that the inductive component is expressed by a circuit diagram representation.

In Appendix B.1, the electrical network viewpoint is concisely elaborated. Further, I recall the relationship between entities at the field theoretical level such as $\bar{P}_{\mathrm{L}} \epsilon$ $\mathbb{R}^{+}$that denotes the time-averaged ohmic loss in $\Omega^{2 D}$, and $\bar{W}_{\mathrm{m}} \in \mathbb{R}^{+}$that denotes the time-averaged magnetic energy in $\Omega^{2 D}$, and entities at the circuit theoretical level such as $R \in \mathbb{R}^{+}$that denotes the resistance, and $L \in \mathbb{R}^{+}$that denotes the inductance.

In order to formulate an optimization problem in accordance with our elaborations in $\S 2.3$, let us introduce the map $\hat{j}_{\bar{P}_{\mathrm{L}}}$ and the map $\hat{Q}_{L}$ such that

$$
\begin{align*}
& \hat{j}_{\bar{P}_{\mathrm{L}}}=\left(\omega, x_{1}, x_{2}\right) \mapsto \bar{P}_{\mathrm{L}} \equiv \hat{\bar{p}}_{\bar{P}_{\mathrm{L}}}\left(\omega, x_{1}, x_{2}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+},  \tag{5.10a}\\
& \hat{Q}_{L}=\left(\omega, x_{1}, x_{2}\right) \mapsto L \equiv \hat{Q}_{L}\left(\omega, x_{1}, x_{2}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}, \tag{5.10b}
\end{align*}
$$

where $x_{1}$ and $x_{2}$ refer to the entities in (i) of Figure 5.1. Next, recalling (3.128), one can construct a new map $\hat{\dot{j}}_{\bar{P}_{\mathrm{L}}, \omega}$ and a new map $\hat{Q}_{L, \omega}$ by currying such that

$$
\begin{align*}
& \hat{\bar{j}}_{\mathrm{L}, \omega}=\omega \mapsto\left(\left(x_{1}, x_{2}\right) \mapsto \hat{\dot{p}}_{\bar{P}_{\mathrm{L}}}\left(\omega, x_{1}, x_{2}\right)\right): \mathbb{R}^{+} \rightarrow \mathbb{R}^{+\mathbb{R}^{+} \times \mathbb{R}^{+}},  \tag{5.11a}\\
& \hat{Q}_{L, \omega}=\omega \mapsto\left(\left(x_{1}, x_{2}\right) \mapsto \hat{Q}_{L}\left(\omega, x_{1}, x_{2}\right)\right): \mathbb{R}^{+} \rightarrow \mathbb{R}^{+\mathbb{R}^{+} \times \mathbb{R}^{+}} . \tag{5.11b}
\end{align*}
$$

If one fixes the operating frequency $f_{0}$ and the angular operating frequency $\omega_{0} \equiv 2 \pi f_{0}$, then one can define the map $\hat{\dot{p}}_{\bar{P}^{\prime}, \omega_{0}}$ and the map $\hat{Q}_{L, \omega_{0}}$ such that

$$
\begin{align*}
& {\hat{\dot{ }} \bar{P}_{L}, \omega_{0}}=\left(x_{1}, x_{2}\right) \mapsto \hat{j}_{\bar{P}_{\mathrm{L}}}\left(\omega_{0}, x_{1}, x_{2}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+},  \tag{5.12a}\\
& \hat{Q}_{L, \omega_{0}}=\left(x_{1}, x_{2}\right) \mapsto \hat{Q}_{L}\left(\omega_{0}, x_{1}, x_{2}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}, \tag{5.12b}
\end{align*}
$$

where ${\hat{\bar{P}_{\bar{L}}, \omega_{0}}} \equiv \hat{j}_{\bar{P}_{\mathrm{L}}, \omega}\left(\omega_{0}\right)$ and $\hat{Q}_{\mathrm{L}, \omega_{0}} \equiv \hat{Q}_{\mathrm{L}, \omega}\left(\omega_{0}\right)$.
Subsequently, the operating frequency $f_{0}$ is set to $f_{0}:=1 \times 10^{5} \mathrm{~Hz}$. This choice is based on a rough heuristic estimate of the scale of the winding losses. This heuristic estimate reads as

$$
\begin{equation*}
e_{\bar{P}_{\mathrm{L}}}=\left(x_{1}, f_{0}\right) \mapsto 2 x_{1} / \delta_{S}\left(f_{0}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}, \tag{5.13}
\end{equation*}
$$

where $\delta_{S}$ has the signature $\mathbb{R}^{+} \rightarrow \mathbb{R}^{+}$and $\delta_{S}\left(f_{0}\right):=\left(\pi f_{0} \mu_{0} \sigma_{\mathrm{Cu}}\right)^{-\frac{1}{2}}$ denotes the socalled skin depth (cf. [103, p. 220]) w.r.t. a good single conductor (recall § 3.1.3) evaluated at the operating frequency $f_{0}$.

If a lower bound of $x_{1}$ is set to $x_{1, \min }:=1 \times 10^{-3} \mathrm{~m}$ and if an upper bound of $x_{1}$ is set to $x_{1, \text { max }}:=3 \times 10^{-3} \mathrm{~m}$, then one can ascertain the corresponding values of the estimate $e_{\bar{P}_{\mathrm{L}}, f_{0}}$ for various operating frequencies in Table 5.1.

In Table 5.1, one can observe that the scale of the winding losses is, very roughly estimated, forty-five times greater at the frequency $f_{0}:=1 \times 10^{5} \mathrm{~Hz}$ than at the frequency $f_{0}:=5 \times 10^{1} \mathrm{~Hz}$. Thus, given the frequency $f_{0}:=1 \times 10^{5} \mathrm{~Hz}$, I conclude that the winding losses are at such a significant level which can be critical in technical applications.

Notice that the estimate in (5.13) can also be used to illustrate the decrease of the magnetic energy due to a conductor's higher magnetic shielding effect at higher operating frequencies. However, the rate of the magnetic energy's decrease is much lower compared to the rate of the winding losses' increase. ${ }^{2}$

From an electromagnetic system point of view, it might be more adequate to regard the time-averaged ohmic loss density, that is, the time-averaged ohmic loss

[^51]TABLE 5.1: Given a lower bound $x_{1, \text { min }}$ and an upper bound $x_{1, \text { max }}$, the rough heuristic estimate of the scale of the winding losses in (5.13) for various operating frequencies $f_{0}$.

| (A) The lower bound $x_{1, \min }$ set to$1 \times 10^{-3} \mathrm{~m}$ |  | (B) The upper bound $x_{1, \text { max }}$ set to$3 \times 10^{-3} \mathrm{~m}$ |  |
| :---: | :---: | :---: | :---: |
| $f_{0}[\mathrm{~Hz}]$ | $e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \min }, f_{0}\right)[1]$ | $f_{0}[\mathrm{~Hz}]$ | $e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \text { max }}, f_{0}\right)[1]$ |
| $5 \times 10^{1}$ | 0.217 | $5 \times 10^{1}$ | 0.651 |
| $1 \times 10^{2}$ | 0.307 | $1 \times 10^{2}$ | 0.920 |
| $1 \times 10^{3}$ | 0.970 | $1 \times 10^{3}$ | 2.910 |
| $1 \times 10^{4}$ | 3.068 | $1 \times 10^{4}$ | 9.204 |
| $1 \times 10^{5}$ | 9.701 | $1 \times 10^{5}$ | 29.104 |
| $1 \times 10^{6}$ | 30.678 | $1 \times 10^{6}$ | 92.035 |
| $1 \times 10^{7}$ | 97.014 | $1 \times 10^{7}$ | 291.041 |
| $1 \times 10^{8}$ | 306.784 | $1 \times 10^{8}$ | 920.353 |

normalized to a volume under test $V_{\mathrm{ut}}$. Hence, the time-averaged ohmic loss density enables a comparison of different solenoids with core (recall Figure 5.1) by taking into account a restriction of the available space within an electromagnetic system.

Given a map $V_{\mathrm{ut}}=\left(x_{1}, x_{2}\right) \mapsto V_{\mathrm{ut}}\left(x_{1}, x_{2}\right)$ with the signature $\mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}$, one can define the map $\hat{\bar{j}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}}$, the map $\hat{\dot{\Gamma}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega^{\prime}}$ and the map $\hat{\bar{j}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega_{0}}$ analogous to (5.10a), (5.11a), and (5.12a), that is,

$$
\begin{gather*}
\hat{\dot{p}}_{\mathrm{P}_{\mathrm{L}}, V_{\mathrm{ut}}}=\left(\omega, x_{1}, x_{2}\right) \mapsto \hat{\dot{p}}_{\bar{P}_{\mathrm{L}}}\left(\omega, x_{1}, x_{2}\right) / V_{\mathrm{ut}}\left(x_{1}, x_{2}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+},  \tag{5.14a}\\
\hat{\dot{j}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega}=\omega \mapsto\left(\left(x_{1}, x_{2}\right) \mapsto \hat{\dot{P}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}}\left(\omega, x_{1}, x_{2}\right)\right): \mathbb{R}^{+} \rightarrow \mathbb{R}^{+\mathbb{R}^{+} \times \mathbb{R}^{+}},  \tag{5.14b}\\
\hat{j}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut},}, \omega_{0}}=\left(x_{1}, x_{2}\right) \mapsto \hat{\dot{p}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}}\left(\omega_{0}, x_{1}, x_{2}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}, \tag{5.14c}
\end{gather*}
$$

where $\hat{j}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega_{0}} \equiv \hat{j}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega}\left(\omega_{0}\right)$.
I conceive $V_{\mathrm{ut}}\left(x_{1}, x_{2}\right) \in \mathbb{R}^{+}$in (5.14) as the volume of the core $V_{\mathrm{c}}\left(x_{1}, x_{2}\right) \in \mathbb{R}^{+}$combined with the volume of the winding $V_{\mathrm{w}}\left(x_{1}, x_{2}\right) \in \mathbb{R}^{+}$, that is,

$$
\begin{equation*}
V_{\mathrm{ut}}\left(x_{1}, x_{2}\right) \equiv V_{\mathrm{c}}\left(x_{1}, x_{2}\right)+V_{\mathrm{w}}\left(x_{1}, x_{2}\right), \tag{5.15}
\end{equation*}
$$

where $V_{\mathrm{c}}\left(x_{1}, x_{2}\right)$ and $V_{\mathrm{w}}\left(x_{1}, x_{2}\right)$ are determined by means of numerical integration. Recall that it is assumed that a cylindrical core and $N$ toroids are associated with the spatial domain in Figure 5.1. Hence, one can also determine $V_{\mathrm{c}}\left(x_{1}, x_{2}\right)$ and $V_{\mathrm{w}}\left(x_{1}, x_{2}\right)$ by means of the following formulae:

$$
\begin{equation*}
V_{\mathrm{c}}\left(x_{1}, x_{2}\right):=\pi\left(0.5 w_{\Omega_{c, 2}^{2 D}}\left(x_{1}, x_{2}\right)\right)^{2} h_{\Omega_{c, 2}^{2 D}}\left(x_{1}, x_{2}\right), \quad V_{\mathrm{w}}\left(x_{1}, x_{2}\right):=N 2 \pi^{2} x_{1}^{2} x_{2} . \tag{5.16}
\end{equation*}
$$

However, let us proceed with the computation of $V_{\mathrm{ut}}\left(x_{1}, x_{2}\right)$ by numerical integration since it is a more general approach. Due to numerical inaccuracies, though, there are slight differences between the computation of $V_{\mathrm{ut}}\left(x_{1}, x_{2}\right)$ by numerical integration and by the formulae in (5.16).

### 5.1.2 Optimization problem I

Without loss of general applicability, in the subsequent exemplification regarding surrogate optimization, let us focus on the term $\hat{\bar{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ in (5.12a) rather than on the term $\hat{\dot{\Gamma}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ in (5.14c); or to put it differently, let us focus on the ohmic loss rather than the ohmic loss density. Hence, I investigate the following high-fidelity optimization problem as a concrete instance of the abstract optimization problem in (2.36) that reads as

$$
\begin{align*}
\min . & \hat{j}_{\bar{P}_{L}, \omega_{0}}\left(x_{1}, x_{2}\right)  \tag{5.17a}\\
\text { s.t. } & x_{1, \min } \leq x_{1} \leq x_{1, \max }  \tag{5.17b}\\
& x_{2, \min } \leq x_{2} \leq x_{2, \max }  \tag{5.17c}\\
& L_{\min }-\hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right) \leq 0,  \tag{5.17d}\\
& \hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)-L_{\max } \leq 0, \tag{5.17e}
\end{align*}
$$

where $\omega_{0}:=2 \pi 100 \mathrm{kHz}, x_{1, \text { min }}:=1 \times 10^{-3} \mathrm{~m}, x_{1, \text { max }}:=3 \times 10^{-3} \mathrm{~m}, x_{2, \text { min }}:=5 \times 10^{-3} \mathrm{~m}$, $x_{2 \text {,max }}:=10 \times 10^{-3} \mathrm{~m}, L_{\text {min }}:=2.5 \times 10^{-6} \mathrm{H}$, and $L_{\text {max }}:=3.5 \times 10^{-6} \mathrm{H}$.
Remark 5.1.1. Due to production considerations in practical applications, there might be only integer length quantities or a mix of a real length quantity and an integer length quantity. Hence, this kind of modeling issue might be covered adequately by an integer optimization problem or a mixed-integer optimization problem. However, in the present work, I regard the problem in (5.17) as a reasonable approximation of potential modeling implications due to production considerations in practical applications.
Remark 5.1.2. Recalling § 2.3.1, one can observe that in (5.17), there is an evaluated reduced parametric quantity of interest in the objective functional and there is an evaluated reduced parametric quantity of interest in the constraints, as well.

It is supposed that the relation in (3.106) holds to be true, at least, in a worst-case sense. Regarding direct solving of the high-fidelity optimization problem in (5.17) by means of an adequate optimization algorithm from $\S 2.3 .3$, a worst-case scenario can be conceived as, e.g., the unfavorable choice of an initial point for a locally convergent algorithm. ${ }^{3}$

For instance, if we apply the COBYLA algorithm (recall § 2.3.3) to the problem in (5.17) with the initial point $x^{(0)}$ and the maximum number of high-fidelity function evaluations $m_{\text {DSO,max }}$ that read as

$$
\begin{equation*}
m_{\mathrm{DSO}, \max } \equiv 30 \quad x^{(0)}:=\left(1.1 \times 10^{-3} \mathrm{~m}, 9.9 \times 10^{-3} \mathrm{~m}\right), \tag{5.18}
\end{equation*}
$$

then one receives the subsequent $\log$ data in an abridged version, i.e.,

$$
\begin{align*}
& m_{\mathrm{DSO}} \equiv 30  \tag{5.19a}\\
& x^{*}:=\left(2.98 \times 10^{-3} \mathrm{~m}, 8.86 \times 10^{-3} \mathrm{~m}\right) \\
& \hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right) \equiv 96.32 \times 10^{-3} \mathrm{~W} \hat{Q}_{\mathrm{L}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right):=2.51 \times 10^{-6} \mathrm{H}, \tag{5.19b}
\end{align*}
$$

where the 3-tuple $\left(\Delta_{\hat{j}_{\bar{T}_{L}, w_{0}}}, \Delta_{\hat{Q}_{L, \omega_{0}}}, \Delta_{x^{*}}\right) \in \mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+}$is set to

$$
\begin{equation*}
\left(\Delta_{{\hat{T_{\bar{P}}^{L}}}, \omega_{0}}, \Delta_{\hat{Q}_{L, \omega_{0}}}, \Delta_{x^{*}}\right):=\left(1.0 \times 10^{-4}, 1.0 \times 10^{-4}, 1.0 \times 10^{-8}\right), \tag{5.20}
\end{equation*}
$$

[^52]which is constituted by the absolute accuracy threshold for the evaluated objective functional $\Delta_{\hat{j}_{\bar{P}_{L}, \omega_{0}}}$, the absolute accuracy threshold for the constraints $\Delta_{\hat{Q}_{L, \omega_{0}}}$, and the relative accuracy threshold for the optimal solution $\Delta_{x^{*}}$. Notice that I do not dwell on these thresholds since they appear problem-dependent. However, let us retain the 3-tuple $\left(\Delta_{\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}}, \Delta_{\hat{Q}_{L, \omega_{0}}}, \Delta_{x^{*}}\right)$ for all optimization problems under consideration. To put it differently: It is assumed that, in some sense adequate, inclusion maps (similarly to, e.g., (3.122)) exist for all optimization problems under consideration.

The initial point $x^{(0)}$ in (5.18) is not an admissible point since the inductance is $\hat{Q}_{L, \omega_{0}}\left(x_{1}^{(0)}, x_{2}^{(0)}\right):=3.98 \times 10^{-6} \mathrm{H}$. However, even if one selects heuristically ${ }^{4}$ an admissible initial point, then the optimization algorithm still detects the solution $x^{*}$ in (5.19a). Hence, I conclude that, at the level of programs (recall Figure 1.4), the implementation of the COBYLA algorithm possesses a coping mechanism to deal with not admissible initial points.

By numerical experiments and theoretical considerations, the optimal solution $x^{*}$ in (5.19a) is plausible. For instance, if one removes the constraints regarding the inductance in (5.17), more precisely, if one removes (5.17d) and (5.17e), then the computed optimal solution is located at $\left(x_{1, \max }, x_{2, \min }\right)$. Recalling Figure 5.1, one can put it in 3D terms: The largest possible radius of the cross-section areas of the $N$ toroids and the smallest possible radius of the $N$ toroids themselves constituted the computed optimal solution.

The computed optimal solution $\left(x_{1, \max }, x_{2, \min }\right)$, though, corresponds to the lowest possible inductance regarding the constraint in (5.17b) and the constraint in (5.17c). Hence, I deem it reasonable to assume that the theoretical optimal solution in (5.17) is at least a member of the level set $\mathfrak{L}_{L_{\text {min }}}\left(\hat{Q}_{L, \omega_{0}}\right)$ that can be defined by set-builder notation as

$$
\begin{equation*}
\mathfrak{L}_{L_{\min }}\left(\hat{Q}_{L, \omega_{0}}\right):=\left\{\left(x_{1}, x_{2}\right) \in \mathbb{R}^{+} \times \mathbb{R}^{+} \mid \hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)-L_{\min }=0\right\} \tag{5.21}
\end{equation*}
$$

Furthermore, if one compares the solution $\left(x_{1, \max }, x_{2, \min }\right)$ with the solution $x^{*}$ in (5.19a), then I prudently infer that, concerning $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ in (5.17), the parameter $x_{1}$ possess probably a higher relevance compared to the parameter $x_{2}$. This inference seems reasonable from a physical viewpoint: If a conceptual approximation is made in the sense that one imagines that the $N$ toroids corresponding to Figure 5.1 are replaced by $N$ cylinders of the length $x_{2}$ and the radius $x_{1}$ of the cross-section area, then one can state, roughly speaking, that $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ is proportional to $x_{2}$ and proportional to $1 / x_{1}^{2}$. These proportional relations furnish us with some indications regarding the order of relevance of the parameters $x_{1}$ and $x_{2}$ w.r.t. $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ in (5.17).

For a surrogate-based optimization, let us use a Sobol quasi-random sequence sampling plan with $m:=21$ and the data-fit low-fidelity models in §3.2.1. In Appendix B.2, I present a visualization of the evaluated data-fit low-fidelity models regarding (5.12), (5.14c), and (5.15). In Appendix B.2, potential scaling issues are taken into account (recall $\S 3.2 .1$ ) by setting up the appropriate units of measure for the corresponding physical dimensions.

Recalling § 3.2.1, one can most likely rule out that the corresponding unknown evaluated high-fidelity models behave like the Ackley function or the Michalewicz function. It is probable that these models behave like one of the other functions

[^53](i.e., the Unit sphere function, the Booth function, the Rosenbrock function or the Modified Branin function) in its outer regions.

Concerning the evaluated data-fit low-fidelity models in Appendix B.2, observe that, in Table 5.2, the normalized global first-order sensitivity measures $S_{1}^{N}$ and $S_{2}^{N}$ (cf. (2.51)) and, given $m_{j} \equiv 50$ and $m_{j-1} \equiv 21$, a low-fidelity models' normalized global first-order sensitivity measures (LFSM) error $\mathfrak{e}_{m_{j}}\left(S_{\tilde{y}, i}^{\mathbb{N}}\right)$ (cf. (3.37)) are computed. I emphasize the data relevant to the exemplification regarding surrogate optimization by means of coloring, i.e., $\tilde{\hat{F}}_{\bar{P}_{L}, \omega_{0}}$ and $\tilde{\hat{Q}}_{\mathrm{L}, \omega_{0}}$. Mind that, as mentioned above, the higher relevance of the parameter $x_{1}$ compared to the parameter $x_{2}$ is plausible from a physical viewpoint.

Table 5.2: (Ia) $S_{i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at (a) $f \equiv \tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}$ and (b) $f \equiv \tilde{V}_{\text {ut }}$ w.r.t. the Figure B.6; (Ib) Given $m_{j}:=50$, LFSM error $\mathfrak{e}_{m_{j}}\left(S_{\hat{y}, i}^{\mathrm{N}}\right)$ w.r.t. (Ia); (IIa) $S_{i}^{\mathrm{N}}$ with $i \in\{1,2\}$ evaluated at (a) $f \equiv \hat{\bar{j}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{Ut}}, \omega_{0}}$ and (b) $f \equiv \tilde{\hat{Q}}_{L, \omega_{0}}$ w.r.t. the Figure B.7; (IIb) Given $m_{j}:=50$, LFSM error $\mathfrak{e}_{m_{j}}\left(S_{\hat{y}, i}^{N}\right)$ w.r.t. (IIa).

| (Ia) | (1a) | (2a) | (3a) | (1b) | (2b) | (3b) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $S_{1}^{\mathrm{N}}(f)$ | 0.6751 | 0.6854 | 0.6860 | 0.8930 | 0.8927 | 0.8631 |
| $S_{2}^{\mathrm{N}}(f)$ | 0.3249 | 0.3146 | 0.3140 | 0.1070 | 0.1073 | 0.1369 |
| $\Sigma_{i=1}^{2} S_{i}(f)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| ( lb ) | (1a) | (2a) | (3) | (1 | (2b) | (3b) |
| $\mathfrak{e}_{m_{j}}\left(S_{\hat{\hat{u}}, 1}^{\mathrm{N}}\right)$ | -0.0066 | -0.0245 | -0.0551 | -0.0009 | -0.0029 | +0.0014 |
| $\mathfrak{e}_{m_{j}}\left(S_{\hat{\hat{y}}, 2}^{\mathrm{N}}\right.$ ) | +0.0134 | +0.0495 | +0.1023 | +0.0074 | +0.0237 | -0.0088 |
| (IIa) | (1a) | (2a) | (3a) | (1b) | (2b) | (3b) |
| $S_{1}^{\mathrm{N}}(f)$ | 0.9720 | 0.9700 | 0.9704 | 0.6074 | . 6166 | 0.6223 |
| $S_{2}^{\mathrm{N}}(f)$ | 0.0280 | 0.0300 | 0.0296 | 0.3926 | 0.3834 | 0.3777 |
| $\Sigma_{i=1}^{2} S_{i}(f)$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| (IIb) | (1a) | (2a) | (3a) | (1b) | (2b) | (3b) |
| $\mathfrak{e}_{m_{j}}\left(S_{\hat{\hat{u}}, 1}^{\mathrm{N}}\right)$ | -0.0016 | -0.0010 | -0.0203 | -0.0018 | -0.0123 | $-0.0114$ |
| $\mathfrak{e}_{m_{j}}\left(S_{\hat{\hat{y}}, 2}^{\mathrm{N}}\right)$ | +0.0541 | +0.0323 | +0.3947 | +0.0028 | +0.0192 | +0.0182 |

Notice well that, due to the examination in $\S$ 3.2.1, if one can most probably rule out that the unknown evaluated high-fidelity models are members of the same class as the Ackley function or members of the same class as the the Michalewicz function, then it is probably reasonable to expect that the behavior of the corresponding lowfidelity models' sensitivity measures w.r.t. the number of training sampling points is relatively fast converging - especially in the case of the kriging low-fidelity model.

Hence, in practical applications such as, e.g., the rapid prototyping part of a product development cycle, the Table 5.2 can serve as an approximate proxy for other indicators (recall the conjecture in § 3.1.1) and it can serve as an overview of the relevance of the parameters.

In the subsequent considerations, let us invoke the kriging low-fidelity model. If we apply the COBYLA algorithm to the corresponding low-fidelity optimization problem with the initial point $\tilde{x}^{(0)}:=\left(1.1 \times 10^{-3} \mathrm{~m}, 9.9 \times 10^{-3} \mathrm{~m}\right)$, then one can compute the optimal solution $\tilde{x}^{*}$ that reads as

$$
\begin{equation*}
\tilde{x}^{*}:=\left(3.00 \times 10^{-3} \mathrm{~m}, 8.87 \times 10^{-3} \mathrm{~m}\right) . \tag{5.22}
\end{equation*}
$$

And if we apply the COBYLA algorithm to the problem in (5.17) with the maximum number of high-fidelity function evaluations $\tilde{m}_{\mathrm{DSO}, \max }:=2$ and the initial point $x^{(0)}:=\tilde{x}^{*}$, one receives the surrogate-based optimization's $\log$ data in an abridged version, i.e.,

$$
\begin{array}{rr}
m_{\mathrm{SBO}} \equiv 23 & x^{*}:=\left(3.00 \times 10^{-3} \mathrm{~m},\right.  \tag{5.23a}\\
\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right) \equiv 96.26 \times 10^{-3} \mathrm{~W} & \hat{Q}_{L, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right):=2.50 \times 10^{-6} \mathrm{H},
\end{array}
$$

$$
x^{*}:=\left(3.00 \times 10^{-3} \mathrm{~m}, 8.87 \times 10^{-3} \mathrm{~m}\right)
$$

where $m_{\mathrm{SBO}}:=m+\tilde{m}_{\mathrm{DSO}, \max }$. The low-fidelity models regarding the entities in (5.17) are constructed in parallel. If one constructs the low-fidelity models sequentially, then $m_{\mathrm{SBO}}$ in (5.23) is determined by $m_{\mathrm{SBO}}:=2 m+\tilde{m}_{\mathrm{DSO}, \max }$.

For additional investigations, one can apply the procedure SBO-DFLF in the neighborhood of the optimal solution in (5.23).

Recalling $\S 4.4 .1$, one can construct simplified-physics low-fidelity models by adapting, e.g., $\mathcal{T}_{h_{1}}^{2 D}$ or ${ }^{1} \Delta_{A x=b}^{2 D}$. Moreover, one can associate $R_{1}$ in $\S 4.4 .1$, e.g., with a physics-oriented approximation formula of $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ and $\hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)$ such as $N\left(2 \pi x_{2}\right)\left(\pi x_{1}^{2}\right)^{-1} \sigma_{\mathrm{Cu}}^{-1} I_{\mathrm{rms}}^{2}$ for the term $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ and $\mu N^{2}\left(\pi x_{2}^{2}\right) h_{\Omega_{c, 2}^{2 D}}\left(x_{1}\right)^{-1}$ for the term $\hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)$. However, in the remaining, let us focus on $\mathcal{T}_{h_{1}}^{2 D}$ and ${ }^{1} \Delta_{A x=b}^{2 D}$.

The entities $\mathcal{T}_{h_{1}}^{2 D}$ and $\mathcal{T}_{h_{2}}^{2 D}$ are associated with $\approx 1 \times 10^{5}$ elements and $\approx 1 \times 10^{3}$ elements, respectively (recall $\S 2.2 .2$ ), i.e., $\mathcal{T}_{h_{1}}^{2 D} \mapsto \approx 1 \times 10^{5}$ elements, and $\mathcal{T}_{h_{2}}^{2 D} \mapsto \approx$ $1 \times 10^{3}$ elements. Further, it is set that ${ }^{1} \Delta_{A x=b}^{2 D}:=1 \times 10^{-16}$ and ${ }^{1} \Delta_{A x=b}^{2 D}:=1 \times 10^{-8}$.

In Table 5.3, given $m:=21$, I present the corresponding mean SSPCC $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k:=5}$ within the $k$-fold cross validation method w.r.t. the simplified-physics low-fidelity model regarding $\hat{j}_{\bar{P}_{L}, \omega_{0}}$ and $\hat{Q}_{L, \omega_{0}}$ in (5.17). The reference setting is defined by the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right) .{ }^{5}$

From the Table 5.3, I infer that one can probably invoke a useful simplifiedphysics low-fidelity model by the combination $\left(\mathcal{T}_{h_{2}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$. Hence, one can proceed analogously to the procedure regarding the log data in (5.23). For additional investigations, one can apply the procedure SBO-SPLF in the neighborhood of the corresponding optimal solution.

If the solving of the optimization problem regarding the selected simplifiedphysics low-fidelity model is perceived as being slow, then one can construct a data-fit low-fidelity model w.r.t. the selected simplified-physics low-fidelity model in addition (see, e.g., procedure SGO-SPLF). Due to the exposition in § 4.4.2, there are also novel construction opportunities for simplified-physics low-fidelity models based on coordinate transformations.

[^54]Table 5.3: Given $\mathcal{T}_{h_{1}}^{2 D} \mapsto \approx 1 \times 10^{5}$ elements, $\mathcal{T}_{h_{2}}^{2 D} \mapsto \approx 1 \times 10^{3}$ elements, ${ }^{1} \Delta_{A x=b}^{2 D}:=1 \times 10^{-16},{ }^{2} \Delta_{A x=b}^{2 D}:=1 \times 10^{-8}$, and $m:=21$, the mean SSPCC $\bar{r}_{\hat{y} \tilde{\tilde{y}}}^{\tilde{\tilde{m}}_{k:=5}}$ within the $k$-fold cross validation method w.r.t. the simplified-physics low-fidelity model regarding $\hat{\dot{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}$ and $\hat{Q}_{L, \omega_{0}}$ in (5.17). The reference setting is indicated by a gray box.
(A) The mean SSPCC $\bar{r}_{\hat{y} \tilde{y}}^{k=5} \mid$
regarding $\hat{\bar{P}}_{\bar{P}_{L}, \omega_{0}}$.

|  | $\mathcal{T}_{h_{1}}^{2 D}$ | $\mathcal{T}_{h_{2}}^{2 D}$ |
| :---: | :---: | :---: |
| ${ }^{1} \Delta_{A x=b}^{2 D}$ | 1.00 | 0.90 |
| ${ }^{2} \Delta_{A x=b}^{2 D}$ | 1.00 | 0.93 |

(в) The mean SSPCC $\left.\vec{r}_{\hat{y} \tilde{\hat{y}}}^{2}\right|_{k=5}$
regarding $\hat{Q}_{L, \omega_{0}}$.

|  | $\mathcal{T}_{h_{1}}^{2 D}$ | $\mathcal{T}_{h_{2}}^{2 D}$ |
| :--- | :---: | :---: |
| ${ }^{1} \Delta_{A x=b}^{2 D}$ | 1.00 | 0.99 |
| ${ }^{2} \Delta_{A x=b}^{2 D}$ | 1.00 | 0.99 |

Subsequently, let us recall explicitly the context that we are in the midst of a rapid prototyping part of a product development cycle. Thus, it is supposed that the concern is mainly to deploy a surrogate-guided optimization in the context of validation and verification of given results of a surrogate-based optimization (recall $\S 3.3)$. Then, it is economical and sustainable to re-use as much of the data of a surrogate-based optimization as possible.

More tangibly, for a co-kriging optimization (recall § 3.3.3), let us re-use the sample of size $m=21$ associated with the data regarding $\tilde{\hat{户}}_{\bar{P}_{L}, \omega_{0}}$ and $\tilde{\hat{Q}}_{L, \omega_{0}}$ w.r.t. the surrogate-based optimization's log data in (5.23). That is, we use a proper subsample of size $m_{K}=15$ as the data of a high-fidelity model and we use a subsample of size $m_{\tilde{K}}=21$ as the data of a low-fidelity model whose output points are defined as

$$
\begin{equation*}
\tilde{y}:=\rho_{\mathrm{ck}} y_{\mathrm{SBO}} \tag{5.24}
\end{equation*}
$$

where $y_{\text {SBO }} \in \mathbb{R}^{m_{\mathrm{K}} \times 1}$ refers to the column vector representing the output points of the sample w.r.t. (5.23) and $\rho_{\mathrm{ck}} \in \mathbb{R}$ refers to a scaling parameter.

The two main reasons for the modeling choice in (5.24) are: (1) It is a correctness check at the level of programs (recall Figure 1.4) that is devised based on the elaborations in [70, p. 173-176]. More precisely, one can expect that the estimate $\hat{\rho}$ in (3.183) is computed as

$$
\begin{equation*}
\hat{\rho}:=\frac{1}{\rho_{\mathrm{ck}}} . \tag{5.25}
\end{equation*}
$$

(2) Due to the examinations in $\S 3.2 .2$, one can expect that, e.g., the SSPCC $r_{\hat{y} \tilde{y}, \tilde{K}_{1}}^{2}$ in (3.193) is close to one such that one can partly emulate the conditions concerning the situation in (3.193).

Notice that it is set that $\rho_{\mathrm{ck}}:=1.05$. Due to numerical issues (see the commentary on (3.183)), let us round the corresponding estimate $\hat{\rho}$ to two decimal places, thus, let us set $\hat{\rho}:=0.95$ in (3.184).

The objective function in (5.17a) is adapted in order to be consistent with the formulation in (3.192) in the sense that a desired value $\hat{\bar{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}, d} \in \mathbb{R}^{+}$(cf. (2.33)) is provided such that one can instantiate $\hat{j}$ as

$$
\begin{equation*}
\hat{j}=\left(x_{1}, x_{2}\right) \mapsto\left\|\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)-\hat{\dot{p}}_{\bar{P}_{\mathrm{L}}, \omega_{0}, d}\right\|_{\bar{L}_{2}}^{2}: X_{0} \rightarrow \mathbb{R}^{+}, \tag{5.26}
\end{equation*}
$$

where $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}, d}:=0.0 \times 10^{-3} \mathrm{~W}$ and $\hat{j}\left(x_{1}, x_{2}\right)=\hat{\hat{j}}\left(\hat{y}\left(x_{1}, x_{2}\right)\right.$ (cf. (2.40)) in (3.192). Due to
the choice of the desired value $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}, d}$, the optimal solution associated with the objective function in (5.17a) and the optimal solution associated with the adapted objective function in (5.26) are essentially the same.

Let us invoke the initial point in (5.18) and we receive the co-kriging optimization's log data in an abridged version, i.e.,

$$
\begin{align*}
m_{\mathrm{SGO}, \mathrm{ck}} & \equiv 15 \\
x^{*} & :=\left(3.00 \times 10^{-3} \mathrm{~m}, 8.86 \times 10^{-3} \mathrm{~m}\right) \\
\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right) \equiv 96.53 \times 10^{-3} \mathrm{~W} & \hat{\mathrm{Q}}_{\mathrm{L}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right):=2.50 \times 10^{-6} \mathrm{H}, \tag{5.27a}
\end{align*}
$$

where $m_{\mathrm{SGO}, \mathrm{ck}}:=m_{\mathrm{K}}$. Mind that, in the counting method regarding $m_{\mathrm{SGO}, \mathrm{ck}}$, the situation is mimicked where one possesses a low-fidelity model whose SSPCC $r_{\hat{y} \tilde{y}, \tilde{K_{1}}}^{2}$ is close to one and which is not equal to the high-fidelity model.

Observe that, compared with (5.23), the relative deviation in a suitable norm between the optimal solution in (5.23) and the optimal solution in (5.27) is well below one percent.

Recalling § 3.3.2, the TRASM algorithm 3.1 is applied by using the co-kriging low-fidelity models of $\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}$ and $\tilde{\hat{Q}}_{L, \omega_{0}}$ corresponding to (5.27). Let us invoke the initial point in (5.18) and we set the remaining input entities of the TRASM algorithm 3.1 as

$$
\begin{align*}
B^{(0)} & :=I & \Delta^{(0)}:=10  \tag{5.28a}\\
\left(\eta_{1}, \eta_{2}, \gamma, \zeta\right) & \equiv\left(1.0 \times 10^{-5}, 1.0 \times 10^{-1}, 0.25,2\right) & \left(k_{\max ,}, \epsilon_{\mathrm{abs}}, \epsilon_{\mathrm{rel}}\right):=\left(2,1.0 \times 10^{-3}, 1.0 \times 10^{-4}\right),
\end{align*}
$$

and we define $F_{0}^{(0)}$ by using (3.149) and by adapting (5.17b) - (5.17e) similarly to (3.154).
Hence, one receives the TRASM algorithm's log data in an abridged version, i.e.,

$$
\begin{align*}
m_{\mathrm{SGO}, \mathrm{sm}} \equiv 19 & x^{*}:=\left(3.00 \times 10^{-3} \mathrm{~m}, 8.86 \times 10^{-3} \mathrm{~m}\right)  \tag{5.29a}\\
\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right) \equiv 96.53 \times 10^{-3} \mathrm{~W} & \hat{Q}_{\mathrm{L}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right):=2.50 \times 10^{-6} \mathrm{H}, \tag{5.29b}
\end{align*}
$$

where $m_{\mathrm{SGO}, \mathrm{sm}}:=m_{\mathrm{sGO}, \mathrm{ck}}+2 k$ with $k \equiv 2$ within the TRASM algorithm 3.1. Let us evaluate the optimal solution $x^{*}$ in (5.29) by using the corresponding co-kriging lowfidelity models.

Similarly to (5.27), observe that, compared with (5.23), the relative deviation in a suitable norm between the optimal solution in (5.23) and the optimal solution in (5.29) is well below one percent.

The entities in (5.28) are partly inspired by choices in [95], but, in general, these entities are chiefly heuristically determined. Therefore, in practical applications, there might be a need for a preprocessing step in which useful values for the entities in (5.28) are determined. Additionally, in practical applications, there might be a need for an adaptive restart strategy in order to prevent a potential low experimental rate of convergence due to, for instance, too small step sizes $h^{(k)}$ within the TRASM algorithm 3.1.

In summary, we have developed a particular relevant application-driven workflow concerning the relation in (3.106). More precisely, we have carved out a use
case such that

$$
\begin{equation*}
\exists m_{\mathrm{DSO}}, m_{\mathrm{SBO}}, m_{\mathrm{SGO}, \mathrm{sm}}, m_{\mathrm{SGO}, \mathrm{ck}} \in \mathbb{N} \backslash\{0\} . m_{\mathrm{DSO}}>m_{\mathrm{SBO}}>m_{\mathrm{SGO}, \mathrm{sm}}>m_{\mathrm{SGO}, \mathrm{ck}} \tag{5.30}
\end{equation*}
$$

holds to be true. To put the statement in (5.30) more poignantly, let us assume that the computational time concerning a high-fidelity model evaluation is approximately 5 min . Then, $m_{\text {DSO }}$ maps to $150 \mathrm{~min}, m_{\text {SBO }}$ maps to $115 \mathrm{~min}, m_{\text {SGO }, \mathrm{sm}}$ maps to 95 min , and $m_{\mathrm{SGO}, \mathrm{ck}}$ maps to 75 min . Thus, we have carved out a use case in which the computational time associated with a high-fidelity optimization problem can be reduced by half. And with some additional effort, the corresponding optimal solution can be validated and verified, too.

Mind that, though, the numbers of the statement in (5.30) should be treated with caution in the light of their potential for generalization such as in (3.106). Furthermore, recalling Figure 1.4, the choice of some entities such as, e.g., the problemdependent entities in (5.28), puts a few limits of comparability at the level of programs. For the issue of comparability at the level of functions, see § 4.3.

### 5.1.3 Optimization problem II

Note that if we extend the optimization problem in (5.17) in such a way that we consider $m_{w} \in \mathbb{N}$ operating frequencies, then one can formulate heuristically an indexed family of high-fidelity optimization problems $O$ whose assignment rule reads as, e.g.,

$$
\begin{align*}
O=i \mapsto \min . & \hat{j}_{\bar{P}_{L}, \omega_{i}}\left(x_{1}, x_{2}\right)  \tag{5.31a}\\
\text { s.t. } & x_{1, \min } \leq x_{1} \leq x_{1, \max },  \tag{5.31b}\\
& x_{2, \min } \leq x_{2} \leq x_{2, \max }  \tag{5.31c}\\
& L_{\min }-\hat{Q}_{L, \omega_{i}}\left(x_{1}, x_{2}\right) \leq 0,  \tag{5.31d}\\
& \hat{Q}_{L, \omega_{i}}\left(x_{1}, x_{2}\right)-L_{\max } \leq 0, \tag{5.31e}
\end{align*}
$$

where $i \in I$ with $I:=\left\{1, \ldots, m_{w}\right\}, \omega_{i} \in W_{I}$ with $W_{I}:=\left\{\omega_{1}, \ldots, \omega_{m_{w}}\right\}, x_{1, \text { min }}:=1 \times 10^{-3} \mathrm{~m}$, $x_{1, \text { max }}:=3 \times 10^{-3} \mathrm{~m}, x_{2, \text { min }}:=5 \times 10^{-3} \mathrm{~m}, x_{2, \text { max }}:=10 \times 10^{-3} \mathrm{~m}, L_{\text {min }}:=2.5 \times 10^{-6} \mathrm{H}$, and $L_{\text {max }}:=3.5 \times 10^{-6} \mathrm{H}$.

A possible aim concerning (5.31) can be to determine the optimal solution ( $\omega_{i^{*}}, x^{*}$ ) regarding all individual high-fidelity optimization problems $O(i)$ in (5.31) for which

$$
\begin{equation*}
\forall\left(w_{i}, x\right) \in W_{I} \times X . \hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{i}{ }^{*}}\left(x^{*}\right) \leq \hat{\dot{j}}_{\bar{P}_{\mathrm{L}}, \omega_{i}}(x) \tag{5.32}
\end{equation*}
$$

holds (cf. (4.18)). If we assume that the entity $\hat{j}_{\bar{T}_{\mathrm{L}}, \omega_{i}}\left(x_{1}, x_{2}\right)$ and the entity $\hat{Q}_{L, \omega_{i}}\left(x_{1}, x_{2}\right)$ are not available quickly for all $\omega_{i} \in W_{I}$ such that one cannot invoke immediately the multivariate vector-valued use case in (3.131a), then we have to consider each highfidelity optimization problem $O(i)$ in (5.31) individually.

By applying the above-mentioned application-driven workflow to each $O(i)$ and by supposing that the numbers $m_{\mathrm{DSO}}, m_{\mathrm{SBO}}, m_{\mathrm{SGO}, \mathrm{sm}}$, and $m_{\mathrm{SGO}, \mathrm{ck}}$ in (5.30) are the same for each $\omega_{i}$, one can roughly estimate the worst-case computational burden by
$\exists m_{\mathrm{DSO}}, m_{\mathrm{SBO}}, m_{\mathrm{SGO}, \mathrm{sm}}, m_{\mathrm{SGO}, \mathrm{ck}} \in \mathbb{N} \backslash\{0\} . m_{w} m_{\mathrm{DSO}}>m_{w} m_{\mathrm{SBO}}>m_{w} m_{\mathrm{SGO}, \mathrm{sm}}>m_{w} m_{\mathrm{SGO}, \mathrm{ck}}$.

In Table 5.4, I present the $\log$ data in an abridged version w.r.t. the setting in (5.19) for the operating frequencies $5 \times 10^{1} \mathrm{~Hz}, 1 \times 10^{5} \mathrm{~Hz}$, and $1 \times 10^{8} \mathrm{~Hz}$ such that it is defined that $\omega_{1}:=2 \pi 50 \mathrm{~Hz}, \omega_{2}:=2 \pi 100 \mathrm{kHz}$, and $\omega_{3}:=2 \pi 100 \mathrm{MHz}$ in (5.31).

Table 5.4: Given the operating frequencies $5 \times 10^{1} \mathrm{~Hz}, 1 \times 10^{5} \mathrm{~Hz}$, and $1 \times 10^{8} \mathrm{~Hz}$, the $\log$ data in an abridged version w.r.t. the setting of the log data in (5.19).

| $f_{0}[\mathrm{~Hz}]$ | $m_{\text {DSO }}[1]$ | $x^{*}[(\mathrm{~m}, \mathrm{~m})]$ | $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right)[\mathrm{W}]$ | $\hat{Q}_{L, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right)[\mathrm{H}]$ |
| :---: | :---: | :---: | :---: | :---: |
| $5 \times 10^{1}$ | 30 | $\left(2.99 \times 10^{-3}, 7.23 \times 10^{-3}\right)$ | $27.63 \times 10^{-5}$ | $3.27 \times 10^{-6}$ |
| $1 \times 10^{5}$ | 30 | $\left(2.98 \times 10^{-3}, 8.86 \times 10^{-3}\right)$ | $96.32 \times 10^{-3}$ | $2.51 \times 10^{-6}$ |
| $1 \times 10^{8}$ | 30 | $\left(1.45 \times 10^{-3}, 9.78 \times 10^{-3}\right)$ | $26.24 \times 10^{-1}$ | $3.39 \times 10^{-6}$ |

Let us set $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{1}}\left(x_{1}^{*}, x_{2}^{*}\right):=27.63 \times 10^{-5} \mathrm{~W}$, and $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{2}}\left(x_{1}^{*}, x_{2}^{*}\right):=96.32 \times 10^{-3} \mathrm{~W}$, and $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{3}}\left(x_{1}^{*}, x_{2}^{*}\right):=26.24 \times 10^{-1} \mathrm{~W}$, then one can define the subsequent ratios regarding $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right)$ in Table 5.4, that is,

$$
\begin{equation*}
\frac{\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{3}}\left(x_{1}^{*}, x_{2}^{*}\right)}{\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{1}}\left(x_{1}^{*}, x_{2}^{*}\right)}:=9496.92, \quad \frac{\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{2}}\left(x_{1}^{*}, x_{2}^{*}\right)}{\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{1}}\left(x_{1}^{*}, x_{2}^{*}\right)}:=348.61, \quad \frac{\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{3}}\left(x_{1}^{*}, x_{2}^{*}\right)}{\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{2}}\left(x_{1}^{*}, x_{2}^{*}\right)}:=27.25 . \tag{5.34}
\end{equation*}
$$

If we invoke Table 5.1 and if we set $e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \text { max }}, f_{1}\right):=0.651$, and $e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{2}\right):=$ 29.104, and $e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{3}\right):=920.353$, then one can define the subsequent ratios regarding $e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \text { max }}, f_{0}\right)$ in Table 5.1, that is,

$$
\begin{equation*}
\frac{e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{3}\right)}{e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{1}\right)}:=1413.75, \quad \frac{e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{2}\right)}{e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{1}\right)}:=44.71, \quad \frac{e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{3}\right)}{e_{\bar{P}_{\mathrm{L}}}\left(x_{1, \max }, f_{2}\right)}:=31.62 . \tag{5.35}
\end{equation*}
$$

Comparing (5.35) with (5.34), one can observe that the rough heuristic estimate of the scale of the winding losses in (5.13) captures at least partly an overall trend that one can expect from the electromagnetic field theory's realm of magnetoquasistatics. Hence, the Table 5.4 furnishes us with a physics-driven plausible optimal solution in the sense of the condition in (5.32).

Notice well that, especially regarding the operating frequency $f_{0}:=1 \times 10^{8} \mathrm{~Hz}$ in Table 5.4, one can observe that the optimal solution determined by the COBYLA algorithm is sensitive to the choice of the initial point $x^{(0)}$.

Table 5.5: Given the operating frequency $1 \times 10^{8} \mathrm{~Hz}$ and the initial points in (5.36), the $\log$ data in an abridged version w.r.t. the setting of the $\log$ data in (5.19).

| $x^{(0)}[(\mathrm{m}, \mathrm{m})]$ | $m_{\mathrm{DSO}}[1]$ | $x^{*}[(\mathrm{~m}, \mathrm{~m})]$ | $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right)[\mathrm{W}]$ | $\hat{Q}_{L, \omega_{0}}\left(x_{1}^{*}, x_{2}^{*}\right)[\mathrm{H}]$ |
| :---: | :---: | :---: | :---: | :---: |
| $x_{1}^{(0)}$ | 30 | $\left(1.45 \times 10^{-3}, 9.78 \times 10^{-3}\right)$ | $26.24 \times 10^{-1}$ | $3.39 \times 10^{-6}$ |
| $x_{2}^{(0)}$ | 30 | $\left(2.79 \times 10^{-3}, 9.38 \times 10^{-3}\right)$ | $23.25 \times 10^{-1}$ | $2.56 \times 10^{-6}$ |

Using the initial points $x_{1}^{(0)}$ and $x_{2}^{(0)}$ that read as

$$
\begin{equation*}
x_{1}^{(0)}:=\left(1.1 \times 10^{-3} \mathrm{~m}, 9.9 \times 10^{-3} \mathrm{~m}\right) \quad x_{2}^{(0)}:=\left(3.0 \times 10^{-3} \mathrm{~m}, 9.9 \times 10^{-3} \mathrm{~m}\right) \tag{5.36}
\end{equation*}
$$

I report in Table 5.5 the corresponding $\log$ data in an abridged version w.r.t. the setting in (5.19) for the operating frequency $1 \times 10^{8} \mathrm{~Hz}$.

From Table 5.5 and the elaborations regarding the log data in (5.19) and a particular relevant application-driven workflow that culminated in the statement in (5.30), I infer that it is likely that the shape of $\hat{j}_{\bar{P}_{\mathrm{L}}, \omega_{3}}\left(x_{1}, x_{2}\right)$ behaves more intricately than the shape of $\hat{\bar{P}}_{\bar{P}_{L}, \omega_{2}}\left(x_{1}, x_{2}\right)$ and that the shape of $\hat{Q}_{L, \omega_{3}}\left(x_{1}, x_{2}\right)$ behaves essentially the same as the shape of $\hat{Q}_{L, \omega_{2}}\left(x_{1}, x_{2}\right)$.

Let us test the inference by constructing corresponding kriging low-fidelity models $\tilde{\hat{p}}_{\bar{P}_{L}, \omega_{i}}$ and $\tilde{\hat{Q}}_{L, \omega_{i}}$ in (5.31) where $m_{w}:=3$. It is supposed that the individual highfidelity models in (5.31) uniformly build upon the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ in Table 5.3. Furthermore, it is assumed that the corresponding kriging low-fidelity models uniformly build upon a Sobol quasi-random sequence sampling plan with $m:=50$ (recall § 3.1.1).

In Figure 5.2, I depict the evaluated kriging low-fidelity models in contour representation for $\omega_{1}:=2 \pi 50 \mathrm{~Hz}, \omega_{2}:=2 \pi 100 \mathrm{kHz}$, and $\omega_{3}:=2 \pi 100 \mathrm{MHz}$.


Figure 5.2: Given the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ in Table 5.3 and using the Sobol quasi-random sequence sampling plan with $m:=50$ and kriging low-fidelity models with (1) $\omega_{1}:=2 \pi 50 \mathrm{~Hz}$, (2) $\omega_{2}:=2 \pi 100 \mathrm{kHz}$, and (3) $\omega_{3}:=2 \pi 100 \mathrm{MHz}$; contour representation of (a) $\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega_{i}}$ and (b) $\tilde{\hat{Q}}_{\mathrm{L}, \omega_{i}}$ in (5.31) where $m_{w}:=3$.
Dark colors indicate low values; bright colors indicate high values (cf. Figure B. 4 and Figure B.5).

Invoking the application-driven workflow concerning (5.30) to each $O(i)$ in (5.31) with $m_{w}:=3$, one can unleash the machinery of surrogate-based optimization and the machinery of surrogate-guided optimization for each $O(i)$ - analogous to the strategies regarding (5.17).

Since this unleashing results in one strategy to tackle the optimization problem in (5.31) within the context of surrogate optimization, let us end the present subsection by discussing briefly some conceivable variations of this strategy based on the examination of the Figure 5.2.

Examining the Figure 5.2, I conclude that the observable behavior regarding $\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega_{i}}$ and $\tilde{\hat{Q}}_{L, \omega_{i}}$ for $\omega_{1}:=2 \pi 50 \mathrm{~Hz}$ passes a physical-driven plausibility check. More precisely, due to the rough heuristic estimate of the scale of the winding losses in (5.13), one can expect that the observable behavior regarding $\tilde{\dot{\hat{P}}}_{\bar{P}_{L}, \omega_{i}}$ and $\tilde{\hat{Q}}_{L, \omega_{i}}$ for $\omega_{1}$ is approximately explainable by the electromagnetic field theory's realm of magnetostatics - besides the realm of magnetoquasistatics.

The similar behavior of $\hat{Q}_{L, \omega_{i}}$ for $\omega_{2}:=2 \pi 100 \mathrm{kHz}$ and $\omega_{3}:=2 \pi 100 \mathrm{MHz}$ is physically deducible from the setup of the underlying boundary value problem (see the commentary on Figure 5.1). Given the material characteristics concerning Figure 5.1, it is to be expected that, after a certain threshold frequency, the electromagnetic field is completely shielded from conducting domains. Hence, it is to be expected that $\hat{Q}_{L}\left(\omega, x_{1}, x_{2}\right)$ in (5.10b) changes more for lower values of the frequency than for higher values of the frequency.

Recalling (5.35) and (5.34), I argue that the essentially different behavior of $\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega}$ for $\omega_{2}:=2 \pi 100 \mathrm{kHz}$ and $\omega_{3}:=2 \pi 100 \mathrm{MHz}$ is explicable by the degree of the action of the skin effect and the proximity effect.

However, a legitimate objection from a numerical analysis viewpoint is whether it is justified to assume that the high-fidelity models in (5.31) uniformly build upon the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ in Table 5.3. Hidden behind this assumption is the assumption that the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ provides a sufficient resolution of the action of the skin effect and the proximity effect for all frequencies under consideration.

From an application-driven viewpoint, the chosen combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ that is based upon the numerical investigations concerning a median w.r.t. the ordered set $W_{I}$ in (5.31) can be interpreted as a compromise between accuracy and speed. In the case of Figure 5.2, it is set that $W_{I}:=\left\{\omega_{1}, \omega_{2}, \omega_{3}\right\}$ such that the median w.r.t. the ordered set $W_{I}$ is $\omega_{2}:=2 \pi 100 \mathrm{kHz}$.

On an alternative path, one could base the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ upon the numerical investigations concerning the highest frequency w.r.t. the ordered set $W_{I}$ in (5.31). This choice puts an emphasis rather on accuracy than on speed since for lower frequencies the same setup is utilized as for the highest frequency.

Another potential path is to choose adaptively the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ for each frequency w.r.t. the ordered set $W_{I}$ in (5.31).

Recalling § 4.4.1, one can conceptualize the above-mentioned paths concerning the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ by employing, for instance, a diagrammatic notation such as in (4.47) for each frequency w.r.t. the ordered set $W_{I}$ in (5.31).

Observing Figure 5.2 and recalling $\S 3.2 .1$, it is possible to choose a different sample size $m$ for each frequency w.r.t. the ordered set $W_{I}$ in (5.31) as well. Notice well that such a choice has an impact on the estimate in (5.33).

### 5.2 Common-Mode Choke

### 5.2.1 Preliminary consideration

The device under test is a common-mode choke (CMC) as another representative of the class of inductive components. The exposition is similar to the notational and methodological exposition in § 5.1. Thus, the focus is primarily on adding other aspects to the discussion regarding surrogate optimization with the magnetoquasistatic model.


FIgURE 5.3: (i) A schematic illustration of a common-mode choke with a longitudinal symmetric domain. (ii) A simplicial triangulation $\mathcal{T}_{h}$ of the space region $\Omega^{2 D}$ via FEMM4.2.
(iii) Magnetic field lines of a common-mode choke due to commonmode (CM) currents. (iv) Magnetic field lines of a common-mode choke due to differential-mode (DM) currents.

For a three-dimensional representation of a common-mode choke, see, e.g., Figure 1.1. In (i) of Figure 5.3, there is a schematic illustration of a two-dimensional
representation of a common-mode choke with a longitudinal symmetric domain. ${ }^{6}$
The subdomains $\Omega_{n c, 1}^{2 D}$ and $\Omega_{n c, 3}^{2 D}$ are behaviorally similar to the subdomain $\Omega_{n c, 1}^{2 D}$ in Figure 5.1; and the subdomain $\Omega_{n c, 2}^{2 D}$ is behaviorally similar to the subdomain $\Omega_{n c, 2}^{2 D}$ in Figure 5.1. Furthermore, the conducting subdomains - indexed by $c$ - are behaviorally similar to the conducting subdomains in Figure 5.1. A notable difference in Figure 5.3 is the existence of a primary winding (that is, pairs of subdomains ( $\Omega_{c, 1 a_{i}}^{2 D} \Omega_{c, 1 b_{i}}^{2 D}$ ) with $i \in\{1, \ldots, N\}$ ) and a secondary winding (i.e., pairs of subdomains $\left(\Omega_{c, 2 a_{i}}^{2 D}, \Omega_{c, 2 b_{i}}^{2 D}\right)$ with $\left.i \in\{1, \ldots, N\}\right)$.

In (iii) of Figure 5.3, the magnetic field lines of a common-mode choke due to common-mode (CM) currents are shown. By abuse of notation regarding (5.6), let us indicate the two possible orthogonal directions of the current density w.r.t. the two-dimensional domain $\Omega_{n c, 1}^{2 d}$ as $I_{0}^{+}$and $I_{0}^{-}$. Hence, let us encode CM currents as the ordered pair ( $I_{0}^{+}, I_{0}^{-}$) for each ordered pair $\left(\Omega_{c, 11_{i}}^{2 d} \Omega_{c, 1 b_{i}}^{2 d}\right)$ and $\left(\Omega_{c, 2 a_{i}}^{2 d}, \Omega_{c, 2 b_{i}}^{2 d}\right)$.

In (iv) of Figure 5.3, the magnetic field lines of a common-mode choke due to differential-mode (DM) currents are shown. Thus, let us encode DM currents as the ordered pair $\left(I_{0}^{+}, I_{0}^{-}\right)$for each ordered pair $\left(\Omega_{c, 11_{i}}^{2 d}, \Omega_{c, 1 b_{i}}^{2 d}\right)$ and the ordered pair $\left(I_{0}^{-}, I_{0}^{+}\right)$ for each ordered pair $\left(\Omega_{c, 2 a_{i}}^{2 d}, \Omega_{c, 2 b_{i}}^{2 d}\right)$. For an in-depth elaboration on the different modes of operation of a common-mode choke, I refer to, e.g., [164, p. 346 - 352] and references therein.

In § 5.1, we discuss two optimization problems with regard to a single inductive component. Hence, I consider these optimization problems as embedded into the component level of numerical investigations concerning inductive components.

If one considers these optimization problems with regard to two or more inductive components or other components such as, e.g., in Appendix B.1, then I conceive the corresponding optimization problems as lifted onto the system level of numerical investigations concerning inductive components.

Instead of lifting the optimization problem in (5.17) and the optimization problem in (5.31) onto the system level, let us consider two other kinds of optimization problems at the system level in order to enlarge the point of view regarding the applications of surrogate optimization for inductive components.

In the subsequent elaborations, let us narrow down our attention of surrogate optimization to surrogate-based optimization, though. Observe that the utilization of surrogate-guided optimization in the context of validation and verification of the given results of a surrogate-based optimization is analogous to the exposition in (5.1.2).

### 5.2.2 Optimization problem I

Recalling the EMC filter in Figure 1.1, let us consider a prototypical version of a simplistic EMC filter in Figure 5.4. ${ }^{7}$

[^55]From a theory-driven viewpoint, examples such as in Figure 5.4 or in Figure B.2b leave the realm of the magnetoquasistatic model (recall $\S$ 2.1.3).

However, from an application-driven viewpoint, I deem it reasonable to argue that, roughly speaking, it depends on the chosen frequency range whether, for the numerical investigation of an application at hand, the magnetoquasistatic subsystem of Maxwell's equations and the complete system of Maxwell's equations itself appear appropriately useful to gain knowledge about the application at hand.

If one considers a frequency range such as, e.g., from 50 Hz to 100 MHz (recall Figure 5.2) or from 0 Hz to 200 MHz , then I claim that, for a given task, it might be beneficial to take into account the magnetoquasistatic subsystem and the complete system of Maxwell's equations. Notice that, though, the boundary of the overlapping modeling region of the subsystem and the complete system is probably fuzzy. For one classification of frequency ranges, I refer to, e.g., [164, p. 19].


FIGURE 5.4: A prototypical version of a simplistic EMC filter created within CST Studio Suite ${ }^{\circledR}$. The nomenclature from (i) of Figure 5.3 is adapted. The value $C_{1}:=1 \times 10^{-5} \mathrm{~F}$ and the value $C_{2}:=2 \times 10^{-5} \mathrm{~F}$ refer to a respective capacitance (see Figure B.1).
The boundary $\partial \Omega^{3 D}$, and the parts relevant to the excitation (four discrete ports and the ground planes), and other ( $x_{1}, x_{2}, x_{3}$ )-dependent geometrical entities (cf. Figure 5.1) are not depicted.

In the present work, however, I do not dwell on epistemological issues regarding the overlapping modeling region of the magnetoquasistatic subsystem and the complete system of Maxwell's equations. Furthermore, I do not dwell on the intricacies of boundary value problems regarding the complete system.

Let us mainly consider the example of application in Figure 5.4 in the spirit of a high-fidelity model as a black-box model (recall Figure 1.2) within a rapid prototyping part of a product development cycle.

Hence, it is assumed that, regarding the application in Figure 5.4, there are eligible parts relevant to the excitation such as, e.g., four discrete ports and ground planes, and suitable conditions at the boundary $\partial \Omega^{3 D}$ such as, e.g., a mix of open boundary conditions and electric boundary conditions. Concerning the domain $\Omega^{3 D}$, let us consider the existence of other ( $x_{1}, x_{2}, x_{3}$ )-dependent geometrical entities (cf. Figure 5.1) than those in Figure 5.4 as given.

[^56]Therefore, one can readout, for instance, the $S_{22}$-parameter depending on the angular frequency $\omega \in \mathbb{R}^{+}$and the three geometrical parameters $x_{1}, x_{2}, x_{3} \in \mathbb{R}^{+}$in Figure 5.4. Recalling our commentary on S-parameters in §4.4.2, let us conceive $S_{22}$ as the magnitude in dB of the corresponding $S_{22}$-parameter. Finally, one can overload the $S_{22}$-parameter in the sense that, similarly to (5.10), one defines the function $S_{22}$ that reads as

$$
\begin{equation*}
S_{22}=\left(\omega, x_{1}, x_{2}, x_{3}\right) \mapsto S_{22}\left(\omega, x_{1}, x_{2}, x_{3}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R} . \tag{5.37}
\end{equation*}
$$

Analogously to (5.11) and (5.12), one can construct a map $S_{22, \omega}$ and, given a fixed angular operating frequency $\omega_{0}$, a map $S_{22, \omega_{0}}$ such that

$$
\begin{align*}
S_{22, \omega} & =\omega \mapsto\left(\left(x_{1}, x_{2}, x_{3}\right) \mapsto S_{22}\left(\omega, x_{1}, x_{2}, x_{3}\right)\right): \mathbb{R}^{+} \rightarrow \mathbb{R}^{\mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+}},  \tag{5.38a}\\
S_{22, \omega_{0}} & =\left(x_{1}, x_{2}, x_{3}\right) \mapsto S_{22}\left(\omega_{0}, x_{1}, x_{2}, x_{3}\right): \mathbb{R}^{+} \times \mathbb{R}^{+} \times \mathbb{R}^{+} \rightarrow \mathbb{R}, \tag{5.38b}
\end{align*}
$$

where $S_{22, \omega_{0}} \equiv S_{22, \omega}\left(\omega_{0}\right)$.
Similarly to the application in Figure 5.1, those geometrical parameters that affect, among other things, the shape of the core of an inductive component, affect inevitably the corresponding inductance (see Figure B.1) of the inductive component (recall, e.g., Figure 5.2), too. And, due to Figure B.3, we possess an indication that a change of the inductance affects the impedance of an inductive component. Supposing an appropriate relationship between the impedance of the inductive component and the corresponding S-parameters, it is reasonable to assume that a change of the input parameters in (5.37) affects the shape of $S_{22}\left(\omega, x_{1}, x_{2}, x_{3}\right)$ in a similar way as the change of parameters affects the shape of $Z_{1}(\omega)$ and $Z_{2}(\omega)$ in Figure B.3. ${ }^{8}$

Thus, by performing a variation of the optimization problem in (5.31), one can formulate an optimization problem regarding the application in Figure 5.4 that reads as

$$
\begin{align*}
O=i \mapsto \min . & S_{22, \omega_{i}}\left(x_{1}, x_{2}, x_{3}\right)  \tag{5.39a}\\
\text { s.t. } & x_{1, \min } \leq x_{1} \leq x_{1, \max }  \tag{5.39b}\\
& x_{2, \min } \leq x_{2} \leq x_{2, \max }  \tag{5.39c}\\
& x_{3, \min } \leq x_{3} \leq x_{3, \max } \tag{5.39~d}
\end{align*}
$$

where $I:=\left\{1, \ldots, m_{w}\right\}, \omega_{i} \in W_{I}$ with $W_{I}:=\left\{\omega_{1}, \ldots, \omega_{m_{w}}\right\}$, and $x_{1, \min }:=40 \times 10^{-3} \mathrm{~m}$, $x_{1, \text { max }}:=50 \times 10^{-3} \mathrm{~m}, x_{2, \text { min }}:=10 \times 10^{-3} \mathrm{~m}, x_{2, \text { max }}:=30 \times 10^{-3} \mathrm{~m}, x_{3, \text { min }}:=2 \times 10^{-3} \mathrm{~m}$, and $x_{3, \text { max }}:=7 \times 10^{-3} \mathrm{~m}$.

A possible aim concerning (5.39) can be to determine the optimal solution $x^{*}$ such that

$$
\begin{equation*}
\forall\left(w_{i}, x\right) \in W_{I} \times X . S_{22, \omega_{i}}\left(x^{*}\right) \leq S_{22, \omega_{i}}(x) \tag{5.40}
\end{equation*}
$$

holds (cf. (4.18)). Observe, though, the slight difference regarding the formulation of the optimal solution in (5.40) compared with (5.32).

In practical applications, it is usual to observe the entity $S_{22, \omega_{i}}$ in (5.39) w.r.t., for instance, a frequency range from 0 Hz to 150 MHz where it is set that $m_{w}:=1001$. Thus, it is assumed that the entity $S_{22, \omega_{i}}$ is quickly available for all $\omega_{i} \in W_{I}$. Therefore,

[^57]the high-fidelity optimization problems $O(i)$ in (5.39) are not considered individually.

Hence, I propose to replace $m_{w}$ in the rough estimate in (5.33) by a problemdependent number $c_{m_{w}} \in \mathbb{N}$ with $c_{m_{w}} \geq 1$. The number $c_{m_{w}}$ encodes a potential differentness of the counting methods regarding the corresponding surrogate optimization methods.

For the usage of surrogate-based optimization w.r.t. (5.39), let us focus on the case $i \in I$ with $I:=\{1,2,3\}$ and $\omega_{i} \in W_{I}$ with $W_{I}:=\{115 \mathrm{MHz}, 120 \mathrm{MHz}, 125 \mathrm{MHz}\}$. This case reflects, for instance, the aim to find a potential minimal value of an S-parameter around a certain frequency or within a frequency subrange.

Thus, using a Maximin LHC (recall Figure 3.1) with $m:=35$, one can construct the corresponding kriging low-fidelity model for each $O(i)$ and one can compute the optimal solution with regard to each kriging low-fidelity model. Let us pick the optimal solution out of three possible optimal solutions that evaluates to the lowest value. Hence, the optimal solution $\tilde{x}^{*}$ reads as

$$
\begin{equation*}
\tilde{x}^{*}:=\left(40.00 \times 10^{-3} \mathrm{~m}, 30.00 \times 10^{-3} \mathrm{~m}, 2.00 \times 10^{-3} \mathrm{~m}\right) \tag{5.41}
\end{equation*}
$$

If we employ the optimal solution in (5.41) as an initial point within the implementation of the NMS algorithm (recall § 2.3.3) of CST Studio Suite ${ }^{\circledR}$ (recall Figure 5.4), then one can observe no significant change w.r.t. the optimal solution in (5.41) after the maximum number of high-fidelity function evaluations $\tilde{m}_{\mathrm{DSO}, \max }:=2$.

Given a random initial point $x^{(0)}$ that can be written as

$$
\begin{equation*}
x^{(0)}:=\left(50.00 \times 10^{-3} \mathrm{~m}, 10.00 \times 10^{-3} \mathrm{~m}, 7.00 \times 10^{-3} \mathrm{~m}\right) \text {, } \tag{5.42}
\end{equation*}
$$

then, on average, $S_{22, \omega_{i}}$ w.r.t. $\tilde{x}^{*}$ is more than $40 \%$ lower than $S_{22, \omega_{i}}$ w.r.t. $\tilde{x}^{(0)}$.
If we utilize the initial point $x^{(0)}$ in (5.42) within the implementation of the NMS algorithm of CST Studio Suite ${ }^{\circledR}$, then, after the maximum number of high-fidelity function evaluations $m_{\mathrm{DSO}, \text { max }}:=40$, it returns an optimal solution such as in (5.41).

Setting $m_{\mathrm{DSO}}:=m_{\mathrm{DSO}, \text { max }}$ and $m_{\mathrm{SBO}}:=m+\tilde{m}_{\mathrm{DSO}, \text { max }}$, one can observe a use case such that

$$
\begin{equation*}
\exists m_{\mathrm{DSO}}, m_{\mathrm{SBO}} \in \mathbb{N} \backslash\{0\} . m_{\mathrm{DSO}}>m_{\mathrm{SBO}} \tag{5.43}
\end{equation*}
$$

holds to be true. However, concerning the statement in (5.43) the same caveats apply as for the statement in (5.30).

For even more complex issues regarding surrogate optimization w.r.t. the example of application in Figure 5.4, the tools from the category theoretical language in ch. 4 can beneficially provide notational and methodological guidance (see, e.g., § 4.4.2).

### 5.2.3 Optimization problem II

In Figure 5.5, a representation of two common-mode chokes within FEMM4.2 and their magnetic field lines due to DM currents is provided. ${ }^{9}$

In the test cases, some relevant choices regarding the geometric modeling of the two CMCs are that both the number of turns of the primary winding and the number

[^58]

Figure 5.5: A representation of two common-mode chokes within FEMM4.2 and their magnetic field lines due to DM currents.
(i) A schematic illustration of the two CMCs with geometrical parameters $\left(x_{1}, x_{2}\right) \in \mathbb{R}^{+} \times \mathbb{R}^{+}$, and $x_{1} \equiv r_{\mathrm{cmc}}$, and $x_{2} \equiv \varphi_{\mathrm{cmc}}$.
(ii) Magnetic field lines for the choice $\left(x_{1}, x_{2}\right):=\left(r_{\mathrm{cm}}, 0^{\circ}\right)$.
(iii) Magnetic field lines for the choice $\left(x_{1}, x_{2}\right):=\left(r_{\mathrm{cm}}, 45^{\circ}\right)$.
(iv) Magnetic field lines for the choice $\left(x_{1}, x_{2}\right):=\left(r_{\mathrm{cm}}, 90^{\circ}\right)$.
of turns of the secondary winding are set to 9 and the radius of the cross-section areas w.r.t. the primary winding and the secondary winding are both set to 2 mm . Furthermore, the geometrical specifications of the cores of both CMCs are defined as subsequently: The inner radius is set to 18 mm , the width is set to 12 mm and the height is set to 15 mm .

However, analogously to the exposition in § 5.2.2, let us mainly consider the example of application in Figure 5.5 in the spirit of a high-fidelity model as a blackbox model within a rapid prototyping part of a product development cycle.

Recalling the short list in $\S 5.1$.1, let us formulate an optimization problem that is concerned with the inductive coupling between the two CMCs.

From a field theoretical perspective, I conceive a situation such as in (ii) of Figure 5.5 as a situation of low inductive coupling between the two CMCs and I conceive a situation such as in (iv) of Figure 5.5 as a situation of high inductive coupling
between the two CMCs. For an electrical network theoretical perspective on inductive coupling, see, e.g., [164, ch. 11] and references therein.

Given this viewpoint, if we provide box constraints regarding the geometric parameters $x_{1}$ and $x_{2}$ that are depicted in (i) of Figure 5.5 such that $x_{1} \in\left[x_{1, \min }, x_{1, \max }\right]$ and $x_{2} \in\left[x_{2, \min }, x_{2, \max }\right]$ where $x_{1, \min }:=81.25 \times 10^{-3} \mathrm{~m}$, and $x_{1, \max }:=146.25 \times 10^{-3} \mathrm{~m}$, and $x_{2, \min }:=0^{\circ}$, and $x_{2, \max }:=90^{\circ}$, then one can expect the solution $x_{\mathrm{H}}^{*}$ that reads as

$$
\begin{equation*}
x_{\mathrm{H}}^{*}:=\left(x_{1, \min }, x_{2, \max }\right) \tag{5.44}
\end{equation*}
$$

to correspond to the situation of highest inductive coupling between the two CMCs. Thus, in practical applications, it is desirable to detect a solution such as in (5.44) in order to avoid it. Mind that, though, from a design viewpoint, it is probably reasonable to assume that

$$
\begin{equation*}
\forall x_{1} \in\left[x_{1, \min }, x_{1, \max }\right] \cdot x_{\mathrm{H}}^{*}:=\left(x_{1}, x_{2, \max }\right) \tag{5.45}
\end{equation*}
$$

holds to be true as solutions to be avoided. That is, all solutions in (5.45) are perceived as equally bad.

Notice well that, analogously to (5.44) and to (5.45), it is probably reasonable to assume that, from a design viewpoint,

$$
\begin{equation*}
\forall x_{1} \in\left[x_{1, \min }, x_{1, \max }\right] . x_{\mathrm{L}}^{*}:=\left(x_{1}, x_{2, \min }\right) \tag{5.46}
\end{equation*}
$$

holds to be true as solutions to be pursued. That is, all solutions in (5.46) are perceived as equally good.

Given these preliminary thoughts, let us anticipate some kind of symmetrical behavior of the evaluated objective function w.r.t. all parameter configurations ( $x_{1}, 45^{\circ}$ ) (see (iii) of Figure 5.5).

Thus, let us choose heuristically a map $\hat{Q}_{L, \omega_{0}}$ as an objective function that is analogously defined to (5.12b). The map $\hat{Q}_{L, \omega_{0}}$ serves as an approximate proxy to encode the considerations about the inductive coupling of the two CMCs in Figure 5.5. Regarding the map $\hat{Q}_{L, \omega_{0}}$, it is supposed that $V \equiv \Omega^{2 D}$ in (B.4) and (B.5).

Hence, let us investigate the following high-fidelity optimization problem

$$
\begin{align*}
\min . & \hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)  \tag{5.47a}\\
\text { s.t. } & x_{1, \min } \leq x_{1} \leq x_{1, \max }  \tag{5.47b}\\
& x_{2, \min } \leq x_{2} \leq x_{2, \max } \tag{5.47c}
\end{align*}
$$

where $\omega_{0}:=2 \pi 50 \mathrm{~Hz}, x_{1, \min }:=81.25 \times 10^{-3} \mathrm{~m}, x_{1, \max }:=146.25 \times 10^{-3} \mathrm{~m}, x_{2, \min }:=0^{\circ}$, $x_{2, \max }:=90^{\circ}$.

In Figure 5.6, I present a contour representation of an evaluated kriging lowfidelity model $\tilde{Q}_{L, \omega_{0}}$ w.r.t. (5.47) by using the Sobol quasi-random sequence sampling plan with $m:=50$.

If we choose $m:=21$ and $\tilde{m}_{\mathrm{DSO}, \max }:=5$, then one needs $m_{\mathrm{SBO}}:=m+\tilde{m}_{\mathrm{DSO}, \max }$ highfidelity function evaluations in order to find the optimal solution

$$
\begin{equation*}
x_{\mathrm{L}}^{*}:=\left(x_{1, \min }, x_{2, \min }\right), \tag{5.48}
\end{equation*}
$$

which satisfies the statement in (5.46). Though, if we set $m_{\mathrm{DSO}, \max }:=30$ and choose, for instance, a random initial point $x^{(0)}$ that can be written as

$$
\begin{equation*}
x^{(0)}:=\left(146.25 \times 10^{-3} \mathrm{~m}, 90^{\circ}\right) \tag{5.49}
\end{equation*}
$$



FIGURE 5.6: Given the combination $\left(\mathcal{T}_{h_{1}}^{2 D},{ }^{2} \Delta_{A x=b}^{2 D}\right)$ in Table 5.3 and using the Sobol quasi-random sequence sampling plan with $m:=50$ and a kriging low-fidelity model with $\omega_{0}:=2 \pi 50 \mathrm{~Hz}$;
contour representation of $\tilde{\hat{Q}}_{L, \omega_{0}}$ in (5.47).
Dark colors indicate low values; bright colors indicate high values (cf. Figure 5.2).
then one can record that, for $m_{\mathrm{DSO}}:=m_{\mathrm{DSO}, \text { max }}$, we do not arrive at the optimal solution in (5.48). Hence, one can observe a use case such that

$$
\begin{equation*}
\exists m_{\mathrm{DSO}}, m_{\mathrm{SBO}} \in \mathbb{N} \backslash\{0\} . m_{\mathrm{DSO}}>m_{\mathrm{SBO}} \tag{5.50}
\end{equation*}
$$

holds to be true. Regarding the statement in (5.50), though, the same prudence ought to be exercised as for the statement in (5.43).

With regard to the example of application in Figure 5.5, the statement in (5.50) reflects rather the focus on a strategy for surrogate optimization than an investigation of formulations of the high-fidelity optimization problem.

The high-fidelity optimization problem in (5.47) is solely one approximate encoding of the issue concerning the inductive coupling of the two CMCs in Figure 5.5 within the mathematical framework of $\S 2.3$. The in-depth investigation of other potential encodings is left for future work.

### 5.3 In closing

Assuming the context of an electrical engineering design workflow, we have developed a strategy of using the tools from ch. 3 in practical applications. Furthermore, we have carved out some relevant spots where the tools from ch. 4 can have a beneficial impact as well.

We have elaborated on four high-fidelity optimization problems that are embedded within the setting of a 2D-LBVP and a 3D-LBVP, respectively.

From the viewpoint of $\S 2.3$, we have narrowed down our attention to optimization problems that have an evaluated reduced parametric quantity of interest in the objective functional and an evaluated reduced parametric quantity of interest in the constraints besides box constraints; and to optimization problems that have an evaluated reduced parametric quantity of interest in the objective functional and box constraints.

Moreover, given the semantics of multiple operating frequencies, we have succinctly addressed its impact on the computational burden in terms of the number of high-fidelity function evaluations and its peculiarity regarding the formulation of a corresponding optimal solution. We have also discussed that if the high-fidelity optimization problems associated with each frequency are considered individually, then, figuratively speaking, the space of options concerning surrogate optimization expands - in the sense that, e.g., a data-fit low-fidelity model's number of sampling plan points can be adaptively chosen for each frequency.

We have seen that various high-fidelity models w.r.t. the magnetoquasistatic model exhibit different behaviors such that the insights regarding the investigations, e.g., in §3.2.1 can be exploited -, for instance, to roughly identify the behavior of the high-fidelity models with the behavior of a test function from Figure 2.2.

Additionally, we have exemplified the quick use of normalized global first-order sensitivity measures as survey tools for the relevance of parameters and as potential approximate proxies for other indicators regarding the quality of a low-fidelity model.

Using the methodological guidance from §4.4.1, we have constructed some simp-lified-physics low-fidelity models and we have computed the corresponding SSPCC of these models.

We have deployed surrogate-guided optimization (see § 3.3) mainly in the context of validation and verification of given results of a surrogate-based optimization. A peculiarity is that we have utilized the co-kriging low-fidelity model in (3.184) within the TRASM algorithm 3.1 in the spirit of hybrid model management strategies (cf. § 3.4).

Mind that, however, we have carved out use cases where the number $m_{\mathrm{DSO}}$ of high-fidelity function evaluations regarding direct solving of a high-fidelity optimization problem exceeds the number $m_{\text {SBO }}$ regarding surrogate-based optimization; and we have carved out a use case where, in addition, the number $m_{\text {SBO }}$ exceeds the number $m_{\text {SGO }}$ regarding surrogate-guided optimization. Concerning these numbers, we have also addressed some caveats and limits of comparability.

## Chapter 6

## Conclusion and outlook

At the end, I distill a conclusion from § 2.4, § 3.4, § 4.6, and § 5.3, more precisely, I select a few particular insights from these sections to illustrate from the frog's-eye view, i.e., at a more technical level, some of the present work's achievements.

Moreover, I adopt a bird's-eye view to illustrate in which respect this work has made some progress in the scientific thicket of full automation of the virtual prototyping of power electronic systems (see chapter 1); and I present an outlook for potential new endeavors that may stem from this work.

### 6.1 Conclusion

The whole zoo of optimization algorithms in § 2.3.3 has proven to be useful for the purpose of finding a solution of a given optimization problem and for the purpose of cross-checking a computed optimal solution in the sense of validation and verification.

The gradient-based interpretation of sensitivity measures is well-suited for functions that permit the determination of derivative information by forward mode automatic differentiation. At the level of programs (see Figure 1.4), this interpretation is especially beneficial if there is a sound interaction between a module for automatic differentiation and a module for numerical integration to facilitate this interpretation's embedding into production-level code.

To my best knowledge, sampling plans constructed by a Sobol quasi-random sequence are not yet widely represented in the literature concerning surrogate optimization. From a theoretical viewpoint, these kinds of sampling plans enable a reproducibility of samples of a given size. From a data management viewpoint, these kinds of sampling plans enable an economical and sustainable handling of data by ensuring reusability of data. Furthermore, such sampling plans may help to lower the computational burden of constructing a co-kriging low-fidelity model.

If we assume a sparse number of sampling plan points concerning the highfidelity model, then using the squared sample Pearson correlation coefficient in combination with the empirical generalization error within the $k$-fold cross validation method requires a minimum number of sampling plan points depending on the number $k$.

By carving out a potential link between the correlation coefficient and the sensitivity measures, I have cautiously formulated a conjecture about the trustworthiness of low-fidelity models' normalized global first-order sensitivity measures. An implication of this conjecture is that, in addition to their feature as survey tools for the relevance of parameters, the sensitivity measures can serve as possible approximate proxies for other indicators regarding the quality of a low-fidelity model.

However, it is unclear whether there exists a reliable complete list of indicators. Despite this unclear point, a benchmark-focused classification of test functions has
been provided that creates the opportunity to classify very roughly the behavior of a corresponding optimization problem within the magnetoquasistatic context.

Mind that use cases have been provided where the number of high-fidelity function evaluations is higher for a direct solving of a high-fidelity optimization problem than for a surrogate-based optimization approach and for a surrogate-guided optimization approach. From an application-driven viewpoint in the context of validation and verification, I have argued that there is an additional value of checking whether the number of high-fidelity model evaluations is higher for a surrogatebased optimization approach than for a surrogate-guided optimization approach. Regarding all these numbers, however, there are some caveats and limits of comparability, too.

Driven by heuristics, a purely formalization-oriented viewpoint has been exploited that has provided us with novel insights of theoretical value (such as potential hybrid model management strategies) and of practical value (such as convergencerelated issues within the space-mapping paradigm and regarding the quality of the low-fidelity model within the co-kriging low-fidelity model).

The formalization-oriented viewpoint has culminated in the exposition of the category theory toolset which represents solely a subset of the large amount of tools available within category theory. The capability of the CT toolset as an algebraic modeling framework for applications in surrogate optimization within the electromagnetics context has been shown. More precisely, the strengths of the CT toolset as a strong notational scaffolding by diagrams of arrows have been illuminated.

In order to quantify the so-called modeling error, I have suggested a heuristicsdriven notion of a problem-dependent degree of forgetfulness as an auxiliary means. Moreover, some classification tools at the level of generalized functions (see Figure 1.4) for the concept of multifidelity model management and the space mapping notion have been propounded. Furthermore, a diagram of arrows has been proposed as a common generic interface of two formalization issues related to coordinate transformations.

From an application-oriented viewpoint, the intuition concerning the CT toolset is especially relevant in order to facilitate the CT toolset's wider acceptance. Hence, there has been an attempt to balance the need for rigor and the need for intuition. Admittedly, however, further investigations are necessary to set the applications of the category theory toolset on an even more rigorous foundation.

Finally, representatives of the class of inductive components have been invoked and I have examined four high-fidelity optimization problems that are embedded within the setting of a two-dimensional linear boundary value problem and a threedimensional linear boundary value problem, respectively.

By supposing the context of an electrical engineering design workflow, I have propounded a strategy of using the surrogate optimization tools of the present work in practical applications. Moreover, some promising spots for a beneficial utilization of the category theory toolset have been illuminated, too.

Finally, let me elucidate briefly in which respect this thesis achieves some progress concerning its ideal long-term goal, that is, the full automation of the virtual prototyping of power electronic systems. This ideal goal can be approximately conceived as the development of a user-independent software system that performs the mathematical modeling, numerical simulation and optimization given an application within the electromagnetics context.

The sheer complexity of such a goal demands expertise from numerous specialties. Thus, it is probable that any endeavor towards this goal is very interdisciplinary by nature.

This thesis has provided some indication that the category theoretical language can be a serious candidate for the important position of a mediator that enables a smooth interplay between diverse fields such as computer science, numerical analysis, and electrical engineering.

Moreover, this thesis has provided some indication that surrogate optimization can be valuable for power electronic applications at a component level as well as at a system level in the context of performance-oriented optimization and in the context of validation and verification.

Category theory's inherent emphasis on formal aspects of a given problem and its closeness to type theory in programming language theory suggest that it is a good companion in the pursuit of a user-independent software system. Furthermore, its high level of mathematical abstraction and its close relationship to logic makes it also a good companion in the pursuit of mathematical modeling, simulation, and optimization of a given application within any physics-inspired semantics.

In a nutshell, the category theoretical toolset offers some help to handle the complexity of this thesis' ideal long-term goal by providing algebraic and visual tools to convey complex formalization ideas that arise naturally in the context of, for instance, multi-fidelity modeling of surrogate optimization. Thus, the CT toolset can surely assist in the ongoing challenging search for novel surrogate-guided optimization methods for power electronic applications.

### 6.2 Outlook

I will first mention some specific potential new endeavors regarding chapter 2, chapter 3, chapter 4, and chapter 5; and then I will mention a general potential new endeavor that is associated with the Disclaimer in § 1.3.

Concerning ch. 2, it is desirable to extend the size of the subset of optimization test functions. Moreover, I also deem it advisable to include more test functions that admit a natural generalization to higher dimensions. With regard to sensitivity measures, it seems worthwhile to compare various interpretations.

Concerning ch. 3 , it might be fruitful to extend the numerical investigations regarding the optimization with test functions by data-fit low-fidelity models. For instance, more data-fit low-fidelity models could be taken into account. The thorough investigation of sequential kriging optimization and sequential co-kriging optimization with regard to the optimization problems in ch. 5 appears as an intriguing venture, too. Furthermore, the proper incorporation of many low-fidelity models in a surrogate-guided optimization approach might be a promising path as well.

Concerning ch. 4, future use cases for the CT toolset have been discussed bit more extensively in § 4.5. Undoubtedly, however, it is very preferable to seek out more examples of applications within the electromagnetics context for the category theory toolset.

Concerning ch. 5 , it is worthwhile to extend the number of parameters and the number of optimization problems. Moreover, given the semantics of multiple operating frequencies, I deem it beneficial to explore exhaustively the many possible options concerning surrogate optimization.

Finally, recalling the Disclaimer in §1.3, it might be fruitful to incorporate uncertainty quantification, parallel computing, and automation aspects all together into the context of surrogate optimization under the guidance of the CT toolset in order to develop a, for lack of a better word, robust, parallel, and automated surrogate optimization guided by category theoretical ideas.

## Appendix A

## Multivariate polynomials (§ 3.1.2)

## A. 1 Reparametrization using mean-centered arguments

With regard to fostering numerical stable computations, some authors (see, e.g., [61, p. 27]) recommend to perform a reparametrization using mean-centered arguments, i.e., the argument $x$ in (3.54) is mapped to $x-\bar{x}$ via the map $T_{\bar{x}}=x \mapsto x-\bar{x}: \mathbb{R}^{d \times 1} \rightarrow \mathbb{R}^{d \times 1}$ where, supposing a sampling plan $X_{s}$, the components of $\bar{x} \in \mathbb{R}^{d \times 1}$ are determined by the componentwise means of the sampling plan points in (3.14). The map $T_{\bar{x}}$ enables to define a map $p_{\bar{x}}:=\left(p \circ T_{\bar{x}}\right)$ that can be graphically represented as

where its assignment is encoded by

$$
\begin{equation*}
\left(p \circ T_{\bar{x}}\right)(x):=\beta_{0}+l_{e}(x-\bar{x})+q_{A}(x-\bar{x}) . \tag{A.2}
\end{equation*}
$$

Analogous to the construction of the matrix $B$ in (3.59), one can define a matrix $B_{\bar{x}}$. In Table A.1, one can observe that there is an improvement regarding the condition number $\kappa\left(B_{\bar{x}}^{\mathrm{T}} B_{\bar{x}}\right)$ w.r.t. $X_{\mathrm{s}, 1}$ and $X_{\mathrm{s}, 3}$. In the case of the sampling plan $X_{\mathrm{s}, 1}$, the absolute difference is $7.67 \times 10^{3}$ and the percental difference is $72.36 \%$. In the case of the sampling plan $X_{\mathrm{s}, 3}$, the absolute difference is $2.2 \times 10^{3}$ and the percental difference is $68.11 \%$. In the case of the sampling plan $X_{s, 3}$, the matrix $B_{\bar{x}}^{\mathrm{T}} B_{\bar{x}}$ is singular which shows that a de facto singular matrix $B$ remains singular in the course of a reparametrization of the form encoded in (A.2).

Table A.1: The condition number w.r.t. a sampling plan from Figure 3.4 without and with reparametrization in (A.2).

|  | Sampling plan | $X_{\mathbf{s}, 1}$ | $X_{\mathbf{s}, 2}$ |
| :---: | :---: | :---: | :---: |
| Condition number | $X_{\mathrm{s}, 3}$ |  |  |
| $\kappa\left(B^{\mathrm{T}} B\right)$ | $1.06 \times 10^{4}$ | $8.57 \times 10^{49}$ | $3.23 \times 10^{3}$ |
| $\kappa\left(B_{\bar{x}}^{\mathrm{T}} B_{\bar{x}}\right)$ | $2.93 \times 10^{3}$ | $\infty$ | $1.03 \times 10^{3}$ |
| $\left\|\kappa\left(B^{\mathrm{T}} B\right)-\kappa\left(B_{\bar{x}}^{\mathrm{T}} B_{\bar{x}}\right)\right\|$ | $7.67 \times 10^{3}$ | $\infty$ | $2.2 \times 10^{3}$ |
| $\frac{\left\|\left(B^{\mathrm{T}} B\right)-\kappa\left(B_{T}^{\mathrm{T}} B_{x}\right)\right\|}{\left\|\kappa\left(B^{\top} B\right)\right\|}$ | $72.36 \times 10^{-2}$ | $\infty$ | $68.11 \times 10^{-2}$ |

## A. 2 Bernstein polynomials

The reparametrization strategy in (A.1) and in (A.2), respectively, inspires to investigate in a small numerical experiment the condition number with regard to polynomials in Bernstein form. For more details on the properties of Bernstein polynomials, see, e.g., [55, p. 205-211].

The Bernstein basis $B_{\zeta} \subseteq \mathrm{P}_{\leq n}$ reads as

$$
\begin{equation*}
B_{\varsigma}:=\left\{\left.\tilde{b}_{i, n}(x) \equiv\binom{n}{i} x^{i}(1-x)^{n-i} \right\rvert\, i \in\{0,1, \ldots, n-1, n\} \wedge x \in[0,1]\right\} . \tag{A.3}
\end{equation*}
$$

In order to transform the domain from $\left[a_{l}, b_{l}\right]^{d} \equiv[0,1]^{d}$ to $\left[\tilde{a}_{l}, \tilde{b}_{l}\right]^{d}$ with $l \in\{1, \ldots, d\}$, let us apply the affine map $\gamma_{l}:\left[a_{l}, b_{l}\right] \rightarrow\left[\tilde{a}_{l}, \tilde{b}_{l}\right]$ and the affine map $v_{l}:\left[\tilde{a}_{l}, \tilde{b}_{l}\right] \rightarrow\left[a_{l}, b_{l}\right]$ such that
where $\operatorname{id}_{\left[q_{l}, b_{l}\right]}$ denotes the identity map on the domain $\left[a_{l}, b_{l}\right]$. The assignments of the affine maps $\gamma_{l}$ and $v_{l}$ are encoded by

$$
\begin{align*}
& \forall l \in\{1, \ldots, d\} \cdot \gamma\left(x_{l}\right):=\left(\left(b_{l}-a_{l}\right) x_{l}+a_{l}\right)  \tag{A.5a}\\
& \forall l \in\{1, \ldots, d\} \cdot v\left(x_{l}\right):=\frac{1}{b_{l}-a_{l}}\left(x_{l}-a_{l}\right) \tag{A.5b}
\end{align*}
$$

Assuming that the Bernstein coefficients are given in their floating-point representation, the de Casteljau's algorithm is a numerically stable tool to evaluate a polynomial $p \in \operatorname{span}\left(B_{\zeta}\right)$. A more elaborated discussion on the properties of de Casteljau's algorithm and, especially, its application in the context of Bézier curves, see, e.g., [55, p. 211-218].

In Listing A.1, I present an example implementation of de Casteljau's algorithm for the evaluation of a univariate Bernstein sum in the Julia PL.

Listing A.1: An example implementation of de Casteljau's algorithm for the evaluation of a univariate Bernstein sum in the Julia PL.

```
function bernstein_deCasteljau_eval_1d(\tilde{c}:: Vector{T}, x: :T) where T<:Real
    N = size(\tilde{c},1) # 1-based indexing
    D = zeros(N,N)
    D[1,:] = copy(\tilde{c})
        for j in 2:N
            for i in 1:N-(j-1)
                D[j,i] = (1-x)*D[j-1,i] + x*D[j-1,i+1]
            end
        end
    return D[N,1]
end
```

Generalizing to the multivariate case by employing the tensor product construction, one can provide a Bernstein basis for the space $\mathrm{P}_{k}^{d}$, however, one cannot provide a Bernstein basis for the space $P_{\leq k}^{d}$.

For a small numerical experiment, let us consider the spaces $\mathrm{P}_{\leq 2}^{2}, \mathrm{P}_{\leq 2}^{3}, \mathrm{P}_{2}^{2}$ and $\mathrm{P}_{2}^{3}$. Regarding the space $\mathrm{P}_{2}^{2}$, exemplarily, let us utilize the corresponding matrix representations of the tensor product basis as column vectors $\tilde{b} \in \mathbb{R}^{9 \times 1}$ and $\tilde{b}_{\varsigma} \in \mathbb{R}^{6 \times 1}$ by invoking the Kronecker product $\otimes$ with the signature $\mathbb{R}^{m \times n} \times \mathbb{R}^{p \times q} \rightarrow \mathbb{R}^{p m \times q n}$. Hence, $\tilde{b}$ and $\tilde{b}_{\varsigma}$ can be written as

$$
\begin{align*}
\tilde{b} & :=\left[\begin{array}{lll}
1 & x_{1} & x_{1}^{2}
\end{array}\right]^{\mathrm{T}} \otimes\left[\begin{array}{lll}
1 & x_{2} & x_{2}^{2}
\end{array}\right]^{\mathrm{T}},  \tag{A.6a}\\
\tilde{b}_{G} & :\left[\begin{array}{lll}
\left(1-x_{1}\right)^{2} & 2 x_{1}\left(1-x_{1}\right) & x_{1}^{2}
\end{array}\right]^{\mathrm{T}} \otimes\left[\begin{array}{lll}
\left(1-x_{2}\right)^{2} & 2 x_{2}\left(1-x_{2}\right) & x_{2}^{2}
\end{array}\right]^{\mathrm{T}} . \tag{A.6b}
\end{align*}
$$

Additionally, let us use sampling plans based on the Sobol quasi-random sequence in order to ensure reproducibility and to avoid averaging such as it would be needed in the case of an Audze-Eglais LHC or a Maximin LHC. In Table A.3, the condition numbers $\kappa\left(B^{\mathrm{T}} B+\lambda I\right)$ and $\kappa\left(B_{\varsigma}^{\mathrm{T}} B_{\zeta}+\lambda I\right)$ are depicted. ${ }^{1}$

In the case of setting the Tikhonov regularization parameter $\lambda$ to zero, one can observe that the condition number decreases by increasing the number of sampling plan points $m$. Adopting a statistics point of view, one can interpret this observation, intuitively, i.e., a larger sample size leads to a better estimate of the so-called true coefficients vector; or, more formally, the best coefficients column vector $\hat{\tilde{c}}$ is consistent for the true coefficients column vector $c$. For more details on this particular notion of consistency, I refer to [82, p. 65f] where the author elaborates on a theorem in which a linkage is presented between the consistency and an approximation check whether $\frac{1}{m}\left(B^{\mathrm{T}} B\right)$ approaches a symmetric positive definite matrix as $m \rightarrow \infty$.

If a symmetric positive definite matrix is given, then its trace is greater than zero; hence, let us check computationally $\operatorname{tr}\left(\frac{1}{m}\left(B^{\mathrm{T}} B\right)\right)$ and $\operatorname{tr}\left(\frac{1}{m}\left(B_{\zeta}^{\mathrm{T}} B_{\zeta}\right)\right)$, respectively. In Table A.2, the corresponding results are reported.

Table A.2: The trace of $\frac{1}{m}\left(B^{\mathrm{T}} B\right)$ and the trace of $\frac{1}{m}\left(B_{\varsigma}^{\mathrm{T}} B_{\zeta}\right)$ w.r.t. the number of sampling plan points $m$ and the Tikhonov regularization parameter $\lambda=0$ assuming the spaces $\mathrm{P}_{\leq 2}^{2}, \mathrm{P}_{\leq 2}^{3}, \mathrm{P}_{2}^{2}$ and $\mathrm{P}_{2}^{3}$ with monomial basis and the spaces $\mathrm{P}_{2}^{2}$ and $\mathrm{P}_{2}^{3}$ with Bernstein basis.

|  | $B_{\mathrm{P}_{\leq 2}^{2}}$ | $B_{\mathrm{P}_{2}^{2}}$ | $B_{c, \mathrm{P}_{2}^{2}}$ | $B_{\mathrm{P}_{\leq 2}^{3}}$ | $B_{\mathrm{P}_{2}^{3}}$ | $B_{\mathrm{c}_{5} \mathrm{P}_{2}^{3}}$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| $m=10$ | 2.131 | 2.310 | 0.267 | 2.862 | 3.328 | 0.133 |
| $m=50$ | 2.147 | 2.304 | 0.270 | 2.899 | 3.427 | 0.140 |
| $m=100$ | 2.162 | 2.334 | 0.278 | 2.909 | 3.573 | 0.145 |
| $m=1000$ | 2.176 | 2.348 | 0.283 | 2.930 | 3.600 | 0.150 |

In Table A.2, one can observe that, for all cases under consideration, the trace is non-negative and it increases very slowly with increasing number of sampling plan points. Thus, these results provide some kind of empirical evidence for the intuition underlying the notion of consistency described above. Note that the trace of a matrix is equal to the sum of its eigenvalues; and a defining property of a symmetric positive definite matrix is that is only positive definite if and only if all of its eigenvalues are positive. In Table A.2, solely the entries with regard to the combinations ( $m=10, B_{\mathrm{P}_{2}^{3}}$ ) and ( $m=10, B_{\zeta, \mathrm{P}_{2}^{3}}$ ) do not correspond to a case where all eigenvalues are positive.

[^59]Table A.3: The condition number $\kappa\left(B^{\mathrm{T}} B+\lambda I\right)$ and $\kappa\left(B_{\zeta}^{\mathrm{T}} B_{\zeta}+\lambda I\right)$ w.r.t. the number of sampling plan points $m$ and the Tikhonov regularization parameter $\lambda$ assuming the spaces $\mathrm{P}_{\leq 2}^{2}, \mathrm{P}_{\leq 2}^{3}, \mathrm{P}_{2}^{2}$ and $\mathrm{P}_{2}^{3}$ with monomial basis and the spaces $\mathrm{P}_{2}^{2}$ and $\mathrm{P}_{2}^{3}$ with Bernstein basis.
(B) The condition number assuming $\mathrm{P}_{\leq 2}^{3}$ with monomial basis.

|  | $m=10$ | $m=50$ | $m=100$ | $m=1000$ |
| :---: | :---: | :---: | :---: | :---: |
| $\lambda=0.0$ | $0.182 \times 10^{6}$ | $2.464 \times 10^{3}$ | $1.782 \times 10^{3}$ | $1.363 \times 10^{3}$ |
| $\lambda=0.2$ | $0.119 \times 10^{3}$ | $0.476 \times 10^{3}$ | $0.716 \times 10^{3}$ | $1.224 \times 10^{3}$ |
| $\lambda=0.5$ | $0.048 \times 10^{3}$ | $0.216 \times 10^{3}$ | $0.378 \times 10^{3}$ | $1.062 \times 10^{3}$ |
| $\lambda=0.8$ | $0.030 \times 10^{3}$ | $0.139 \times 10^{3}$ | $0.257 \times 10^{3}$ | $0.938 \times 10^{3}$ |


| (D) The condition number assuming $\mathrm{P}_{2}^{3}$ with monomial basis. |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $m=10$ | $m=50$ | $m=100$ | $m=1000$ |
| $\lambda=0.0$ | $7.994 \times 10^{19}$ | $6.938 \times 10^{8}$ | $3.702 \times 10^{8}$ | $1.563 \times 10^{8}$ |
| $\lambda=0.2$ | $0.130 \times 10^{3}$ | $0.668 \times 10^{3}$ | $1.388 \times 10^{3}$ | $1.395 \times 10^{4}$ |
| $\lambda=0.5$ | $0.052 \times 10^{3}$ | $0.268 \times 10^{3}$ | $0.555 \times 10^{3}$ | $5.581 \times 10^{3}$ |
| $\lambda=0.8$ | $0.033 \times 10^{3}$ | $0.167 \times 10^{3}$ | $0.347 \times 10^{3}$ | $3.488 \times 10^{3}$ |


| (F) The condition number assuming $\mathrm{P}_{2}^{3}$ with Bernstein basis. |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $m=10$ | $m=50$ | $m=100$ | $m=1000$ |
| $\lambda=0.0$ | $8.471 \times 10^{18}$ | $4.578 \times 10^{3}$ | $2.269 \times 10^{3}$ | $1.063 \times 10^{3}$ |
| $\lambda=0.2$ | $0.003 \times 10^{3}$ | $0.010 \times 10^{3}$ | $0.019 \times 10^{3}$ | $0.158 \times 10^{3}$ |
| $\lambda=0.5$ | $0.001 \times 10^{3}$ | $0.004 \times 10^{3}$ | $0.008 \times 10^{3}$ | $0.070 \times 10^{3}$ |
| $\lambda=0.8$ | $0.001 \times 10^{3}$ | $0.003 \times 10^{3}$ | $0.005 \times 10^{3}$ | $0.045 \times 10^{3}$ |


| (C) The condition number assuming $P_{2}^{2}$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | with monomial basis.

(E) The condition number assuming $\mathrm{P}_{2}^{2}$ with Bernstein basis.

|  | $m=10$ | $m=50$ | $m=100$ | $m=1000$ |
| :---: | :---: | :---: | :---: | :---: |
| $\lambda=0.0$ | $1.092 \times 10^{3}$ | $0.129 \times 10^{3}$ | $0.104 \times 10^{3}$ | $0.098 \times 10^{3}$ |
| $\lambda=0.2$ | $0.006 \times 10^{3}$ | $0.023 \times 10^{3}$ | $0.036 \times 10^{3}$ | $0.084 \times 10^{3}$ |
| $\lambda=0.5$ | $0.003 \times 10^{3}$ | $0.011 \times 10^{3}$ | $0.019 \times 10^{3}$ | $0.069 \times 10^{3}$ |
| $\lambda=0.8$ | $0.002 \times 10^{3}$ | $0.007 \times 10^{3}$ | $0.013 \times 10^{3}$ | $0.058 \times 10^{3}$ |

Given a fixed $\lambda>0$ in Table A.3, one can observe that the condition number is increasing as the number of sampling plan points $m$ is increasing. I cannot back up this particular behavior in a formal way. Given a fixed $m$, one can observe that, the condition number is decreasing as the regularization parameter is increasing. This behavior is matching the expectation regarding the regularization parameter.

If we exclude pathological several orders of magnitude from consideration, then, in practical application, the assessment of a condition number as an indication for an ill-conditioned problem is, mostly, incumbent upon the judgment of the user and its desired accuracy for a problem under investigation.

Therefore, the key empirical insight from Table A. 3 is that for a low number of sampling plan points, by invoking the regularization parameter $\lambda$, one can achieve moderately small condition numbers by monomial basis for the space $\mathrm{P}_{\leq 2}^{2}$ and $\mathrm{P}_{\leq 2}^{3}$. These condition numbers are comparable to those condition numbers associated with their tensor product polynomial basis counterparts.

## A. 3 Chebyshev polynomials

Another possibility to mitigate by design the multicollinearity with respect to the chosen basis is, instead of a monomial basis $B \subseteq \mathrm{P}_{\leq n}$ such as in (3.40), to apply a finite basis of orthogonal polynomials $Q \subseteq \mathrm{P}_{\leq n}$ in which the polynomials are pairwise orthogonal with respect to an inner product $\langle\cdot, \cdot\rangle_{\mathbb{R}^{x}}$ that can be considered in a continuous form or in a discrete form (see, e.g., [48, p. 251f]), more precisely,

$$
\begin{align*}
& \langle\cdot, \cdot\rangle_{\mathbb{R}^{X}}:=(f, g) \mapsto \int_{a}^{b} f(x) g(x) w(x) \mathrm{d} x: \mathbb{R}^{X} \times \mathbb{R}^{X} \rightarrow \mathbb{R} \quad \text { or }  \tag{A.7a}\\
& \langle\cdot \cdot \cdot\rangle_{\mathbb{R}^{X_{s}}}:=(f, g) \mapsto \sum_{i=1}^{m} f\left(x_{i}\right) g\left(x_{i}\right) w\left(x_{i}\right): \mathbb{R}^{X_{s}} \times \mathbb{R}^{X_{\mathbf{s}}} \rightarrow \mathbb{R}, \tag{A.7b}
\end{align*}
$$

where $x \in X:=[a, b] \subset \mathbb{R}, x_{i} \in X_{\mathbf{s}} \subseteq X^{m}$ with $m \in \mathbb{N}$ and $i \in\{1,2, \ldots, m\}$, and $w$ denotes a weight function. Let us focus on the discrete form in (A.7b).

Using orthogonal polynomials $q_{i} \in Q$, one can concisely define a matrix $Q \in \mathbb{R}^{m} \times \mathbb{R}^{s}$ with respect to the sampling plan points $x_{i}$ :

$$
Q:=\left[\begin{array}{c}
\tilde{q}_{x_{1}}^{\mathrm{T}}  \tag{A.8}\\
\tilde{q}_{x_{2}}^{\mathrm{T}} \\
\vdots \\
\tilde{q}_{x_{m}}^{\mathrm{T}}
\end{array}\right],
$$

such that, in the absence of the floating point approximation error, $Q^{\mathrm{T}} Q$ is exactly diagonal in the sense of

$$
\begin{equation*}
Q^{T} Q \equiv \operatorname{diag}\left(\left\langle\tilde{q}_{0}, \tilde{q}_{0}\right\rangle_{\mathbb{R}^{X_{s}}}, \ldots,\left\langle\tilde{q}_{s-1}, \tilde{q}_{s-1}\right\rangle_{\mathbb{R}^{X_{s}}}\right), \tag{A.9}
\end{equation*}
$$

where $\tilde{q}_{i}$ denote the components of a basis vector $\tilde{q} \in \mathbb{R}^{s \times 1}$. If $Q$ is orthonormal, then $Q^{\mathrm{T}} Q=I$ where $I \in \mathbb{R}^{s} \times \mathbb{R}^{s}$ denotes the identity matrix.

Let us aim our attention at univariate Chebyshev polynomials of the first kind $T_{n}(x)$ with the leading coefficient $2^{n-1}$ for $n \geq 1$ and the domain $X:=[-1,1]$. For more details on the properties of those sequences of orthogonal polynomials called Chebyshev polynomials, see, e.g., [208].

The corresponding recurrence formula reads as

$$
T_{n}(x):= \begin{cases}1, & \text { if } n \equiv 0,  \tag{A.10}\\ x, & \text { if } n \equiv 1, \\ 2 x T_{n-1}(x)-T_{n-2}(x), & \text { if } n \geq 2 .\end{cases}
$$

Based on the recurrence formula in (A.10), one can construct a basis for the space $P_{\leq 2}$; furthermore, one can construct a basis for the space $\mathrm{P}_{<2}^{d}$ in the same manner as in (3.45). Hence, one can define the column vector $\tilde{q} \in \mathbb{R}^{s \times 1^{〔 2}}$ that represents the basis of a $d$-variate Chebyshev polynomial of the first kind of degree at most two:

$$
\tilde{q}:=\left[\begin{array}{llllllllll}
\tilde{q}_{0} & \tilde{q}_{1} & \ldots & \tilde{q}_{d} & \tilde{q}_{d+1} & \ldots & \tilde{q}_{2 d} & \tilde{q}_{2 d+1} & \ldots & \tilde{q}_{s-1} \tag{A.11}
\end{array}\right]^{\mathrm{T}},
$$

where $\tilde{q}_{0}=T_{0}\left(x_{1}\right) \cdots T_{0}\left(x_{d}\right), \tilde{q}_{1}=T_{1}\left(x_{1}\right), \tilde{q}_{d}=T_{1}\left(x_{d}\right), \tilde{q}_{d+1}=T_{2}\left(x_{1}\right), \tilde{q}_{2 d}=T_{2}\left(x_{d}\right)$, $\tilde{q}_{2 d+1}=T_{1}\left(x_{1}\right) T_{1}\left(x_{2}\right)$, and $\tilde{q}_{s-1}=T_{1}\left(x_{d-1}\right) T_{1}\left(x_{d}\right)$; more specifically, $\tilde{q}_{0}=1, \tilde{q}_{1}=x_{1}$, $\tilde{q}_{d}=x_{d}, \tilde{q}_{d+1}=2 x_{1}^{2}-1, \tilde{q}_{2 d}=2 x_{d}^{2}-1, \tilde{q}_{2 d+1}=x_{1} x_{2}$, and $\tilde{q}_{s-1}=x_{d-1} x_{d}$. Thus, analogously to (3.57), one can define a polynomial as

$$
\begin{equation*}
p(x):=\tilde{q}^{\mathrm{T}} \tilde{c} . \tag{A.12}
\end{equation*}
$$

Mind that the sampling plan $X_{\mathbf{s}} \subset\left([-1,1]^{d}\right)^{m_{1} \times \cdots \times m_{d}}$ is constituted by the Chebyshev nodes

$$
\begin{equation*}
\forall l \in\{1, \ldots, d\} . \forall k \in\left\{1, \ldots, m_{l}\right\} \cdot x_{k l}:=\cos \left(\frac{2 k-1}{2 m} \pi\right), \tag{A.13}
\end{equation*}
$$

where $\prod_{l=1}^{d} m_{l}:=m$. Let us refer to this specific construction of a sampling plan $X_{\mathbf{s}}$ as a Chebyshev grid in which the sampling plan $X_{s}$ can be conceived as a tensor in the sense of a multidimensional array (see, e.g., [79]). For a brief and succinct comment on the smoothness requirements for a high-fidelity model and on the influence of the positioning of the nodes in polynomial interpolation (the Runge phenomenon, the Faber theorem), see, e.g., [209].

In Figure A.1, three Chebyshev grids with $d:=2$ are illustrated.


Figure A.1: Instances of a Chebyshev grid as a sampling plan $X_{\mathbf{s}} \subset\left(X^{2}\right)^{m_{1} \times m_{2}}$.
(i) $X:=[-1,1], m_{1} \equiv m_{2}:=5$, (ii) $X:=[0,1], m_{1} \equiv m_{2}:=5$, (iii) $X:=[0,1], m_{1} \equiv m_{2}:=10$.

In order to transform the domains from $\left[a_{l}, b_{l}\right]^{2} \equiv[-1,1]^{2}$ to $\left[\tilde{a}_{l}, \tilde{b}_{l}\right]^{2} \equiv[0,1]^{2}$ with $l \in\{1,2\}$, let us overload the maps and override the domains and the variables in (A.4) and in (A.5). Thus, let us apply the affine map $\gamma_{l}:\left[a_{l}, b_{l}\right] \rightarrow\left[\tilde{a}_{l}, \tilde{b}_{l}\right]$ and the
affine map $v_{l}:\left[\tilde{a}_{l}, \tilde{b}_{l}\right] \rightarrow\left[a_{l}, b_{l}\right]$ such that
where $\mathrm{id}_{\left[a_{l}, b_{l}\right]}$ denotes the identity map on the domain $\left[a_{l}, b_{l}\right]$. Invoking (A.13), the assignments of the affine maps $\gamma_{l}$ and $v_{l}$ are encoded by

$$
\begin{align*}
& \forall l \in\{1, \ldots, d\} . \forall k \in\left\{1, \ldots, m_{l}\right\} \cdot \gamma\left(x_{k l}\right):=\frac{1}{2}\left(\left(b_{l}-a_{l}\right) x_{k l}+\left(a_{l}+b_{l}\right)\right),  \tag{A.15a}\\
& \forall l \in\{1, \ldots, d\} . \forall k \in\left\{1, \ldots, m_{l}\right\} \cdot v\left(x_{k l}\right):=\frac{1}{b_{l}-a_{l}}\left(2 x_{k l}-\left(a_{l}+b_{l}\right)\right) . \tag{A.15b}
\end{align*}
$$

If we use the orthogonality condition in the discrete form for the univariate case, i.e.,

$$
\forall l \in\{1, \ldots, d\} \cdot \sum_{k=1}^{m_{l}} T_{i}\left(x_{k l}\right) T_{j}\left(x_{k l}\right):= \begin{cases}0, & \text { if } i \neq j,  \tag{A.16}\\ m_{l}, & \text { if } i \equiv j \wedge i=0 \\ \frac{m_{l}}{2}, & \text { if } i \equiv j \wedge i \neq 0\end{cases}
$$

one can express the orthogonality condition in the discrete form for the multivariate case in terms of the components $\tilde{q}_{i}$ in (A.11), i.e.,

$$
\left\langle\tilde{q}_{i}, \tilde{q}_{j}\right\rangle_{\mathbb{R}^{X_{\mathbf{s}}}}:= \begin{cases}0, & \text { if } i \neq j  \tag{A.17}\\ m, & \text { if } i \equiv j \wedge i=0 \\ \frac{m}{2}, & \text { if } i \equiv j \wedge i \in[1,2 d] \\ \frac{m}{4}, & \text { if } i \equiv j \wedge i \in[2 d+1, s-1]\end{cases}
$$

Analogously to the Listing 3.1, if we have the optimal coefficients as floatingpoint numbers, then a numerically stable approach to evaluate the function in (A.12) is Clenshaw's recurrence formula.

In Listing A.2, I present an example implementation of Clenshaw's algorithm for the evaluation of a univariate Chebyshev sum in the Julia PL. For a neat graph representation of the structure of Clenshaw's algorithm, see, e.g., [78, p. 199].

To avoid clutter, a modified version of this algorithm (see [75, p. 78f]) is left out. The modified version attempts to alleviate the rounding errors' hazard in the cases in which the argument is close to the domain's boundary, e.g., the case $x:=-1+1 \times 10^{3} \epsilon$ or the case $x:=1-1 \times 10^{3} \epsilon$ where the number $\epsilon$ encodes the machine epsilon in floating point arithmetic.

Listing A.2: An example implementation of Clenshaw's algorithm for the evaluation of a univariate Chebyshev sum in the Julia PL.

```
function chebyshev_clenshaw_eval_1d(\tilde{c}:: Vector{T},x::T) where {T<:Real}
    N = size(\tilde{c},1) - 1 # 1-based indexing
    d = zeros(N+2)
    d[N+2]=0
    d[N+1] = \tilde{c}[N+1]
        for i in N:-1:2
            d[i] = 2x*d[i+1] - d[i+2] + \tilde{c}[i]
        end
    return x*d[2] - d[3] + \tilde{c}[1]
end
```

Concerning the multivariate case, if we possess the space $P_{\leq 2}^{2}$ with $\operatorname{dim}\left(P_{\leq 2}^{2}\right) \equiv 6$, then one can introduce the column vectors $\tilde{\tau}\left(x_{2}\right) \in \mathbb{R}^{6 \times 1}$ and $\tilde{\eta}\left(x_{1}\right) \in \mathbb{R}^{6 \times 1}$ and the diagonal matrix $\tilde{\Delta} \in \mathbb{R}^{6} \times \mathbb{R}^{6}$ such that

$$
\begin{align*}
& \tilde{\tau}\left(x_{2}\right):=\left[\begin{array}{lllll}
T_{0}\left(x_{2}\right) & T_{0}\left(x_{2}\right) & T_{0}\left(x_{2}\right) & T_{1}\left(x_{2}\right) & T_{2}\left(x_{2}\right)
\end{array} T_{1}\left(x_{2}\right)\right]^{\mathrm{T}},  \tag{A.18a}\\
& \tilde{\eta}\left(x_{1}\right):=\left[\begin{array}{lllll}
T_{0}\left(x_{1}\right) & T_{1}\left(x_{1}\right) & T_{2}\left(x_{1}\right) & T_{0}\left(x_{1}\right) & T_{0}\left(x_{1}\right)
\end{array} T_{1}\left(x_{1}\right)\right]^{\mathrm{T}},  \tag{A.18b}\\
& \tilde{\Delta}:=\operatorname{diag}\left(\tilde{c}_{1}, \tilde{c}_{2}, \tilde{c}_{3}, \tilde{c}_{4}, \tilde{c}_{5}, \tilde{c}_{6}\right),  \tag{A.18c}\\
& p(x):=\tilde{\tau}\left(x_{2}\right)^{\mathrm{T}} \tilde{\Delta} \tilde{\eta}\left(x_{1}\right) . \tag{A.18d}
\end{align*}
$$

Finally, similarly to (3.69), one can apply an evaluation scheme based on multiple nested hierarchical evaluations of univariate Chebyshev sums to the formulation

$$
\begin{equation*}
p(x):=\sum_{i=1}^{\operatorname{dim}\left(\mathrm{P}_{\leq 2}^{2}\right)} \tilde{c}_{j} \tilde{\tau}_{j}\left(x_{2}\right) \tilde{\eta}_{j}\left(x_{1}\right) . \tag{A.19}
\end{equation*}
$$

For more details on evaluation algorithms for multivariate orthogonal polynomials such as Chebyshev polynomials, see, e.g., [17].

Noteworthily, the authors in [206] allude to a numerically stable evaluation scheme for a univariate Chebyshev sum based on the second (true) form of the barycentric formula in rational interpolation where the number of Chebyshev nodes of the second kind has to be sufficiently large (cf. [22]). This evaluation scheme only relies on the information about the nodes and the corresponding values of the highfidelity model. Though, since the basic presumption in the present work is to keep the number of evaluation points fairly low, this formula is solely employed for testing purposes.

Regarding the usage of a deterministic data-fit low-fidelity model in an interpolation context, an additional aspect that is restricted by the basic presumption in the present work is the reduction of the empirical surrogate modeling error (see Definition 3.1.2) by increasing sufficiently the sample size and by positioning appropriately the sampling plan points. Some kind of adaptive interpolation in the sense that an initial sampling plan is extended sequentially in a controlled manner, though, it is solely regarded for the probabilistic data-fit low-fidelity model, that is, the kriging low-fidelity model.

Regarding the condition number $\kappa\left(Q^{\mathrm{T}} Q\right)$, if $Q$ is orthonormal, then, by design, it holds that $\kappa\left(Q^{\mathrm{T}} Q\right) \equiv 1$; and if $Q$ is only orthogonal, then, in principle, it holds that $\kappa\left(Q^{\mathrm{T}} Q\right)<9$. Comparing these observations to the observations in Table A.3, it can be argued that a change of basis, especially, a change to a basis of Chebyshev polynomials, can have a favorable effect on the condition number. Mind that, from an application-oriented view, the regularization approach offers more flexibility concerning the choice of a basis and the choice of a sampling plan.

## Appendix B

## Solenoid with a core (§ 5.1)

## B. 1 An electrical network viewpoint

In Figure B.1, I depict the circuit diagram representation of the three fundamental passive electrical components where the circles encode external terminal nodes of the passive electrical components.


Figure B.1: Circuit diagram representation of the three fundamental passive electrical components: resistance $R$, inductance $L$, and capacitance $C$.

If we use the map $\underline{U} \in \mathbb{C}^{\mathbb{R}^{+}}$and the map $\underline{I} \in \mathbb{C}^{\mathbb{R}^{+}}$to denote a complex-valued voltage drop and a complex-valued current intensity, respectively, where both maps depend on the angular frequency $\omega$, then, with respect to the passive electrical components in Figure B.1, one can state the following equations to hold to be true:

$$
\begin{equation*}
\underline{U}_{R}(\omega)=R \cdot \underline{I}_{R}(\omega), \quad \underline{U}_{L}(\omega)=j \omega L \cdot \underline{I}_{L}(\omega), \quad \underline{I}_{C}(\omega)=j \omega C \cdot \underline{U}_{C}(\omega), \tag{B.1}
\end{equation*}
$$

where it is assumed that the entities $R, L, C \in \mathbb{C}$ with $\operatorname{Re}(R), \operatorname{Re}(L), \operatorname{Re}(C) \in \mathbb{R}^{+}$and $\operatorname{Im}(R):=0, \operatorname{Im}(L):=0$, and $\operatorname{Im}(C):=0$. Hence, all the multiplication maps in (B.1) have the same signature $\mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$ and the corresponding assignment rules refer to the common algebraic rules for complex numbers.

In Figure B.2, two representatives from the class of circuit diagrams for real inductive components (cf. [112, p. 520ff]) are depicted where a rough colloquial equivalence relation can be defined by "has the same four-tuple of fundamental passive electrical components ( $L_{0}, R_{\mathrm{w}}, R_{\mathrm{c}}, C_{\mathrm{p}}$ ) as".

Mind that in Figure B.2, due to the exposition in § 2.1.3, one can solely consider the resistance $R_{\mathrm{W}}$ (associated with the losses in the winding), the resistance $R_{\mathrm{c}}$ (associated with the losses in the core) and the inductance $L_{0}$ (associated with the magnetic energy) within the magnetoquasistatic model where, in practical applications, the entity $L_{0}$ refers to the nominal inductance provided by a choke manufacturer's data sheet.

In the context of electromagnetic compatibility, though, it is common to take the capacitance $C_{\mathrm{p}}$ as a parasitic component into account, too, in order to reconstruct properly a real inductive component's impedance map $\underline{Z} \in \mathbb{C}^{\mathbb{R}^{+}}$with $\underline{Z}=\omega \mapsto \frac{U(\omega)}{\frac{U}{I}(\omega)}$ over a wide range of frequencies. Notice well that, in contrast to the $2 \mathrm{D}-\mathrm{L} \overline{\mathrm{B} V P}$ in $\S 5.1$, the aim for proper reconstruction of a real inductive component's impedance map prompts one to also incorporate the resistance $R_{\mathrm{c}}$. The assignment rules for the


Figure B.2: Two representatives from the equivalence class of circuit diagrams for real inductive components with the equivalence relation "has the same four-tuple of fundamental passive electrical components $\left(L_{0}, R_{\mathrm{W}}, R_{\mathrm{c}}, C_{\mathrm{p}}\right)$ as".
representatives in Figure B. 2 read as

$$
\begin{equation*}
\underline{Z}_{1}(\omega):=\frac{\left(\left(R_{\mathrm{c}}+R_{\mathrm{w}}\right)+j \omega L_{0}\right) \frac{1}{j \omega C_{\mathrm{p}}}}{\left(R_{\mathrm{w}}+R_{\mathrm{c}}\right)+j \omega L_{0}+\frac{1}{j \omega C_{\mathrm{p}}}}, \quad \underline{Z}_{2}(\omega):=\frac{1}{\frac{1}{R_{\mathrm{w}}+j \omega L_{0}}+\frac{1}{R_{\mathrm{c}}}+j \omega C_{\mathrm{p}}}, \tag{B.2}
\end{equation*}
$$

where $\underline{Z}_{1}(\omega)$ corresponds to the circuit diagram in Figure B.2a and $\underline{Z}_{2}(\omega)$ corresponds to the circuit diagram in Figure B.2b.

If we overload the impedance map $\underline{Z}$ (and the maps $\underline{U}$ and $\underline{I}$ as well) in the sense that $\underline{Z} \in \mathbb{C}^{C}$ with $\underline{Z}=s \mapsto \frac{U(s)}{I(s)}$, then one can rewrite the assignment rules in (B.2) by substituting the term $j \omega$ with the term $s$ - which we conceive as $s:=\sigma_{s}+j \omega_{s}$ with $\sigma_{s}, \omega_{s} \in \mathbb{R}$ - such that
$\underline{Z}_{1}(s):=\frac{\left(R_{\mathrm{c}}+R_{\mathrm{w}}\right)+L_{0} s}{1+\left(R_{\mathrm{w}}+R_{\mathrm{c}}\right) C_{\mathrm{p}} s+C_{\mathrm{p}} L_{0} s^{2}}, \quad \underline{Z}_{2}(\omega):=\frac{R_{\mathrm{c}} R_{\mathrm{W}}+R_{\mathrm{c}} L_{0} s}{R_{\mathrm{c}}+R_{\mathrm{W}}+\left(L_{0}+C_{\mathrm{p}} R_{\mathrm{c}} R_{\mathrm{w}}\right) s+C_{\mathrm{p}} R_{\mathrm{c}} L_{0} s^{2}}$,

$$
\begin{array}{cc}
\underline{Z}_{1}(s):=\frac{N_{1}(s)}{\underline{D}_{1}(s)}, & \underline{Z}_{2}(s):=\frac{N_{2}(s)}{\underline{D}_{2}(s)}, \\
\underline{Z}_{1}(s):=k_{1} \frac{s-z_{1_{1}}}{\left(s-p_{1_{1}}\right)\left(s-p_{1_{2}}\right)}, & \underline{Z}_{2}(s):=k_{2} \frac{s-z_{2_{1}}}{\left(s-p_{2_{1}}\right)\left(s-p_{2_{2}}\right)}, \tag{B.3c}
\end{array}
$$

where $\underline{N}_{i} \in \mathbb{C}^{\mathrm{C}}$ with $i \in\{1,2\}$ denotes the complex numerator polynomial with real coefficients of the respective impedance map - and the map $\underline{D}_{i} \in \mathbb{C}^{\mathrm{C}}$ denotes the corresponding complex denominator polynomial with real coefficients. The Julia PL package SymPy.jl (see [110]) is utilized in order to perform symbolic computations with regard to (B.2) and (B.3).

In (B.3), the non-negative numbers $k_{1}$ and $k_{2}$ are defined as $k_{1}:=\frac{L}{L C_{\mathrm{p}}}$ and $k_{2}:=\frac{L R_{\mathrm{c}}}{L C_{\mathrm{p}} R_{\mathrm{p}}}$. Moreover, $z_{i_{1}} \in \mathbb{C}$ with $i \in\{1,2\}$ refers to the zero of the respective impedance map and $p_{i_{j}} \in \mathbb{C}$ with $i \in\{1,2\}$ and $j \in\{1,2\}$ refers to the pole of the respective impedance map. If the poles have non-zero imaginary parts, i.e., $\forall i . \forall j \operatorname{Im}\left(p_{i_{j}}\right) \neq 0$, then one can define the resonance frequency $f_{r_{i}} \in \mathbb{R}^{+}$for each $i$ as $f_{r_{i}}:=\left|\frac{\operatorname{Im}\left(p_{i_{1}}\right)}{2 \pi}\right|$.

A useful approximation of the resonance frequency is given by $f_{r_{i}}:=\frac{1}{2 \pi} \frac{1}{\sqrt{C_{\mathrm{p}} L}}$. Observe that we limit our consideration of a real inductive component's impedance $\operatorname{map} \underline{Z}$ to the case of one resonance frequency.

In Figure B.3, I illustrate the magnitude (or modulus) $Z(\omega)$ and the phase (or argument) $\theta(\omega)$ of the impedances in (B.2) associated with the representatives in Figure B. 2 for synthetic data w.r.t. the four-tuple ( $L_{0}, R_{\mathrm{w}}, R_{\mathrm{c}}, \mathrm{C}_{\mathrm{p}}$ ) and the frequency range $\left[1 \times 10^{2} \mathrm{~Hz}, 1 \times 10^{8} \mathrm{~Hz}\right]$. ${ }^{1}$


Figure B.3: Given the frequency range $\left[1 \times 10^{2} \mathrm{~Hz}, 1 \times 10^{8} \mathrm{~Hz}\right]$, the magnitude $Z(\omega)$ and the phase $\theta(\omega)$ of the impedances in (B.2) associated with the representatives in Figure B. 2 for synthetic data w.r.t. the four-tuple $\left(L_{0}, R_{W}, R_{c}, C_{\mathrm{p}}\right)$.

[^60]In order to move from the field theoretical level in (2.1) to the circuit theoretical level in (B.1), let us invoke Poynting's theorem (see, e.g., [139, p. 108ff]) for the frequency domain such that one can determine the three fundamental passive electrical components by the following identifications:

$$
\begin{array}{rlr}
\bar{P}_{\mathrm{L}} \equiv \int_{V} \frac{1}{2} \mathbf{J}_{\mathrm{cond}} \cdot \overline{\mathbf{E}} \mathrm{~d} V, & \bar{W}_{\mathrm{m}} \equiv \int_{V} \frac{1}{4} \mathbf{B} \cdot \overline{\mathbf{H}} \mathrm{~d} V, & \bar{W}_{\mathrm{e}} \equiv \int_{V} \frac{1}{4} \mathbf{E} \cdot \overline{\mathbf{D}} \mathrm{~d} V, \\
R \equiv \frac{\bar{P}_{\mathrm{L}}}{I_{\mathrm{rms}}^{2}}, & L \equiv \frac{2 \bar{W}_{\mathrm{m}}}{I_{\mathrm{rms}}^{2}}, & C \equiv \frac{2 \bar{W}_{\mathrm{e}}}{I_{\mathrm{rms}}^{2}}, \tag{B.5}
\end{array}
$$

where $\bar{P}_{\mathrm{L}} \in \mathbb{R}^{+}$denotes the time-averaged ohmic loss in $\Omega^{2 D}, \bar{W}_{\mathrm{m}} \in \mathbb{R}^{+}$denotes the time-averaged magnetic energy in $\Omega^{2 D}$, and $\bar{W}_{\mathrm{e}} \in \mathbb{R}^{+}$denotes the time-averaged electric energy in $\Omega^{2 D}$. In accordance with the elaborations in $\S 2.2 .1$, one can conceive the 3-tuple ( $\bar{P}_{\mathrm{L}}, \bar{W}_{\mathrm{m}}, \bar{W}_{\mathrm{e}}$ ) and the 3-tuple ( $R, L, C$ ) as 3-tuples of evaluated quantities of interest.

Notice that $\bar{W}_{\mathrm{e}}$ and $C$ are excluded in the considerations with regard to the 2DLBVP in § 5.1. Expressing it in terms of the circuit diagram representative in Figure B.2a, one can set $R_{\mathrm{c}} \equiv 0 \mathrm{~m} \Omega, C_{\mathrm{p}} \equiv 0 \mathrm{pF}, R_{\mathrm{w}} \equiv R$, and $L_{0} \equiv L$ such that the circuit diagram representative reduces to a series connection of the impedances associated with the resistance $R_{w}$ and the inductance $L_{0}$.

Furthermore, if we consider the map $\underline{U} \in \mathbb{C}^{\mathbb{R}^{+}}$and the map $\underline{I} \in \mathbb{C}^{\mathbb{R}^{+}}$with respect to the external terminal nodes, then one can express the resistance $R$ and the inductance $L$ as

$$
\begin{equation*}
R \equiv \operatorname{Re}\left(\frac{\underline{U}(\omega)}{\underline{I}(\omega)}\right), \quad L \equiv \operatorname{Re}\left(\frac{\underline{\Psi}(\omega)}{\underline{I}(\omega)}\right) \tag{B.6}
\end{equation*}
$$

where the map $\underline{\Psi} \in \mathbb{C}^{\mathbb{R}^{+}}$denotes the complex-valued total magnetic flux. By invoking the definition (ii) in (2.5) and applying the theorem of Stokes, the assignment rule of the map $\underline{\Psi}$ can be stated as

$$
\begin{equation*}
\underline{\Psi}=\omega \mapsto \int_{\partial A} \mathbf{A} \cdot \mathrm{~d} \mathbf{s} . \tag{B.7}
\end{equation*}
$$

Notice well that due to the numerical integration that is involved in determining the entities such as $\bar{W}_{\mathrm{m}}$ in (B.4) or such as $\underline{\Psi}(\omega)$ in (B.6), there might be slight differences in the decimal places with regard to the resistance $R$ and the inductance $L$ depending on the method of computation - even if we assume that the domain of integration is properly chosen.

## B. 2 A visualization of evaluated data-fit low-fidelity models regarding (5.12), (5.14c), and (5.15)

(1a)

(1b)

(2a)

(2b)


(A) Representation of $z:=\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$.
(1a)
(2a)
(3a)

(1b)





(в) Representation of $z:=\tilde{V}_{\mathrm{ut}}\left(x_{1}, x_{2}\right)$.

Figure B.4: By using the Sobol quasi-random sequence sampling plan with $m:=21$ and the data-fit low-fidelity models in $\S$ 3.2.1, i.e., (1) Polynomial, (2) TPS RBF, and (3) Kriging; representations (surface (a) and contour (b)) of $\tilde{\dot{j}}_{\bar{P}_{\mathrm{L}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ and $\tilde{V}_{\mathrm{ut}}\left(x_{1}, x_{2}\right)$
where $\omega_{0}:=2 \pi 100 \mathrm{kHz}$.


Figure B.5: By using the Sobol quasi-random sequence sampling plan with $m:=21$ and the data-fit low-fidelity models in $\S$ 3.2.1, i.e., (1) Polynomial, (2) TPS RBF, and (3) Kriging; representations (surface (a) and contour (b)) of $\tilde{\hat{j}}_{\bar{P}_{\mathrm{L}}, V_{\mathrm{ut}}, \omega_{0}}\left(x_{1}, x_{2}\right)$ and $\tilde{\hat{Q}}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)$ where $\omega_{0}:=2 \pi 100 \mathrm{kHz}$.

## B.2. A visualization of evaluated data-fit low-fidelity models regarding (5.12),


(A) Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ w.r.t. Figure B.4a.

(B) Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ w.r.t. Figure B. 4 b .

Figure B.6: Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ as a projection on the contour representation of the data-fit low-fidelity models for the functions in

(A) Depicting $\operatorname{grad}(\tilde{K})\left(x_{1}, x_{2}\right)$ w.r.t. Figure B.5a.

(в) Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ w.r.t. Figure B.5b.

Figure B.7: Depicting $\operatorname{grad}(\tilde{\mathrm{K}})\left(x_{1}, x_{2}\right)$ as a projection on the contour representation of the data-fit low-fidelity models for the functions in Figure B.5.

## Bibliography

[1] S. Abramsky and N. Tzevelekos. "Introduction to Categories and Categorical Logic". In: New Structures for Physics. Ed. by B. Coecke. Vol. 813. Lecture Notes in Physics. Springer, 2011, pp. 3-94.
[2] N. M. Alexandrov and M. Y. Hussaini (editors). Multidisciplinary Design Optimization: State of the Art. SIAM, 1997.
[3] N. M. Alexandrov and R. M. Lewis. "First-Order Approximation and Model Management in Optimization". In: Large-Scale PDE-Constrained Optimization. Ed. by L.T. Biegler, M. Heinkenschloss, O. Ghattas, and B. van Bloemen Waanders. Vol. 30. Lecture Notes in Computational Science and Engineering. Springer, 2003, pp. 63-79.
[4] N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. "Approximation and model management in aerodynamic optimization with variable-fidelity models". In: Journal of Aircraft 38.6 (2001), pp. 10931101.
[5] S. Arlot and A. Celisse. "A survey of cross-validation procedures for model selection". In: Statistics Surveys 4 (2010), pp. 40-79.
[6] D. N. Arnold. Finite Element Exterior Calculus. SIAM, 2018.
[7] D. N. Arnold. "Stability, Consistency, and Convergence of Numerical Discretizations". In: Encyclopedia of Applied and Computational Mathematics. Ed. by B. Engquist. Springer, 2015.
[8] D. N. Arnold, R. S. Falk, and R. Winther. "Finite element exterior calculus, homological techniques, and applications". In: Acta Numerica 15 (2006), pp. 1155.
[9] K. Atkinson and W. Han. Theoretical Numerical Analysis: A Functional Analysis Framework. 3rd ed. Springer, 2009.
[10] C. Audet and M. Kokkolaras (editors). "Special Issue on Blackbox and Deri-vative-Free Optimization". In: Optimization and Engineering 17.1 (2016), pp. 1262.
[11] S. Awodey. Category Theory. Oxford University Press, 2nd ed., 2009.
[12] I. Babuška and J. T. Oden. "Verification and validation in computational engineering and science: basic concepts". In: Comput. Methods Appl. Mech. Eng. 139.36 (2004), pp. 4057-4066.
[13] J. C. Baez and B. Fong. "A Compositional Framework for Passive Linear Networks". In: Theory and Applications of Categories 33.38 (2018), pp. 1158-1222.
[14] J. W. Bandler, R. M. Biernacki, S. H. Chen, R. H. Hemmers, and K. Madsen. "Electromagnetic optimization exploiting aggressive space mapping". In: IEEE Trans. Microwave Theory Tech. 43.12 (1995), pp. 2874-2882.
[15] H. P. Barendregt. The Lambda Calculus - Its Syntax and Semantics. revised ed. North-Holland, 1985.
[16] M. Barr and C. Wells. "Category Theory for Computing Science". In: Reprints in Theory and Applications of Categories 22 (2012), pp. 1-538.
[17] R. Barrio, J. M. Peña, and T. Sauer. "Three term recurrence for the evaluation of multivariate orthogonal polynomials". In: Journal of Approximation Theory 162.2 (2010), pp. 407-420.
[18] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. Nonlinear Programming: Theory and Algorithms. 3rd ed. Wiley, 2006.
[19] R. K. Beatson, M. J. D. Powell, and A. M. Tan. "Fast evaluation of polyharmonic splines in three dimensions". In: IMA Journal of Numerical Analysis 27.3 (2007), pp. 427-450.
[20] P. Benner, S. Gugercin, and K. Willcox. "A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems". In: SIAM Review 57.4 (2015), pp. 483-531.
[21] P. Benner, M. Hinze, and E. J. W. ter Marten (editors). Model Reduction for Circuit Simulation. 1st ed. Springer, 2011.
[22] J.-P. Berrut and L. N. Trefethen. "Barycentric Lagrange Interpolation". In: SIAM Review 46.3 (2004), pp. 501-517.
[23] L. Bessi. Surrogates in Julia. https://github.com/SciML/Surrogates.j1. [Online; accessed 12-February-2020]. 2020.
[24] H.-G. Beyer and H.-P. Schwefel. "Evolution strategies: A comprehensive introduction". In: Natural Computing 1 (2002), pp. 3-52.
[25] J. Bezanson, J. Chen, B. Chung, S. Karpinski, V. B. Shah, J. Vitek, and L. Zoubritzky. "Julia: Dynamism and Performance Reconciled by Design". In: Proc. ACM Program. Lang. 2.OOPSLA (2018), pp. 1-23.
[26] J. Bezanson, J. Edelman, S. Karpinski, and V. B. Sha. "Julia: A Fresh Approach to Numerical Computing". In: SIAM Review 59.1 (2017), pp. 65-98.
[27] M. E. Biancolini. Fast Radial Basis Functions for Engineering Applications. 1st ed. Springer, 2017.
[28] P. B. Bochev and A. C. Robinson. "Matching algorithms with physics: exact sequences of finite element spaces". In: Collected lectures on preservation of stability under discretization. Ed. by D. Estep and S. Tavener. Workshop on the preservation of stability under discretization. SIAM, 2001, pp. 145-166.
[29] D. E. Bockelman and W. R. Eisenstadt. "Combined Differential and CommonMode Scattering Parameters: Theory and Simulation". In: IEEE Transactions on Microwave Theory and Techniques 43.7 (1995), pp. 1530-1539.
[30] Z. Bontinck. Simulation and Robust Optimization for Electric Devices with Uncertainties. Technische Universität Darmstadt, PhD thesis. 2018.
[31] A. J. Booker, J. E. Dennis, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset. "A rigorous framework for optimization of expensive functions by surrogates". In: Structural Optimization 17.1 (1999), pp. 1-13.
[32] A. Bossavit. Computational Electromagnetism: Variational Formulations, Complementarity, Edge Elements. Academic Press, 1998.
[33] A. Bossavit. "Discretization of Electromagnetic Problems: The "Generalized Finite Differences" Approach". In: Numerical Methods in Electromagnetics. Ed. by P. G. Ciarlet. Vol. 13. Handbook of Numerical Analysis. Elsevier, 2005, pp. 105-197.
[34] J. P. Boyd and K. W. Gildersleeve. "Numerical experiments on the condition number of the interpolation matrices for radial basis functions". In: Applied Numerical Mathematics 61.4 (2011), pp. 443-459.
[35] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.
[36] R. Brown and T. Porter. "Category Theory: an abstract setting for analogy and comparison". In: What is Category Theory? Advanced Studies in Mathematics and Logic, Polimetrica Publisher. 2006, pp. 257-274.
[37] R. Brown and T. Porter. "The methodology of mathematics". In: The Mathematical Gazette 79.485 (1995), pp. 321-334.
[38] M. D. Buhmann. Radial Basis Functions - Theory and Implementation. 1st ed. Cambridge University Press, 2003.
[39] H.-J. Bungartz and M. Griebel. "Sparse grids". In: Acta Numerica 13 (2004), pp. 147-269.
[40] S. Burgard, O. Farle, P. Loew, and R. Dyczij-Edlinger. "Fast Shape Optimization of Microwave Devices Based on Parametric Reduced-Order Models". In: IEEE Transactions on Magnetics 50.2 (2014), pp. 629-632.
[41] R. M. Burkart. Advanced Modeling and Multi-Objective Optimization of Power Electronic Converter Systems. ETH Zürich, PhD thesis. 2016.
[42] W. Burke. Applied Differential Geometry. 1st ed. Cambridge University Press, 1985.
[43] J. M. F. Castillo. "The Hitchhiker Guide to Categorical Banach Space Theory. Part I". In: Extracta Mathematicae 25.2 (2010), pp. 103-149.
[44] L. Codecasa. "Novel Approach to Model Order Reduction for Nonlinear EddyCurrent Problems". In: IEEE Transactions on Magnetics 51.3 (2015), pp. 1-4.
[45] B. Coecke and A. Kissinger. Picturing Quantum Processes - A First Course in Quantum Theory and Diagrammatic Reasoning. Cambridge University Press, 2017.
[46] B. Coecke (editor). New Structures for Physics. Springer, 2011.
[47] A. R. Conn, K. Scheinberg, and L. N. Vicente. Introduction to Derivative-Free Optimization. 1st ed. SIAM, 2009.
[48] S. D. Conte and C. de Boor. Elementary Numerical Analysis: An Algorithmic Approach. 3rd ed. McGraw-Hill, 1980.
[49] G. Crevecoeur. Numerical Methods for Low Frequency Electromagnetic Optimization and Inverse Problems using Multi-Level Techniques. Ghent University, PhD thesis. 2009.
[50] G. Crevecoeur and R. H. De Staelen. "On cost function transformations for the reduction of uncertain model parameters' impact towards the optimal solutions". In: J. Comput. App. Math. 289 (2015), pp. 392-399.
[51] F. Cucker and D.-X. Zhou. Learning Theory: An Approximation Theory Viewpoint. Cambridge University Press, 2007.
[52] J. Culbertson and K. Sturtz. "A Categorical Foundation for Bayesian Probability". In: Applied Categorical Structures 22.4 (2014), pp. 647-662.
[53] A. Dean, M. Morris, J. Stufken, and D. Bingham (editors). Handbook of Design and Analysis of Experiments. 1st ed. CRC Press, 2015.
[54] M. C. Delfour and J.-P. Zolésio. Shapes and Geometries - Metrics, Analysis, Differential Calculus, and Optimization. 2nd ed. SIAM, 2011.
[55] P. Deuflhard and A. Hohmann. Numerical Analysis in Modern Scientific Computing: An Introduction. 2nd ed. Springer, 2003.
[56] D. Echeverría Ciaurri. Multi-Level Optimization: Space Mapping and Manifold Mapping. Amsterdam University, PhD thesis. 2007.
[57] M. S. Eldred and D. M. Dunlavy. "Formulations for surrogate-based optimization with data-fit, multifidelity and reduced-order models". In: 2006 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. Portsmouth, U.S.A., September 6-8, 2006, pp. 1-20.
[58] C. Elliott. "The simple essence of automatic differentiation". In: 2018 23rd ACM SIGPLAN International Conference on Functional Programming (ICFP). St. Louis, United States, September 23-29, 2018, pp. 1-29.
[59] M. H. van Emden and B. Moa. "Termination Criteria in the Moore-Skelboe Algorithm for Global Optimization by Interval Arithmetic". In: Frontiers in Global Optimization. Ed. by C. A. Floudas and P. Pardalos. Springer, 2004.
[60] R. W. Erickson and D. Maksimovic. Fundamentals of Power Electronics. 2nd ed. Springer, 2001.
[61] K.-T. Fang, R. Li, and A. Sudjianto. Design and Modeling for Computer Experiments. 1st ed. CRC Press, 2006.
[62] G. E. Fasshauer. Meshfree Approximation Methods with Matlab. 1st ed. World Scientific Publishing, 2007.
[63] G. E. Fasshauer and M. J. McCourt. "Stable Evaluation of Gaussian Radial Basis Function Interpolants". In: SIAM Journal on Scientific Computing 34.2 (2012), pp. 737-762.
[64] R. Feldt. Black-box optimization package for Julia. https://github.com/robert feldt/BlackBoxOptim.jl. [Online; accessed 12-February-2020]. 2020.
[65] J. L. Fiadeiro. Categories for Software Engineering. Springer, 2005.
[66] H. Flanders. Differential Forms with Applications to the Physical Sciences. 1st ed. Academic Press, 1963.
[67] C. A. Floudas. Deterministic Global Optimization: Theory, Methods and Applications. 1st ed. Springer, 2000.
[68] B. Fong, D. I. Spivak, and R. Tuyéras. "Backprop as Functor: A compositional perspective on supervised learning". In: 2019 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS). Vancouver, Canada, June 2427, 2019, pp. 1-13.
[69] A. Forrester and A. Keane. "Recent advances in surrogate-based optimization". In: Progress in Aerospace Sciences 45.1-3 (2009), pp. 50-79.
[70] A. Forrester, A. Sóbester, and A. Keane. Engineering Design via Surrogate Modelling - A Practical Guide. Wiley, 2008.
[71] P. Freyd. Abelian categories. An Introduction to the Theory of Functors. 1st ed. Harper and Row, 1964.
[72] F. Genovese, A. Gryzlov, J. Herold, A. Knispel, M. Perone, E. Post, and A. Videla. "idris-ct: A library to do category theory in Idris". In: 2019 2nd ACT Applied Category Theory Conference. Oxford, United Kingdom, July 15-19, 2019, 1-13.
[73] R. Geroch. Mathematical Physics. University of Chicago Press, 1985.
[74] R. G. Ghanem and P. D. Spanos. Stochastic Finite Elements: A Spectral Approach. 1st ed. Springer, 1991.
[75] A. Gil, J. Segura, and N. M. Temme. Numerical Methods for Special Functions. 1st ed. SIAM, 2007.
[76] D. Ginsbourger, R. Le Riche, and L. Carraro. "Kriging Is Well-Suited to Parallelize Optimization". In: Computational Intelligence in Expensive Optimization Problems. Ed. by Y. Tenne and C. K. Goh. Springer, 2010.
[77] L. Giraldi, A. Litvinenko, D. Liu, Matthies H. G., and A. Nouy. "To Be or Not to Be Intrusive? The Solution of Parametric and Stochastic Equations - the "Plain Vanilla" Galerkin Case". In: SIAM Journal on Scientific Computing 36.6 (2014), pp. 2720-2744.
[78] R. Goldman. Pyramid Algorithms: A Dynamic Programming Approach to Curves and Surfaces for Geometric Modeling. 1st ed. Morgan Kaufmann, 2003.
[79] L. Grasedyck, D. Kressner, and C. Tobler. "A literature survey of low-rank tensor approximation techniques". In: GAMM-Mitteilungen 36.1 (2013), pp. 5378.
[80] R. Griesse and B. Vexler. "Numerical Sensitivity Analysis for the Quantity of Interest in PDE-Constrained Optimization". In: SIAM J. Sci. Comput. 29.1 (2007), pp. 22-48.
[81] A. Griewank and A. Walther. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. 2nd ed. SIAM, 2008.
[82] J. Groß. Linear Regression. 1st ed. Springer, 2003.
[83] P. W. Gross and P. R. Kotiuga. Electromagnetic Theory and Computation: A Topological Approach. 1st ed. Cambridge University Press, 2004.
[84] NEOS Guide. NEOS Server: State-of-the-Art Solvers for Numerical Optimization. https://neos-guide .org/. [Online; accessed 12-February-2020]. 2020.
[85] B. Gustavsen and A. Semlyen. "Rational approximation of frequency domain responses by vector fitting". In: IEEE Transactions on Power Delivery 14.3 (1999), pp. 1052-1061.
[86] S. Gutsche. Constructive Category Theory and Applications to Algebraic Geometry. Universität Siegen, PhD thesis. 2017.
[87] R. T. Haftka, D. Villanueva, and A. Chaudhuri. "Parallel surrogate-assisted global optimization with expensive functions - a survey". In: Struct Multidisc Optim 54 (2016), pp. 3-13.
[88] R. Harper. Practical Foundations for Programming Languages. Cambridge University Press, 2nd ed., 2015.
[89] D. W. Hart. Introduction to Power Electronics. 1st ed. Prentice Hall, 1996.
[90] B. Hashemi and L. N. Trefethen. "Chebfun in Three Dimensions". In: SIAM Journal on Scientific Computing 39.5 (2017), pp. C341-C363.
[91] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. 2nd ed. Springer, 2008.
[92] A. Hatcher. Algebraic Topology. 1st ed. Cambridge University Press, 2002.
[93] F. W. Hehl and Y. N. Obukhov. Foundations of Classical Electrodynamics. 1st ed. Birkhäuser, 2003.
[94] J. S. Hesthaven, G. Rozza, and B. Stamm. Certified Reduced Basis Methods for Parametrized Partial Differential Equations. Springer, 2015.
[95] M. Hintermüller and L. Vicente. "Space mapping for optimal control of partial differential equations". In: SIAM J. Optim 15.4 (2005), pp. 1002-1025.
[96] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. Optimization with PDE Constraints. Springer, 2009.
[97] R. Hiptmair, F. Krämer, and J. Ostrowski. "Robust Maxwell Formulations for all Frequencies". In: IEEE Transactions on Magnetics 44.6 (2008), pp. 682-685.
[98] R. Horst and H. Tuy. Global optimization: Deterministic approaches. 3rd ed. Springer, 1995.
[99] A. S. Householder. The Theory of Matrices in Numerical Analysis. Dover, 2006.
[100] B. G. M. Husslage, G. Rennen, E. R. van Dam, and D. den Hertog. "Spacefilling Latin hypercube designs for computer experiments". In: Optimization and Engineering 12 (2011), pp. 611-630.
[101] Julia Interop. An interface for using MATLAB ${ }^{\text {TM }}$ from Julia using the MATLAB C api. https:// github. com/JuliaInterop/MATLAB.jl. [Online; accessed 12-February-2020]. 2020.
[102] C. Ionescu and P. Jansson. "Domain-specific languages of mathematics: Presenting mathematical analysis using functional programming". In: 4th and 5th International Workshop on Trends in Functional Programming in Education. 2016, pp. 1-15.
[103] J. D. Jackson. Classical Electrodynamics. 3rd ed. John Wiley \& Sons, 1999.
[104] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. Applied Interval Analysis: With Examples in Parameter and State Estimation, Robust Control and Robotics. 1st ed. Springer, 2001.
[105] S. G. Johnson. HCubature.jl - A pure-Julia multidimensional h-adaptive integration. https://github. com/JuliaMath/HCubature . j1. [Online; accessed 12-February-2020]. 2020.
[106] S. G. Johnson. Sobol low-discrepancy-sequence (LDS) package for Julia. https: //github . com/stevengj / Sobol . jl. [Online; accessed 12-February-2020]. 2020.
[107] S. G. Johnson. The NLopt nonlinear-optimization package. http://github.com/ stevengj/nlopt. [Online; accessed 12-February-2020]. 2020.
[108] D. R. Jones. "A Taxonomy of Global Optimization Methods Based on Response Surfaces". In: Journal of Global Optimization 21.4 (2001), pp. 345-383.
[109] D. R. Jones, C. D. Perttunen, and B. E. Stuckman. "Lipschitzian optimization without the Lipschitz constant". In: J. Optimization Theory and Application 79.1 (1993), pp. 157-181.
[110] JuliaPy. A Julia interface to SymPy (a Python library for symbolic mathematics) via PyCall. https://github. com/JuliaPy/SymPy .j1. [Online; accessed 12-February-2020]. 2020.
[111] J. Kangas, T. Tarhasaari, and L. Kettunen. "Maxwell equations and finite element software systems: object-oriented coding needs well defined objects". In: IEEE Transactions on Magnetics 36.4 (2000), pp. 1645-1648.
[112] M. K. Kazimierczuk. High-Frequency Magnetic Components. 2nd ed. Wiley, 2014.
[113] P. Kerschke, H. H. Hoos, F. Neumann, and H. Trautmann. "Automated Algorithm Selection: Survey and Perspectives". In: Evolutionary Computation 27.1 (2019), pp. 3-45.
[114] L. Kettunen, T. Kovanen, and T. Tarhasaari. "Electromagnetism and crossdisciplinary problems". In: 2016 URSI International Symposium on Electromagnetic Theory (EMTS). Espoo, Finland, August 14-18, 2016, pp. 500-501.
[115] D. Klis, M. Jochum, O. Farle, and R. Dyczij-Edlinger. "Application of nonlinear model-order reduction to 3D eddy current problems". In: 2013 International Conference on Electromagnetics in Advanced Applications (ICEAA). Torino, Italy, September 09-13, 2013, pp. 344-347.
[116] M. J. Kochenderfer and T. A. Wheeler. Algorithms for Optimization. The MIT Press, 2019.
[117] I. Kolář, P. W. Michor, and J. Slovák. Natural operations in differential geometry. Springer, 1993.
[118] Y. Konkel, O. Farle, A. Köhler, A. Schultschik, and R. Dyczij-Edlinger. "Adaptive strategies for fast frequency sweeps". In: COMPEL 30.2 (2011), pp. 18551869.
[119] S. Koziel. "Space mapping: Performance, reliability, open problems and perspectives". In: 2017 IEEE MTT-S International Microwave Symposium (IMS). Honolulu, U.S.A., June 4-9, 2017, pp. 1512-1514.
[120] S. Koziel and J. W. Bandler. "Coarse and Surrogate Model Assessment for Engineering Design Optimization with Space Mapping". In: 2007 IEEE/MTTS International Microwave Symposium (IMS). Honolulu, U.S.A., June 3-8, 2007, pp. 107-110.
[121] S. Koziel, J. W. Bandler, and K. Madsen. "Quality assessment of coarse models and surrogates for space mapping optimization". In: Optimization and Engineering 9 (2008), pp. 375-391.
[122] S. Koziel and A. Bekasiewicz. "Sequential approximate optimisation for statistical analysis and yield optimisation of circularly polarised antennas". In: IET Microw. Antennas Propag 12.13 (2018), pp. 2060-2064.
[123] S. Koziel and A. Bekasiewicz. "Variable-fidelity design optimization of antennas with automated model selection". In: 2016 10th European Conference on Antennas and Propagation (EuCAP). Davos, Switzerland, April 10-15, 2016, pp. 1-5.
[124] S. Koziel and D. Echeverría Ciaurri. "Reliable Simulation-Driven Design Optimization of Microwave Structures Using Manifold Mapping". In: Progress In Electromagnetics Research B 26.1 (2010), pp. 361-382.
[125] S. Koziel, D. Echeverría Ciaurri, and L. Leifsson. "Chapter 3: Surrogate-Based Methods". In: Computational Optimization, Methods and Algorithms. Ed. by S. Koziel and X. S. Yang. Vol. 356. Studies in Computational Intelligence. Springer, 2011, pp. 33-59.
[126] S. Koziel, L. Leifsson, and X. S. Yang (editors). Solving Computationally Expensive Engineering Problems: Methods and Applications. Springer, 2014.
[127] S. Koziel and L. Leifsson (editors). Surrogate-Based Modeling and Optimization: Applications in Engineering. Springer, 2013.
[128] D. Kraft. "Algorithm 733: TOMP-Fortran modules for optimal control calculations". In: ACM Transactions on Mathematical Software 20.3 (1994), pp. 262281.
[129] S. Kucherenko and B. Iooss. "Derivative-Based Global Sensitivity Measures". In: Handbook of Uncertainty Quantification. Ed. by R. Ghanem, D. Higdon, and H. Owhadi. Cham: Springer, 2017. Chap. 36, pp. 1241-1263.
[130] J. Kuipers, A. Plaat, J. A. M. Vermaseren, and H. J. van den Herik. "Improving multivariate Horner schemes with Monte Carlo tree search". In: Computer Physics Communications 184.11 (2013), pp. 2391-2395.
[131] S. Kurz and B. Auchmann. "Differential Forms and Boundary Integral Equations for Maxwell-Type Problems". In: Fast Boundary Element Methods in Engineering and Industrial Applications. Ed. by U. Langer, M. Schanz, O. Steinbach, and W. Wendland. Springer, 2012.
[132] V. Lahtinen. Searching for Frontiers in Contemporary Eddy Current Model Based Hysteresis Loss Modelling of Superconductors. Tampere University of Technology, PhD thesis. 2014.
[133] V. Lahtinen, P. R. Kotiuga, and A. Stenvall. An electrical engineering perspective on missed opportunities in computational physics. arXiv:1809.01002v2. 2018.
[134] S. Lang. Differential manifolds. 2nd ed. Springer, 1985.
[135] J. Larson, M. Menickelly, and S. M. Wild. "Derivative-free optimization methods". In: Acta Numerica 28 (2019), pp. 287-404.
[136] F. W. Lawvere and S. H. Schanuel. Conceptual Mathematics. 2nd ed. Cambridge University Press, 2009.
[137] L. Lebensztajn, C. A. R. Marretto, M. C. Costa, and J.-L. Coulomb. "Kriging: A Useful Tool for Electromagnetic Device Optimization". In: IEEE Transactions on Magnetics 40.2 (2004), pp. 1196-1199.
[138] M. C. Lehmann, M. Hadžiefendić, A. Piwonski, and R. Schuhmann. "Encoding Electromagnetic Transformation Laws for Dimensional Reduction". In: International Journal of Numerical Modelling: Electronic Networks, Devices and Fields 33.1 (2020). https://doi . org/10. 1002/jnm.2747, e2747.
[139] G. Lehner. Electromagnetic Field Theory for Engineers and Physicists. 1st ed. Springer, 2010.
[140] P. Linz. Theoretical Numerical Analysis: An Introduction to Advanced Techniques. 1st ed. John Wiley \& Sons, 1979.
[141] E. Ljungskog. Interpolation of scattered data in Julia. https://github.com/el jungsk/ScatteredInterpolation. j1. [Online; accessed 12-February-2020]. 2020.
[142] D. G. Luenberger. Optimization by Vector Space Methods. Wiley, 1969.
[143] H. D. Macedo and J. N. Oliveira. "Typing linear algebra: A biproduct-oriented approach". In: Science of Computer Programming 78.11 (2013), pp. 2160-2191.
[144] S. MacLane. Categories for the Working Mathematician. Springer, 1971.
[145] N. Marheineke and R. Pinnau. "Model hierarchies in space mapping optimization: Feasibility study for transport processes". In: J. Comput. Meth. Sci. Eng 12.1,2 (2012), pp. 63-74.
[146] R. T. Marler and J. S. Arora. "Survey of Multi-Objective Optimization Methods for Engineering". In: Structural and Multidisciplinary Optimization 26.6 (2004), pp. 369-395.
[147] J. R. R. A. Martins and A. B. Lambe. "Multidisciplinary Design Optimization: A Survey of Architectures". In: AIAA Journal 51.9 (2013), pp. 2049-2075.
[148] P. McCullagh. "What is a statistical model?" In: The Annals of Statistics 30.5 (2002), pp. 1225-1310.
[149] D. Meeker. Finite Element Method Magnetics (FEMM4.2). http://www.femm. info/. 2017.
[150] K. Miettinen. Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, 1999.
[151] P. K. Mogensen and A. N. Riseth. "Optim: A mathematical optimization package for Julia". In: Journal of Open Source Software 3.24 (2018), p. 615.
[152] N. Mohan, T. M. Undeland, and W. P. Robbins. Power Electronics: Converters, Applications, and Design. 3rd ed. Wiley, 2002.
[153] P. Monk. Finite Element Methods for Maxwell's Equations. 1st ed. Oxford University Press, 2003.
[154] J. Mühlethaler. Modeling and multi-objective optimization of inductive power components. ETH Zürich, PhD thesis. 2012.
[155] K. P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
[156] R. D. Neidinger. "Introduction to Automatic Differentiation and MATLAB Object-Oriented Programming". In: SIAM Review 52.3 (2010), pp. 545-563.
[157] A. Neumaier. "Complete search in continuous global optimization and constraint satisfaction". In: Acta Numerica 13 (2004), pp. 271-369.
[158] J. Nocedal and S. J. Wright. Numerical Optimization. 2nd ed. Springer, 2006.
[159] W. L. Oberkampf and C. J. Roy. Verification and Validation in Scientific Computing. 1st ed. Cambridge University Press, 2010.
[160] J. T. Oden and S. Prudhomme. "Estimation of Modeling Error in Computational Mechanics". In: Journal of Computational Physics 182.2 (2002), pp. 496515.
[161] J. Oliver. "Rounding error propagation in polynomial evaluation schemes". In: Journal of Computational and Applied Mathematics 5.2 (1979), pp. 85-97.
[162] S. Patel. LaTeX Templates: Masters/Doctoral Thesis. https : / / www . latextem plates. com/template/masters-doctoral-thesis. [Online; accessed 12-February-2020]. 2020.
[163] E. Patterson. Catlab - A framework for applied category theory in the Julia language. https://github . com/AlgebraicJulia/ Catlab.j1. [Online; accessed 12-February-2020]. 2020.
[164] C. R. Paul. Introduction to Electromagnetic Compatibility. 2nd ed. Wiley, 2006.
[165] B. Peherstorfer, K. Willcox, and M. Gunzburger. "Optimal model management for multifidelity monte carlo estimation". In: SIAM Journal on Scientific Computing 38.5 (2016), pp. 3163-3194.
[166] B. Peherstorfer, K. Willcox, and M. Gunzburger. "Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization". In: SIAM Review 60.3 (2018), pp. 550-591.
[167] B. C. Pierce. Basic Category Theory for Computer Scientists. MIT Press, 1991.
[168] S. Posur. Constructive Category Theory and Applications to Equivariant Sheaves. Universität Siegen, PhD thesis. 2017.
[169] M. J. D. Powell. Approximation Theory and Methods. 1st ed. Cambridge University Press, 1981.
[170] M. J. D. Powell. "Direct search algorithms for optimization calculations". In: Acta Numerica 7 (1998), pp. 287-336.
[171] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. Numerical Recipes - The Art of Scientific Computing. 3rd ed. Cambridge University Press, 2007.
[172] C. Psarras, H. Barthels, and P. Bientinesi. The Linear Algebra Mapping Problem. arXiv:1911.09421v1. 2019.
[173] C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning. MIT Press, 2006.
[174] P. Raumonen. Mathematical structures for dimensional reduction and equivalence classification of electromagnetic boundary value problems. Tampere University of Technology, PhD thesis. 2009.
[175] J. Revels, M. Lubin, and T. Papamarkou. Forward-Mode Automatic Differentiation in Julia. arXiv:1607.07892. 2016.
[176] J. A. Richardson and J. L. Kuester. "The complex method for constrained optimization". In: Commun. ACM 16.8 (1973), pp. 487-489.
[177] E. Riehl. Category Theory in Context. Dover, 2016.
[178] U. Römer. Numerical Approximation of the Magnetoquasistatic Model with Uncertainties and its Application to Magnet Design. Technische Universität Darmstadt, PhD thesis. 2015.
[179] A. A. Rodríguez and A. Valli. Eddy Current Approximation of Maxwell Equations - Theory, algorithms and applications. 1st ed. Springer, 2010.
[180] S. Roman. An Introduction to the Language of Category Theory. Birkhäuser, 2017.
[181] G. Roussos and B. J. C. Baxter. "Rapid evaluation of radial basis functions". In: Journal of Computational and Applied Mathematics 180.1 (2005), pp. 51-70.
[182] C. J. Roy and W. L. Oberkampf. "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing". In: Computer Methods in Applied Mechanics and Engineering 200.25-28 (2011), pp. 2131-2144.
[183] W. Rudin. Principles of Mathematical Analysis. 3rd ed. McGraw-Hill, 1976.
[184] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. "Design and Analysis of Computer Experiments". In: Statistical Science 4.4 (1989), pp. 409-435.
[185] B. S. Saini, M. Lopez-Ibanez, and K. Miettinen. "Automatic surrogate modelling technique selection based on features of optimization problems". In: 2019 The Genetic and Evolutionary Computation Conference (GECCO). Prague, Czech Republic, July 13-17, 2019, pp. 1765-1772.
[186] D. P. Sanders. Rigorous global optimisation package for Julia. https://githu b. com/JuliaIntervals / IntervalOptimisation . jl. [Online; accessed 12-February-2020]. 2020.
[187] R. Schaback and H. Wendland. "Kernel techniques: From machine learning to meshless methods". In: Acta Numerica 15 (2006), pp. 543-639.
[188] M. Scheuerer, R. Schaback, and M. Schlather. "Interpolation of spatial data - A stochastic or a deterministic problem?" In: European Journal of Applied Mathematics 24.4 (2013), pp. 601-629.
[189] W. H. Schilders, H. A. van der Vorst, and J. Rommes (editors). Model Order Reduction: Theory, Research Aspects and Applications. 1st ed. Springer, 2008.
[190] R. B. Schnabel. "Parallel Nonlinear Optimization: Limitations, Challenges, and Opportunities". In: Algorithms for Continuous Optimization: The State of the Art. Ed. by E Spedicato. Vol. 434. NATO ASI Series (Series C: Mathematical and Physical Sciences). Springer, 1994, pp. 531-559.
[191] S. Y. Serovajsky. "Differentiation Functor and Its Application in the Optimization Control Theory". In: Fourier Analysis. Ed. by Michael Ruzhansky and Ville Turunen. Springer International Publishing, 2014, pp. 335-347.
[192] S. Shalev-Shwartz and S. Ben-David. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.
[193] C. H. da Silva Santos, M. S. Gonçalves, and H. E. Hernandez-Figueroa. "Designing Novel Photonic Devices by Bio-Inspired Computing". In: IEEE Photonics Technology Letters 22.15 (2010), pp. 1177-1179.
[194] J. Søndergaard. Optimization using surrogate models - by the space mapping technique. Technical University of Denmark, PhD thesis. 2003.
[195] J. C. Spall. Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control. 1st ed. Wiley, 2003.
[196] P. Spirtes, C. Glymour, and R. Scheines. Causation, Prediction, and Search. 2nd ed. MIT Press, 2001.
[197] D. I. Spivak. "Categories as Mathematical Models". In: Categories for the Working Philosopher (ed. by Elaine Landry), Oxford University Press. 2017, pp. 381-401.
[198] T. Steinmetz, S. Kurz, and M. Clemens. "Domains of validity of quasistatic and quasistationary field approximations". In: COMPEL 30.4 (2011), pp. 12371247.
[199] A. Stenvall and V. Lahtinen. "The Methodology of HTS AC-Loss Modeling". In: IEEE Transactions on Applied Superconductivity 29.5 (2019), pp. 1-7.
[200] G. Strang. Computational Science and Engineering. 1st ed. Wellesley-Cambridge Press, 2007.
[201] T. J. Sullivan. Introduction to Uncertainty Quantification. 1st ed. Springer, 2015.
[202] S. Surjanovic and D. Bingham. Virtual Library of Simulation Experiments: Test Functions and Datasets. http://www.sfu.ca/ ${ }^{\text {ssurjano. [Online; accessed }}$ 12-February-2020]. 2020.
[203] K. Svanberg. "A class of globally convergent optimization methods based on conservative convex separable approximations". In: SIAM J. Optim. 12.2 (2002), pp. 555-573.
[204] W. P. Thurston. "On proof and progress in mathematics". In: Bulletin of the American Mathematical Society 30.2 (1994), pp. 161-177.
[205] E. Tonti. The Mathematical Structure of Classical and Relativistic Physics: A General Classification Diagram. Springer, 2013.
[206] A. Townsend and L. N. Trefethen. "An Extension of Chebfun to Two Dimensions". In: SIAM Journal on Scientific Computing 35.6 (2013), pp. C495-C518.
[207] J. F. Traub and A. G. Werschulz. Complexity and Information. 1st ed. Cambridge University Press, 1998.
[208] L. N. Trefethen. Approximation Theory and Approximation Practice. 1st ed. SIAM, 2013.
[209] L. N. Trefethen. "Six myths of polynomial interpolation and quadrature". In: Mathematics Today 47.4 (2012), pp. 184-188.
[210] F. Tröltzsch. Optimal Control of Partial Differential Equations - Theory, Methods and Applications. Graduate Studies in Mathematics, Volume: 112. American Mathematical Society, 2010.
[211] F. Tröltzsch and A. Valli. "Optimal control of low-frequency electromagnetic fields in multiply connected conductors". In: Optimization 65.9 (2016), pp. 16511673.
[212] R. Trobec and G. Kosec. Parallel Scientific Computing: Theory, Algorithms, and Applications of Mesh Based and Meshless Methods. Springer, 2015.
[213] W. Tucker. Validated Numerics: A Short Introduction to Rigorous Computations. Princeton University Press, 2011.
[214] The Univalent Foundations Program. Homotopy Type Theory: Univalent Foundations of Mathematics. Institute for Advanced Study: https://homotopytype theory .org/book, 2013.
[215] M. Urquhart. Julia package for the creation of optimised Latin Hypercube Sampling Plans. https://github com/MrUrq/LatinHypercubeSampling.jl. [Online; accessed 12-February-2020]. 2020.
[216] M. Urquhart, E. Ljungskog, and S. Sebben. "Surrogate-based optimisation using adaptively scaled radial basis functions". In: Applied Soft Computing 88 (2020), pp. 1-17.
[217] S. Voß. "Meta-heuristics: The State of the Art". In: Local Search for Planning and Scheduling (LSPS 2000). Ed. by A. Nareyek. Springer, 2001.
[218] P. Šolin. Partial Differential Equations and the Finite Element Method. 1st ed. John Wiley \& Sons, 2006.
[219] R. F. C. Walters. Categories and Computer Science. Cambridge Computer Science Texts, 1991.
[220] Q. Wang, X. Zhang, R. Burgos, D. Boroyevich, A. White, and M. Kheraluwala. "Design and optimization of a high performance isolated three phase AC/DC converter". In: 2016 IEEE Energy Conversion Congress and Exposition (ECCE). Milwaukee, U.S.A., September 18-22, 2016, pp. 1-10.
[221] R. Webster and M. A. Oliver. Geostatistics for Environmental Scientists. 2nd ed. John Wiley \& Sons, 2007.
[222] B. Wen et al. "Integrated Design by Optimization of Electrical Power Systems for More Electric Aircraft". In: 2015 More Electric Aircraft (MEA). Toulouse, France, February 3-5, 2015, pp. 1-4.
[223] T. Wittig. Zur Reduzierung der Modellordnung in elektromagnetischen Feldsimulationen. Technische Universität Darmstadt, PhD thesis. 2004.
[224] D. H. Wolpert and W. G. Macready. "No Free Lunch Theorems for Optimization". In: IEEE Transactions on Evolutionary Computation 1.1 (1997), pp. 67-82.
[225] N. S. Yanofsky. "Towards a Definition of an Algorithm". In: Journal of Logic and Computation 21.2 (2010), pp. 253-286.
[226] K. Yosida. Functional Analysis. 6th ed. Springer, 1995.
[227] S. Zaglmayr. High Order Finite Element Methods for Electromagnetic Field Computation. Johannes Kepler Universität Linz, PhD thesis. 2006.
[228] Z.-H. Zhan, J. Zhang, Y. Li, and H. S. Chung. "Adaptive Particle Swarm Optimization". In: IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39.6 (2009), pp. 1362-1381.


[^0]:    ${ }^{1}$ Notice that the present dissertation's typesetting builds upon a free and open-source $\mathrm{LA}_{\mathrm{E}} \mathrm{X}$ typesetting template provided by LaTeX Templates (see [162]).
    ${ }^{2}$ Mind that, for the purpose of drawing figures in the present dissertation, I invoke the free and open-source vector graphics editor Inkscape (see version 1.0.1 at https://inkscape.org/), the free vector graphics editor Ipe (see version 7.2.20 at http://ipe.otfried.org/), and the free and opensource 3D computer graphics software Blender (see version 2.91.0 at https ://www.blender. org/).

[^1]:    ${ }^{3} \mathrm{CST}$ Studio Suite ${ }^{\circledR}$ is a proprietary commercial 3D electromagnetic analysis software package by Dassault Systèmes (see version 2019 at https ://www .3ds . com/).

[^2]:    ${ }^{4}$ Note that I employ the affix "guided" in the term "surrogate-guided". This term is by no means a member of the usual terminology in which "surrogate-guided" would be a synonym for "surrogatebased". However, I consider the terminological supplement as a helpful tool to enable better conceptional differentiation of the corresponding mechanisms.

[^3]:    ${ }^{5}$ In the literature (see, e.g., [49, p. 45]), data-fit low-fidelity models are also called metamodels. Moreover, in [57], the authors suggest the wording data-fit, multifidelity, and reduced-order.

[^4]:    ${ }^{6}$ Unfortunately, there is some ambiguity regarding the term "ohmic loss". On the one hand, it refers to the non-negative real number resulting from the ohmic loss integral computation; on the other hand, it refers to the ohmic loss integral itself as a map. In the present context, I refer to the map.

[^5]:    ${ }^{7}$ MATLAB ${ }^{\circledR}$ is a proprietary commercial programming language by MathWorks (see version R2019b at https://www mathworks.com/).
    ${ }^{8}$ Julia is a free and open-source programming language (see version v1.5.3 at https:// julialang. org/). For more details on the dynamically-checked programming language Julia, I refer to, e.g., [26] or [25].

[^6]:    ${ }^{9}$ In [199], the authors offer a general classification of research methodologies. Although their field of application is the numerical modeling of AC-loss in high-temperature superconductors, their classification has an application-independent general validity. Note that their classification can be regarded as an extension of the classification in [37] of the methodology of mathematics.

[^7]:    ${ }^{1}$ The bullet points are a direct quotation of the listing in [205, p. 273], but the italic face and bold face are my emphasis.

[^8]:    ${ }^{2}$ Borrowing from programming language theory (see, e.g., [88]), I conceive the term "function signature" similarly to the sense of the term "type signature". For example, given a function called $f$ that maps a real number $x$ to a real number $x \bullet_{\mathbb{R}} x$ where the function $\bullet_{\mathbb{R}}$ indicates the real-valued binary multiplication map, then one can write $f=x \mapsto x \cdot \mathbb{R} x: \mathbb{R} \rightarrow \mathbb{R}$ such that $\mathbb{R} \rightarrow \mathbb{R}$ is the type signature (or type annotation) of the function $f$ (cf. the discussion in [32, p. 279-282]). Setting $A \equiv \mathbb{R}$ and $B \equiv \mathbb{R}$ in this example, then, roughly speaking, the type of $x$ refers to $A$ (such that $x: A$ ), the type of $x \cdot{ }_{A} x$ refers to $B$ (such that $x \cdot{ }_{A} x: B$ ), and the type of $f$ refers to $A \rightarrow B$ or $B^{A}$ (such that $f: A \rightarrow B$ and $f: B^{A}$, respectively).

[^9]:    " scalar field $\xrightarrow{\text { grad }}$ vector field $\xrightarrow{\text { curl }}$ vector field $\xrightarrow{\text { div }}$ scalar field $"$

[^10]:    ${ }^{3}$ Often, the factor 100 is used instead of the factor 5 . However, the factor 5 is employed analogously to [116, p. 431].
    ${ }^{4}$ Mind that the global minimum (2.20,1.57) is just an approximation (see, e.g., [116, p. 430]).
    ${ }^{5}$ The additional term $5 x_{1}$ is the modification (cf. [70, p. 196]) to the Branin function (see, e.g., [116, p. 427]). The global minimum ( $-3.70,13.63$ ) is just an approximation.

[^11]:    ${ }^{6}$ The map grad in (2.7) is overloaded in the sense that grad is equipped with the signature $(U \rightarrow$ $\mathbb{R}) \rightarrow\left(U \rightarrow \mathbb{R}^{2}\right)$. Given a point $p \in U$, one can extract the components of $\operatorname{grad}(f)(p)$, that is, $\partial_{\mathbf{e}_{x_{1}}}(f)(p)$ and $\partial_{\mathbf{e}_{x_{2}}}(f)(p)$, by setting $\partial_{\mathbf{e}_{x_{1}}}(f)(p):=\operatorname{grad}(f)(p) \cdot \mathbb{R}^{2} \mathbf{e}_{x_{1}}$ and by setting $\partial_{\mathbf{e}_{x_{2}}}(f)(p):=\operatorname{grad}(f)(p) \cdot \mathbb{R}^{2}$ $\mathbf{e}_{x_{2}}$ where $\cdot \mathbb{R}^{2}$ denotes the Euclidean inner product w.r.t. $\mathbb{R}^{2}$; and $\mathbf{e}_{x_{1}}$ and $\mathbf{e}_{x_{2}}$ refer to the unit vectors w.r.t. $x_{1}$ and $x_{2}$, respectively.
    ${ }^{7}$ For more details on gradient-based sensitivity measures such as, e.g., other possible definitions than the definition in (2.49), I refer to [129] and references therein.

[^12]:    ${ }^{8}$ Due to numerical inaccuracies, it is needed to set $U \equiv[-29.9999,30.0] \times[-30.0,30.0]$ and $U \equiv[-1.9999,2.0] \times[-2.0,2.0]$, respectively, in the case of the test function (i), i.e., Ackley, in order to ensure that, for all test functions, the estimated error with regard to the estimated integral is at least below $1 \times 10^{-4}$.

[^13]:    ${ }^{9}$ Commonly, the term meta-heuristic (see, e.g., [217]) refers to a strategic search by trial and error without a theoretical guarantee of global optimality.

[^14]:    ${ }^{10}$ For more details on gradient-based optimization methods, I refer to, e.g., [158], [18]; on derivativefree optimization algorithms, I refer to, e.g., [47], [10], [135]; and on deterministic and stochastic global optimization algorithms, I refer to, e.g., [98], [67], and, e.g., [24], [195], respectively.
    ${ }^{11}$ For further deliberations on deterministic global optimization using interval arithmetic, I refer to, e.g., the survey in [157].

[^15]:    ${ }^{1}$ An identification problem is also frequently named recovery problem (see, e.g., [187, p. 551f]). A direct problem is often called a forward problem as well (see, e.g., [49, p. 5]).

[^16]:    ${ }^{2}$ The notation in (3.7) follows the customary index convention from multilinear algebra in order to emphasize the linear combination of members of a vector space. The operation ${ }^{\mathbb{F}}$ implies a signature $\mathbb{F} \times \operatorname{hom}_{\mathbb{F}}(X, Y) \rightarrow \operatorname{hom}_{\mathbb{F}}(X, Y)$. The operation's concrete implementation depends on the selected concrete function space like, for example, $\mathrm{P}_{\leq n}$.

[^17]:    ${ }^{3}$ I regard $Y^{X}:=L^{2}$ solely as a powerful interface. The most important use case is given by $X:=\mathbb{R}^{d}$ with $d \in \mathbb{N}$ and $Y:=\mathbb{R}$. For all the subtleties regarding the constructions such as Borel $\sigma$-algebra, Lebesgue measure, the extended real numbers, the Lebesgue integral, and similar, I refer to the literature (see, e.g., [201, ch. 3]).

[^18]:    ${ }^{4}$ I do not invoke the package's functionality to skip the initial portion of the Sobol sequence which, allegedly, could further improve the uniform spreading of the sampling plan's points.
    ${ }^{5}$ From the viewpoint of statistical learning theory, this error is a representative of mean squared errors (see, e.g., [116, p. 265]). Mean squared errors along with root mean squared errors (see, e.g., [70, p. 37]) can be considered as parts of squared error loss functions (see, e.g., [91, p. 219]).

[^19]:    ${ }^{6}$ In statistics vernacular (see, e.g., [61, p. 249-252]), the input entity, the output entity, and the error entity are conceived as random variables $X, Y$, and $\varepsilon$, respectively. Hence, the high-fidelity model K is regarded as an unknown smooth regression function, i.e., as the conditional expectation $\mathbb{E}(Y \mid X=x)=: \mathrm{K}(x)$. Regarding the conditional variance $\mathbb{V}_{\mu \equiv 0}(\varepsilon \mid X=x)$, it is assumed that the homoscedasticity property holds, that is, $\forall i . \mathbb{V}_{\mu \equiv 0}\left(\varepsilon_{i} \mid X=x\right) \equiv \sigma_{\varepsilon \mid X}^{2}$ with $\sigma_{\varepsilon \mid X}^{2} \in \mathbb{R}^{+} \cup\{+\infty\}$. In addition, noise is only modeled by additive Gaussian noise.

[^20]:    ${ }^{7}$ Apart from these heuristic values, there is a lack of rigorously proven lower or upper bounds for the number of runs or the number of folds.

[^21]:    ${ }^{8}$ It is implicitly supposed that the limit considerations are with regard to some appropriate norms.

[^22]:    ${ }^{9}$ In span $(B)$, technically, the object under the given predicate allows to meaningfully state the predicate $p(x) \in \mathrm{P}_{\leq n}$ where $\mathrm{P}_{\leq n} \equiv \operatorname{span}(B)$. In order to meaningfully state the predicate $p \in \mathrm{P}_{\leq n}$ with $p=x \mapsto p(x): X \rightarrow \mathbb{R}$, one should understand the object in $\operatorname{span}(B)$ as a shorthand notation for $x \mapsto \sum_{i=0}^{n} \tilde{c}^{i} \cdot \mathbb{R} \tilde{\varphi}_{i}(x): X \rightarrow \mathbb{R}$.
    ${ }^{10}$ Generally, constructing multivariate functions as tensor products of univariate functions involves many subtelties such as a quotient space or the universal property. For some subtleties of such constructions, see, e.g., [201, p. 45-52].

[^23]:    ${ }^{11}$ Technically, the representation of a vector $y \in \mathbb{R}^{d}$ as a column vector $y \in \mathbb{R}^{d \times 1}$ and the representation of a $1 \times 1$ matrix $\gamma \in \mathbb{R}^{1 \times 1}$ as a scalar $\gamma \in \mathbb{R}$ involves some kind of isomorphisms, in order to state $\mathbb{R}^{d} \cong \mathbb{R}^{d \times 1}$ and $\mathbb{R} \cong \mathbb{R}^{1 \times 1}$.

[^24]:    ${ }^{12}$ Let us conceive the rank of the $m \times s$ matrix $B$ in the sense that $\operatorname{rank}(B) \leq \min \{m, s\}$. If $m=q$ and $\operatorname{rank}(B)=m$, then we say that the rank of the matrix $B$ is full.
    ${ }^{13}$ Let us comprehend the trace of a square matrix $A \in \mathbb{R}^{n} \times \mathbb{R}^{n}$ as the map $\operatorname{tr}: \mathbb{R}^{n} \times \mathbb{R}^{n} \rightarrow \mathbb{R}^{+}$with $\operatorname{tr}(A):=\sum_{i=1}^{n} a_{i, i}$.

[^25]:    ${ }^{14}$ Given a column vector $d \in \mathbb{R}^{n \times 1}$ and a square diagonal matrix $A \in \mathbb{R}^{n \times n}$ where $\forall i, j \in\{1,2, \ldots, n\}$. $i \neq j \Longrightarrow a_{i, j}:=0$, let us comprehend diag as the map with the signature $\mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{n \times n}$ and the assignment $\operatorname{diag}(d):=\left[a_{i, i} \equiv d_{i}\right]$. Note well that, in the specific context in which the map diag lives, the term $\operatorname{diag}\left(d_{1}, \ldots, d_{n}\right)$ is treated as a rewriting of the term $\operatorname{diag}\left(\left[d_{1}, \ldots, d_{n}\right]^{\mathrm{T}}\right)$.

[^26]:    ${ }^{15}$ Let us comprehend the condition number for inversion of a matrix $A \in \mathbb{R}^{m} \times \mathbb{R}^{n}$ with $m \geq n$ as the map $\kappa: \mathbb{R}^{m} \times \mathbb{R}^{n} \rightarrow \mathbb{R}^{+}$with $\kappa(A):=\|A\|_{2}\left\|A^{+}\right\|_{2}$ where $\|\cdot\|_{2}: \mathbb{R}^{m} \times \mathbb{R}^{n} \rightarrow \mathbb{R}^{+}$denotes the matrix norm induced by the $l_{2}$-norm for vectors. If we possess $A^{\prime}$ 's largest singular value $\sigma_{\max }$ and its smallest singular value $\sigma_{\min }$, then one can set $\kappa(A) \equiv \frac{\sigma_{\max }}{\sigma_{\min }}$. Given a positive integer $k$ where $\kappa(A) \propto 10^{k}$, then, very roughly speaking, it is supposed that there are only $16-k$ or $16-\log _{10}(\kappa(A))$ significant digits of an output's accuracy in a double-precision floating-point format.

[^27]:    ${ }^{16}$ Note that the determination of the coefficients of the extended radial basis low-fidelity model $\hat{\psi}$ results in a block matrix which incorporates additional constraints in order to uniquely determine all the coefficients. By considering the corresponding Schur complements, it is probably beneficial to experiment with various combinations of polynomials $p \in \mathrm{P}_{\leq k}^{d}$ of varying degree $k$ and radial basis functions $\psi$ in order to achieve desirable properties such as a symmetric positive definite block matrix.

[^28]:    ${ }^{17}$ Let us set $|\Sigma|:=\operatorname{det}(\Sigma)$ where we comprehend the map det as the determinant of a square matrix, more precisely, det $=\Sigma \mapsto \operatorname{det}(\Sigma): \mathbb{R}^{m \times m} \rightarrow \mathbb{R}$.

[^29]:    ${ }^{18}$ Given $|\Psi|:=\operatorname{det}(\Psi)$ and if $\lambda_{i}$ denote the eigenvalues of the matrix $\Psi \in \mathbb{R}^{m \times m}$, then one can invoke the statement $\operatorname{det}(\Psi) \equiv \prod_{i=1}^{m} \lambda_{i}$. Hence, if the matrix $\Psi$ is positive definite and all of its eigenvalues are positive, then $\exists \Psi . \operatorname{det}(\Psi)>0$ holds.
    ${ }^{19}$ For a more elaborated treatment of some aspects regarding the interpretation of the likelihood function, I refer to [201, p. 29ff].

[^30]:    ${ }^{20}$ More precisely, in statistics vernacular, $\hat{y}(x)$ indicates the best linear unbiased predictor (BLUP) (see, e.g., [61, p. 146f]).

[^31]:    ${ }^{21}$ The Figure 3.7 is partly inspired by the depictions in [166].

[^32]:    ${ }^{22}$ Let us comprehend an unwanted pole as a spurious pole in the sense that it captures a singularity that does not correspond to a non-essential singularity of the high-fidelity model.

[^33]:    ${ }^{23}$ In the case of a Chebyshev polynomial, the common k-fold cross-validation method breaks the regular pattern of the Chebyshev grid (recall Figure A.1).

[^34]:    ${ }^{24}$ For more details on the implementation of the NMS algorithm from (Opkg1) and the NMS algorithm from (Opkg3), I refer to the respective package documentation and references therein.

[^35]:    ${ }^{25}$ For computational reasons in a manner similar to (3.97), it can be useful to add the smallest positive normalized floating-point number $2^{-1022}$ to the assignment in (3.114) and to consider the logarithm with base 10 of this extended version, i.e., to utilize rather $\log _{10}\left(E I(x)+2^{-1022}\right)$ than $E I(x)$ itself (see, e.g., the book website associated with [70]).

[^36]:    ${ }^{26}$ In [138], multivariate functions are examined in a computational context for investigating partial derivatives and the corresponding chain rule for multivariate calculus.

[^37]:    ${ }^{27}$ If we set $\left[C^{1}(U, \mathbb{R})\right]^{m}:=C^{1}(U, \mathbb{R}) \times \cdots{ }_{m-1} \times C^{1}(U, \mathbb{R})$, and if we assume a map $f \in\left[C^{1}(U, \mathbb{R})\right]^{m}$ with $U \subset \mathbb{R}^{n}$ being an open set, and $f_{i} \in C^{1}(U, \mathbb{R})$ with $i \in\{1, \ldots, m\}$ denote the components of the map $f$, and $x_{j} \in \mathbb{R}$ with $j \in\{1, \ldots, n\}$ denote the components of the component maps $f_{i}$, and if we suppose a fixed argument $p \in U$, then let us conceive the Jacobi matrix $J_{f}(p) \in \mathbb{R}^{m \times n}$ w.r.t. $f$ and evaluated at $p$ as $J_{f}(p):=\left[j_{f}(p)_{i, j}\right]$ with $j_{f}(p)_{i, j}:=\partial_{\mathbf{e}_{x_{j}}}\left(f_{i}\right)(p)$ for each $i \in\{1, \ldots, m\}$ and for each $j \in\{1, \ldots, n\}$ with $\partial_{\mathbf{e}_{x_{j}}}\left(f_{i}\right)(p):=\operatorname{grad}\left(f_{i}\right)(p) \cdot \mathbb{R}^{n} \mathbf{e}_{x_{j}}$ where $\cdot \mathbb{R}^{n}$ denotes the Euclidean inner product w.r.t. $\mathbb{R}^{n}$; and $\mathbf{e}_{x_{j}}$ refer to the unit vectors w.r.t. $x_{j}$, respectively. See the commentary on the overloading of the map grad in $\S$ 2.3.3 as well.

[^38]:    ${ }^{28}$ The definition of the $k$-th step $h^{(k)}$ in (3.155) aligns itself with the approach in [95]. However, in [56, p. 18f] and in [49, p. 68f], another approach for computing $h^{(k)}$ is discussed within the context of the Levenberg-Marquardt method for least-squares problems (see, e.g., [158, p. 258-262]). Hence, $h^{(k)}$ is determined by solving the linear system of equations $\left(\left(B^{(k)}\right)^{\mathrm{T}} B^{(k)}+\lambda I\right) h^{(k)}=-\left(B^{(k)}\right)^{\mathrm{T}} e^{(k)}$ with $e^{(k)}:=\tilde{\mathrm{P}}_{v}\left(x^{k}\right)-\tilde{x}^{*}$ where $\tilde{x}^{*}$ refers to an existing optimal solution of the low-fidelity optimization problem and $\lambda \in \mathbb{R}^{+}$plays a similar role as the regularization parameter w.r.t. (3.65). For more details on this approach for computing $h^{(k)}$, I refer to, e.g., [56, p. 18f], [49, p. 68f], [158, p. 69ff], and [158, p. 258-262] and references therein.

[^39]:    ${ }^{29}$ Since some kind of sub-structure is supposed between $X_{0}$ and $X_{1}$ as well as between $Y_{1}$ and $Y_{0}$ such as in (3.122), one can denote $\left.\tilde{\mathrm{K}}\right|_{X_{0}} \in \operatorname{hom}\left(X_{0}, Y_{1}\right)$ as the restriction of $\tilde{\mathrm{K}}$ to $X_{0}$ at the function level (recall Figure 1.4). Furthermore, the map $\left.\tilde{K}\right|_{X_{0}}$ can be composed with $\tau_{\tilde{R}}$ such that one can construct the map $\left.\iota_{\tilde{\mathrm{R}}}{ }^{\circ} Y_{10} \tilde{\mathrm{~K}}\right|_{X_{0}} \in \operatorname{hom}\left(X_{0}, Y_{0}\right)$. I argue, therefore, that it is reasonable to conceive the column vector $\left.\tilde{y}\right|_{X_{\mathrm{s}}}$ within the context of the map $\left.\iota_{\tilde{R}}{ }^{\circ} Y_{10} \tilde{\mathrm{~K}}\right|_{X_{0}}$. Keep in mind that if we solely operate with various forms of the real numbers such as, e.g., $\mathbb{R}, \mathbb{R}^{n}$, and $\mathbb{R}^{n \times m}$ with $n, m \in \mathbb{N}$, then a lot of valuable conceptional distinction is probably lost.
    ${ }^{30}$ The operation \is overloaded with the signature $\mathbb{R}^{m_{\bar{\kappa}} \times d} \times \mathbb{R}^{m \times d} \rightarrow \mathbb{R}^{\left(m_{\bar{\kappa}}-m\right) \times d}$ where a resulting difference sampling plan $X_{\mathbf{s}_{\bar{K}}} \backslash X_{\mathbf{s}}$ contains all the sampling points of $X_{\mathbf{s}_{\bar{K}}}$ that are not contained in $X_{\mathbf{s}}$.

[^40]:    ${ }^{31}$ Regarding some applications, there might be numerical issues that are presumably caused predominantly by the estimate $\hat{\rho}$. In order to mitigate such potential numerical issues, it is advisable to round the numerical value associated with the estimate $\hat{\rho}$. However, an in-depth analysis of the propagation of, e.g., the corresponding round-off error is out of the scope of the present work.

[^41]:    ${ }^{1}$ In [138], the Agda programming language is partially used for investigating partial derivatives and the corresponding chain rule for multivariate calculus.

[^42]:    ${ }^{2}$ Expressing formally the notion of natural transformation for applications in the field of algebraic topology is mostly considered as the starting point of category theory (see, e.g., [177, p. 1f]).

[^43]:    ${ }^{3}$ Regarding the signature of F , see Remark § 4.2.4.

[^44]:    ${ }^{4}$ For some applications of $\lambda$-calculus with connection to electromagnetics, I refer to, e.g., [138].

[^45]:    ${ }^{5}$ Regarding the signature of F , see Remark § 4.2.4.

[^46]:    ${ }^{6}$ Notice well that, compared with (3.124), there is a slight semantical change regarding the notation in (4.34) in order to harmonize the notation a bit with the notation w.r.t. the CT toolset (recall § 4.2). However, the identification of the entities in (3.124) with the entities in (4.34) should be clear.

[^47]:    ${ }^{7}$ The elaborations are part of a joint publication in preparation for submission called "Formalization Issues of Surrogate Modeling in Electromagnetic Compatibility" (M. Hadžiefendić, R. S. Rezende, R. Schuhmann).

[^48]:    ${ }^{8}$ Technically, an $S_{i j}$-parameter is defined as a member of the complex numbers, i.e., $S_{i j} \in \mathbb{C}$. However, I utilize the common abbreviated interpretation in which the term $S_{i j}$ refers to the magnitude of the corresponding $S_{i j}$-parameter. And since, roughly speaking, an $S_{i j}$-parameter encodes the ratio of waves, it is usual to conceive the term $S_{i j}$ in the relative unit of measurement decibel (dB). Therefore, I interpret the term $S_{i j}$ as a member of the real numbers, i.e., $S_{i j} \in \mathbb{R}$.
    ${ }^{9}$ Respecting the notation in (3.131), the entry $S_{i j \mathrm{k}}$ has the signature $Y_{0}^{X_{0}}$, i.e., $S_{i j \mathrm{~K}}: Y_{0}^{X_{0}}$, where the matrix $S_{\mathrm{K}}:=\left[s_{i, j}\right]_{\mathrm{K}} \in\left(Y_{0}^{X_{0}}\right)^{4 \times 4}$ and $S_{i j \mathrm{~K}}:=\left[s_{i, j}\right]_{\mathrm{K}}$. Analogously, the entry $S_{i j_{\mathrm{K}}}$ has the signature $Y_{0}^{X_{0}}$, i.e., $S_{i j_{\tilde{\mathrm{K}}}}: Y_{0}^{X_{0}}$, where the matrix $S_{\tilde{\mathrm{K}}}:=\left[s_{i, j}\right]_{\tilde{\mathrm{K}}} \in\left(Y_{0}^{X_{0}}\right)^{4 \times 4}$ and $S_{i j_{\mathrm{K}}}:=\left[s_{i, j}\right]_{\tilde{\mathrm{K}}}$.

[^49]:    ${ }^{10}$ For more details regarding expressions such as in (4.64), I refer to, e.g., [46, p. 173-177] and references therein.

[^50]:    ${ }^{1}$ Mind that parts of the corresponding simulation code at an early stage have been developed during a bachelor thesis called "Numerische Optimierung in der Magnetoquasistatik mit dem Space Mapping Ansatz" [in English: "Numerical Optimization in Magnetoquasistatics using the Space Mapping Approach"] (Albert Piwonski, summer term 2017; unpublished) under my scientific supervision and reviewed by the first reviewer of the present work and Prof. Dr.-Ing. Ronald Plath (TU Berlin).

[^51]:    ${ }^{2}$ For more details on computations w.r.t. magnetic fields in the magnetoquasistatic model, I refer to, e.g., [103, p. 218-224] and references therein.

[^52]:    ${ }^{3}$ Even if some potential physical intuition is available concerning a suitable candidate for an initial point, such a worst-case scenario is highly probable in practical applications where the shape of the admissible set of solutions and the landscape of the evaluated objective functional are usually unknown.

[^53]:    ${ }^{4}$ A strategy to determine a feasible point regarding the optimization problem in (5.17) is, e.g., to solve an optimization problem where the evaluated objective function is defined as $\| \hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)$ $L_{\min } \|_{l_{2}}^{2}$ or as $\left\|\hat{Q}_{L, \omega_{0}}\left(x_{1}, x_{2}\right)-L_{\max }\right\|_{l_{2}}^{2}$ and the constraints are defined by (5.17b) and (5.17c).

[^54]:    ${ }^{5}$ In [123], the authors investigate among others the relationship between the SSPCC and the degree of the grid discretization in the context of antenna design. However, mind that an in-depth examination of relationships such as, e.g., the relationship between $\mathcal{T}_{h_{1}}^{2 D}$ and $\left.\bar{r}_{\hat{y} \tilde{y}}^{2}\right|_{k:=5}$ in the light of re-meshing (see (5.9)), is out of the scope of the present work.

[^55]:    ${ }^{6}$ Mind that parts of the corresponding simulation code at an early stage have been developed during a student project called "Spulen-Optimierung mit Ersatzmodellen" [in English: "Coil optimization with surrogate models"] (Marie Krause, Mandy Domke, winter term 2017/2018; unpublished) under my scientific supervision and reviewed by the first reviewer of the present work.
    ${ }^{7}$ Mind that parts of the corresponding simulation code at an early stage have been developed during a master thesis called "Multi-fidelity modeling for electromagnetic compatibility problems" (Rodrigo Silva Rezende, summer term 2020; unpublished) under my scientific supervision and the scientific supervision of Dr. Jan Hansen (Robert Bosch GmbH) and reviewed by the first reviewer of the present work and Prof. Dr. Stefan Kurz (TU Darmstadt). Moreover, the example in Figure 5.4 is part of a series of numerical studies regarding joint publications in preparation for submission called "VectorValued Multi-Fidelity Surrogate Modeling for Microwave Components Design" (R. S. Rezende, M.

[^56]:    Hadžiefendić, R. Schuhmann) and "Multi-Output Variable-Fidelity Bayesian Optimization of a Common Mode Choke" (R. S. Rezende, M. Hadžiefendić, J. Hansen, R. Schuhmann).

[^57]:    ${ }^{8} \mathrm{I}$ conceive $S_{22}\left(\omega, x_{1}, x_{2}, x_{3}\right)$ as the magnitude in dB of the corresponding $S_{22}\left(\omega, x_{1}, x_{2}, x_{3}\right)$ parameter whereas I conceive $Z_{1}(\omega)$ and $Z_{2}(\omega)$ as the magnitude in $\Omega$ of the corresponding impedance $\underline{Z}_{1}(\omega)$ and $\underline{Z}_{2}(\omega)$, respectively. Hence, it is assumed that suitable maps exist in order to compare the shapes of the entities $S_{22}\left(\omega, x_{1}, x_{2}, x_{3}\right), Z_{1}(\omega)$ and $Z_{2}(\omega)$.

[^58]:    ${ }^{9}$ The example in Figure 5.5 is part of a series of numerical studies regarding a joint publication in preparation for submission called "Surrogate-guided Optimization based on the Space-Mapping Paradigm and the Co-Kriging Approach with Application in Electromagnetic Compatibility" (M. Hadžiefendić, R. S. Rezende, R. Schuhmann).

[^59]:    ${ }^{1}$ Given $m=1000$, the time needed to construct the sampling plan based on the Sobol quasi-random sequence is approximately 198 s for $d=2$ and 227 s for $d=3$ on a notebook with an Intel ${ }^{\circledR}$ Core $^{\mathrm{TM}} \mathrm{i}^{7}$ 6500 U CPU @ 2.50 GHz . This time represents the main bottleneck during the construction of the matrix $\left(B^{\mathrm{T}} B+\lambda I\right)$ and $\left(B_{\zeta}^{\mathrm{T}} B_{\zeta}+\lambda I\right)$, respectively.

[^60]:    ${ }^{1}$ It is supposed that the identification $\operatorname{abs}(\underline{Z}) \equiv Z(\omega)$ is given where abs refers to the single-valued absolute value function with the signature $\mathbb{C}^{\mathbb{R}^{+}} \rightarrow \mathbb{R}^{+}$and it is supposed that the identification $\arg (\underline{Z}) \equiv \theta(\omega)$ is given where $\arg$ refers to the single-valued argument function with the signature $\mathbb{C}^{\mathbb{R}^{+}} \rightarrow \mathbb{R}^{+}$.

