Thesis for the Degree of Doctor of Philosophy

Numerical Analysis of Evolution Problems in Multiphysics

Anna Persson

Division of Mathematics
Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden, 2018

Numerical Analysis of Evolution Problems in Multiphysics
Anna Persson
ISBN 978-91-7597-713-3

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4394
ISSN 0346-718X

Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg
SE–412 96 Gothenburg
Sweden
Phone: +46 (0)31-772 10 00

Printed in Gothenburg, Sweden, 2018.

Abstract

In this thesis we study numerical methods for evolution problems in multiphysics. The term multiphysics is commonly used to describe physical phenomena that involve several interacting models. Typically, such problems result in coupled systems of partial differential equations.

This thesis is essentially divided into two parts, which address two different topics with applications in multiphysics. The first topic is numerical analysis for multiscale problems, with a particular focus on heterogeneous materials, like composites. For classical finite element methods such problems are known to be numerically challenging, due to the rapid variations in the data.

One of our main goals is to develop a numerical method for the thermoelastic system with multiscale coefficients. The method we propose is based on the localized orthogonal decomposition (LOD) technique introduced in [17]. This is performed in three steps, first we extend the LOD framework to parabolic problems (Paper I) and then to linear elasticity equations (Paper II). Using the theory developed in these two papers we address the thermoelastic system (Paper III).

In addition, we aim to extend the LOD framework to differential Riccati equations where the state equation is governed by a multiscale operator. The numerical solution of such problems involves solving many parabolic equations with multiscale coefficients. Hence, by applying the method developed in Paper I to Riccati equations the computational gain may be significantly large. In this thesis we show that this is indeed the case (Paper IV).

The second part of this thesis is devoted to the Joule heating problem, a coupled nonlinear system describing the temperature and electric current in a material. Analyzing this system turns out to be difficult due to the low regularity of the nonlinear term. We overcome this issue by introducing a new variational formulation based on a cut-off functional. Using this formulation, we prove (Paper V) strong convergence of a large class of finite element methods for the Joule heating problem with mixed boundary conditions on nonsmooth domains in three dimensions.

Keywords: Thermoelasticity, parabolic equations, linear elasticity, Riccati equations, multiscale, generalized finite element, localized orthogonal decomposition, Joule heating, thermistor, finite element method, regularity.

List of included papers

The following papers are included in this thesis:

- **Paper I.** Axel Målqvist and Anna Persson: *Multiscale techniques for parabolic equations*, Numer. Math. 138 (2018), no. 1, pp. 191-217.

- **Paper II.** Patrick Henning and Anna Persson: *A multiscale method for linear elasticity reducing Poisson locking*, Comput. Methods Appl. Mech. Engrg., 310 (2016), pp. 156-171.

- **Paper III.** Axel Målqvist and Anna Persson: *A generalized finite element method for linear thermoelasticity*, ESAIM Math. Model. Numer. Anal. 51 (2017), no. 4, pp. 1145–1171.

- **Paper IV.** Axel Målqvist, Anna Persson, and Tony Stillfjord: *Multiscale differential Riccati equations for linear quadratic regulator problems*, Preprint (Submitted).

- **Paper V.** Max Jensen, Axel Målqvist, and Anna Persson: *Finite element convergence for the time-dependent Joule heating problem with mixed boundary conditions*, Preprint (Submitted).

Contribution:

- **Paper I, II, III, and V.** The author of this thesis had the main responsibility to write and prepare the manuscript, performed most of the analysis, made the necessary implementations and performed all numerical experiments. The ideas were developed in close collaborations between the authors.
- **Paper IV.** The author of this thesis contributed mainly to the error analysis of the method. The ideas were developed in close collaboration between the authors.

Acknowledgments

First of all I would like to thank my supervisor Axel Målqvist for his guidance throughout my work on this thesis. Thank you for introducing me to the subject, for sharing your expertise and knowledge, and for teaching me how to do research. I would also like to thank my co-advisors Stig Larsson and Patrick Henning for many fruitful discussions and your helpful advice.

I wish to thank all my co-authors, Patrick Henning, Max Jensen, Axel Målqvist, and Tony Stillfjord, for great collaborations and for all the valuable discussions during our joint work. Also, thank you Daniel Peterseim for introducing me to many interesting problems and for inviting me to Bonn.

Furthermore, I am grateful to my colleagues at the mathematics department for creating such a friendly working environment. Thank you all!

Since I moved to Stockholm in 2016, I have also spent a lot of time at the mathematics department at KTH. Many thanks to everyone there for arranging an extra office space for me and making me feel very welcome at your department.

Many thanks also goes to my family for always believing in me and giving me encouragement when I need it.

Last, but not least, thank you Joakim for your love and immense support.

Contents

Part 1

Introduction

CHAPTER 1

Multiscale methods

1.1. Background

In many applications it is of great importance to understand how different materials interact and respond to external forces and temperature changes. For instance, it may be crucial when designing parts for aircrafts or when constructing a bridge.

In this thesis we study numerical solutions to partial differential equations (PDEs) describing displacement and temperature changes in materials over time. In particular, we are interested in applications where the material under consideration is strongly heterogeneous, e.g. composites. Composite materials are constructed using two or more different constituents with different physical properties. Typically, the material properties vary on a very fine scale, as in, for instance, fiber reinforced materials. Modeling physical behavior in these materials results in equations with highly varying and oscillating coefficients. Such problems, that exhibit a lot of variations in the data, often on multiple scales, are commonly referred to as *multiscale* problems.

One of the most common methods to obtain numerical solutions to PDEs is the finite element method (FEM) based on continuous piecewise polynomials. These methods work well for homogeneous media or media that are not varying too much in space. However, for highly varying media, like composite materials, the classical FEMs struggle to approximate the solution accurately unless the mesh width is sufficiently small. Indeed, the mesh width must be small enough to resolve all the fine variations in the data. In practice, this leads to issues with computational cost and available memory.

Today's increasing interest in, and usage of, composite materials thus pose a demand for other types of numerical methods. Several such methods have been proposed over the last two decades, see, for instance, [12, 8, 3, 13]. However, the analysis of many of these methods requires restrictive assumptions on the data, such as periodicity or separation of scales.

In [17] a generalized finite element method (GFEM), cf. [4], is proposed and analyzed. Convergence of the method is proven for an arbitrary positive and bounded coefficient, that is, no assumptions on periodicity or separation of scales are needed. The method is often referred to as *localized orthogonal decomposition* (LOD).

The purpose of this thesis is to generalize the method proposed in [17] to solve time dependent PDEs, with highly varying and oscillating coefficients. The main focus is on equations describing temperature and displacement in materials.

In Paper I we extend the method to parabolic problems which can be used to model the evolution of the temperature in a material over time. In Paper II we consider (stationary) linear elasticity equations describing the displacement in an elastic body. In Paper III we combine the results in Paper I and Paper II to address the thermoelastic system, a coupled multiphysics system modeling the interaction between temperature and displacement in a material. Finally, in Paper IV we consider differential Riccati equations (DREs) used to solve linear quadratic regulator (LQR) problems. The LQR problem can, for instance, be used to model problems where the temperature in a material is subject to a control input. In all four papers we prove convergence of optimal order, except for a logarithmic factor in the Riccati case, for highly varying coefficients and we provide several numerical examples that confirm the analysis.

In the upcoming section we provide some background on the thermoelastic system and the Riccati equation and define their respective variational formulations. In Section 1.3 the issue with applying the classical FEM to multiscale problems is described in more detail. In Section 1.4 we introduce the GFEM proposed in [17] for elliptic equations and discuss the main idea behind the extension to linear thermoelasticity and Riccati equations. Finally, in Section 1.5 we summarize the appended papers and highlight the main results.

1.2. General setting and notation

Throughout this chapter Ω denotes a domain in $\mathbb{R}^d$, for $d = 1, 2$, or 3. We use $(\cdot, \cdot)$ to denote the inner product in $L_2(\Omega)$ and $\|\cdot\|$ the corresponding norm. Let $H^1(\Omega) := W_2^1(\Omega)$ denote the classical Sobolev space with norm $\|v\|_{H^1(\Omega)}^2 = \|v\|^2 + \|\nabla v\|^2$ and let $H_0^1(\Omega)$ denote the functions in $H^1(\Omega)$ that vanish on the boundary $\partial\Omega$. We also use the notation $H^{-1}(\Omega)$ to denote the dual space to $H_0^1(\Omega)$. We refer to [2] for further details on Sobolev spaces.

Furthermore, let $L_p([0,T];X)$ denote the Bochner space with norm

$$\|v\|_{L_p([0,T];X)} = \Big(\int_0^T \|v\|_X^p \,\mathrm{dt} \Big)^{1/p}, \quad 1 \le p < \infty,$$
$$\|v\|_{L_\infty([0,T];X)} = \operatorname*{ess\,sup}_{0 \le t \le T} \|v\|_X,$$

where X is a Banach space equipped with the norm $\|\cdot\|_X$. The dependence on the interval $[0,T]$ and the domain Ω is frequently suppressed and we write, for instance, $L_2(L_2)$ for $L_2([0,T];L_2(\Omega))$. We also use the double-dot product notation to denote the Frobenius inner product of two matrices A and B

$$A : B = \sum_{i,j=1}^{d} A_{ij} B_{ij}, \quad A, B \in \mathbb{R}^{d \times d}.$$

Finally, we use $\mathcal{L}(X, Y)$ to denote the space of linear bounded operators from a Hilbert space X to another Hilbert space Y. The notation $\|\cdot\|_{\mathcal{L}(X,Y)}$ denotes the corresponding operator norm.

1.2.1. Linear thermoelasticity. Linear thermoelasticity refers to a coupled system of PDEs describing the displacement and temperature of an elastic body, see [6, 7]. To introduce the mathematical formulation of this system we let Ω describe the initial configuration of an elastic medium. For a given simulation time $T > 0$, we let the vector valued function $u : [0, T] \times \Omega \to \mathbb{R}^d$ denote the displacement field and $\theta : [0, T] \times \Omega \to \mathbb{R}$ denote the temperature. To define boundary conditions for u we let Γ_D^u and Γ_N^u be two disjoint parts of the boundary such that $\Gamma_D^u \cup \Gamma_N^u = \partial\Omega$, where $\partial\Omega$ denotes the boundary of Ω. On the part denoted Γ_D^u we impose Dirichlet boundary conditions corresponding to a clamped part of the material. On Γ_N^u, corresponding to the traction boundary, we impose Neumann boundary conditions. Similarly, we define Γ_D^θ and Γ_N^θ to be the drained and flux part of the boundary for the temperature θ.

Under the assumption that the displacement gradients are small, the strain tensor is given by the following linear relation

$$\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^\intercal).$$

For isotropic materials, the total stress tensor is given by

$$\bar{\sigma} = 2\mu\varepsilon(u) + \lambda(\nabla \cdot u)I - \alpha\theta I,$$

where I is the d-dimensional identity matrix and α is the thermal expansion coefficient. The first part of $\bar{\sigma}$, involving u, represents the mechanical stress and the second part, involving θ, represents the thermal stress.

Furthermore, μ and λ denote the Lamé coefficients satisfying

$$\mu = \frac{E}{2(1+\nu)}, \quad \lambda = \frac{E\nu}{(1+\nu)(1-2\nu)},$$

where ν denotes Poisson's ratio and E denotes Young's elastic modulus. Poisson's ratio is a measure on the materials tendency to shrink (expand) when stretched (compressed) and Young's modulus describes the stiffness of the material. The coefficients α, λ, and μ, are all material dependent and thus rapidly varying in space for strongly heterogeneous (multiscale) materials.

Cauchy's equilibrium equations states that

$$-\nabla \cdot \bar{\sigma} = f,$$

where $f : \Omega \to \mathbb{R}^d$ denotes the external body forces. Furthermore, the temperature in the material can be described by the parabolic equation

$$\dot{\theta} - \nabla \cdot \kappa\nabla\theta + \alpha\nabla \cdot \dot{u} = g,$$

where $\kappa : \Omega \to \mathbb{R}^{d\times d}$ is the heat conductivity parameter and g denotes internal heat sources. The term $\alpha\nabla \cdot \dot{u}$ corresponds to the internal heating due to the dilation rate. Note that also κ is material dependent and thus rapidly varying.

To summarize, the linear thermoelastic system is given by the following system of equations

$$-\nabla \cdot (2\mu\varepsilon(u) + \lambda\nabla \cdot uI - \alpha\theta I) = f, \quad \text{in } (0,T] \times \Omega, \tag{1.2.1a}$$

$$\dot{\theta} - \nabla \cdot \kappa\nabla\theta + \alpha\nabla \cdot \dot{u} = g, \quad \text{in } (0,T] \times \Omega, \tag{1.2.1b}$$

$$u = 0, \quad \text{on } (0,T] \times \Gamma_D^u, \tag{1.2.1c}$$

$$\bar{\sigma} \cdot n = 0, \quad \text{on } (0,T] \times \Gamma_N^u. \tag{1.2.1d}$$

$$\theta = 0, \quad \text{on } (0,T] \times \Gamma_D^\theta, \tag{1.2.1e}$$

$$\kappa\nabla\theta \cdot n = 0, \quad \text{on } (0,T] \times \Gamma_N^\theta. \tag{1.2.1f}$$

$$\theta(0) = \theta_0, \quad \text{in } \Omega, \tag{1.2.1g}$$

where we for simplicity assume homogeneous boundary conditions. Note that the equations (1.2.1a)-(1.2.1b) are coupled. In [19] a comprehensive review of the literature on this system, and similar versions of it, is given.

Remark 1.2.1. The system (1.2.1) is formally equivalent to a linear model for poroelasticity, see, e.g., [19, 22, 3]. In this case θ denotes the fluid pressure, κ the hydraulic conductivity, and α the Biot-Willis coupling-deformation coefficient. Hence, the results in this thesis also apply to the linear poroelastic system.

To define a FEM (and a GFEM) for (1.2.1) we define the corresponding variational (or weak) formulation. For this purpose define the following spaces

$$V^1 := \{v \in (H^1(\Omega))^d : v = 0 \text{ on } \Gamma_D^u\}, \quad V^2 := \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_D^\theta\}.$$

Multiplying (1.2.1a) with $v_1 \in V^1$ and (1.2.1b) with $v_2 \in V^2$ and using Green's formula together with the boundary conditions (1.2.1c)-(1.2.1f) we arrive at the following variational formulation; find $u(t,\cdot) \in V^1$ and $\theta(t,\cdot) \in V^2$ such that, for a. e. $t > 0$,

$$(\sigma(u) : \varepsilon(v_1)) - (\alpha\theta, \nabla \cdot v_1) = (f, v_1), \quad \forall v_1 \in V^1, \tag{1.2.2}$$

$$(\dot{\theta}, v_2) + (\kappa\nabla\theta, \nabla v_2) + (\alpha\nabla \cdot \dot{u}, v_2) = (g, v_2), \quad \forall v_2 \in V^2, \tag{1.2.3}$$

and the initial value $\theta(0,\cdot) = \theta_0$ is satisfied. Here $\sigma(u) := 2\mu\varepsilon(u) + \lambda\nabla \cdot uI$ is the first (mechanical) part of $\bar{\sigma}$ involving only the displacement u.

Two functions u and θ are weak solutions if (1.2.2)-(1.2.3) are satisfied and $u \in L_2(V^1)$, $\nabla \cdot \dot{u} \in L_2(H^{-1})$, $\theta \in L_2(V^2)$, and $\dot{\theta} \in L_2(H^{-1})$. Existence and uniqueness of such weak solutions are proved in, e.g., [22, 21], and in [19] within the framework of linear degenerate evolution equations in Hilbert spaces. In [19] it is also proved that the system is of parabolic type, meaning that it is well posed for nonsmooth initial data with regularity estimates depending on negative powers of t.

1.2.2. Riccati equations. Differential Riccati equations (DREs) typically arise when solving linear quadratic regulator (LQR) problems. The LQR

problem is a common problem in optimal control theory and has a wide range of applications.

In LQR problems, the goal is to control the output y given a state x of a system whose evolution may be influenced through the input u. The relation between these quantities is given by the state and output equations

$$\dot{x} = \mathcal{A}x + \mathcal{B}u, \quad x(0) = x_0, \tag{1.2.4}$$

$$y = \mathcal{C}x, \tag{1.2.5}$$

where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are given operators. The output y denotes a measurable quantity, specified by $\mathcal{C}$, of the system. The goal is to minimize the following cost functional

$$J(u) = \int_0^T (\mathcal{Q}y, y) + (\mathcal{R}u, u)\,\mathrm{d}t + (\mathcal{G}y(T), y(T)), \tag{1.2.6}$$

where $\mathcal{Q}$, $\mathcal{R}$, and $\mathcal{G}$, are given weighting operators. The first term in (1.2.6) penalizes the output, the second term penalizes the control effort, and the third term penalizes the final state of the output. It can be proved (see e.g. [1], [16]) that the optimal input u^* is given by $u^* = -\mathcal{R}^{-1}\mathcal{B}^* X(T-t)x(t)$ where X solves the operator-valued DRE

$$\dot{X}(t) = \mathcal{A}^* X(t) + X(t)\mathcal{A} + \mathcal{C}^*\mathcal{Q}\mathcal{C} - X(t)\mathcal{B}\mathcal{R}^{-1}\mathcal{B}^* X(t), \tag{1.2.7}$$

$$X(0) = \mathcal{G}. \tag{1.2.8}$$

In the context of materials, the input may correspond to a heat source applied to (parts of) the domain or the boundary and the output a measurable quantity, like the average of the temperature. In multiscale applications it is typically the operator $\mathcal{A}$ that exhibits multiscale features. This is the case if the state equation x models physical behavior in a heterogeneous material, such as a composite.

Let $V = H_0^1(\Omega)$ and define the domain of $\mathcal{A}$ by $\mathcal{D}(\mathcal{A}) = \{u \in V | \mathcal{A}u \in L_2\}$. In Paper IV we consider settings where the operator $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to L_2$ is given by $(\mathcal{A}u, v) = -a(u, v)$ and $a(u, v) = \int \kappa \nabla u \cdot \nabla v$. Here κ may describe, for instance, the conductivity of a composite material.

Furthermore, we let U and Z denote the control and observation space, respectively, such that $\mathcal{B} : U \to L_2$, $\mathcal{C} : L_2 \to Z$, $\mathcal{Q} : Z \to Z$, $\mathcal{R} : U \to U$, and for the final state operator $\mathcal{G} : L_2 \to L_2$. The notation $(\cdot, \cdot)_U$ and $(\cdot, \cdot)_Z$ are used for the corresponding inner products. In this notation, the weak form of (1.2.7)-(1.2.8) is to find $X \in \mathcal{L}(L_2, L_2)$ such that

$$\left(\dot{X}x, y\right) = (Xx, \mathcal{A}y) + (Xy, \mathcal{A}x) + (\mathcal{Q}\mathcal{C}x, \mathcal{C}y)_Z - \left(\mathcal{R}^{-1}\mathcal{B}^* Xx, \mathcal{B}^* Xy\right)_U, \tag{1.2.9}$$

for all $x, y \in \mathcal{D}(\mathcal{A})$.

If $\mathcal{A}$ generates a strongly continuous semigroup $\mathrm{e}^{t\mathcal{A}}$ on L_2 and the involved operators are bounded, then there exists a unique solution to (1.2.9), see [5, Part IV, Ch. 1, Theorem 2.1]. In addition, the solution X is self-adjoint. If $\mathcal{A}$ generates an analytic semigroup $\mathrm{e}^{t\mathcal{A}}$ on L_2, as in Paper IV, then X is also a solution in a classical sense, see [5, Part IV, Ch. 1, Theorem 3.1]. In particular,

this means that $\mathcal{A}^*X + X\mathcal{A}$ is a well defined operator in $\mathcal{L}(L_2, L_2)$ and X satisfies (1.2.7)-(1.2.8).

1.3. Classical finite element

In this section we explain more carefully why the classical FEM fails to approximate the solution to problems with rapidly varying data. To simplify the discussion we start by considering elliptic equations of Poisson type.

1.3.1. Elliptic equations. Consider the elliptic equation

$$\begin{aligned} -\nabla \cdot A\nabla u &= f, \quad \text{in } \Omega, \\ u &= 0, \quad \text{on } \partial\Omega, \end{aligned}$$

with the variational formulation; find $u \in V$, such that

$$a(u, v) = (f, v), \quad \forall v \in V, \tag{1.3.1}$$

where $V = H_0^1(\Omega)$ and $a(u, v) := (A\nabla u, \nabla v)$. Here the diffusion coefficient $A : \Omega \to \mathbb{R}^{d\times d}$ is assumed to be rapidly oscillating. We also assume that A is symmetric and bounded such that $\alpha\|v\|_{H^1}^2 \leq a(v, v) \leq \beta\|v\|_{H^1}^2$, $\forall v \in V$, for some positive constants α, β. In particular, this means that $a(\cdot, \cdot)$ defines an inner product on V.

To define a FEM we need a triangulation of the domain. Let $\{\mathcal{T}_h\}_{h>0}$ be a family triangulations of Ω with the mesh size $h_K := \operatorname{diam}(K)$, for $K \in \mathcal{T}_h$ and denote the largest diameter in the triangulation by $h := \max_{K\in\mathcal{T}_h} h_K$. Now let $V_h \subseteq V$ denote the space of continuous piecewise affine functions on the triangulation $\mathcal{T}_h$. The finite element formulation then reads; find $u_h \in V_h$, such that,

$$a(u_h, v) = (f, v), \quad \forall v \in V_h. \tag{1.3.2}$$

Classical a priori error analysis gives the bound

$$\|u_h - u\|_{H^1} \leq Ch\|D^2u\|, \tag{1.3.3}$$

where D^2u denotes the second order (weak) derivatives of u. Not only does this bound require additional regularity of the solution, the norm $\|D^2u\|$ may also be very large if A is rapidly oscillating. Indeed, if A varies with frequency ϵ^{-1} for some $\epsilon > 0$, then, typically, $\|D^2u\| \sim \epsilon^{-1}$ and we need $h < \epsilon$ in (1.3.3) to obtain accurate approximations.

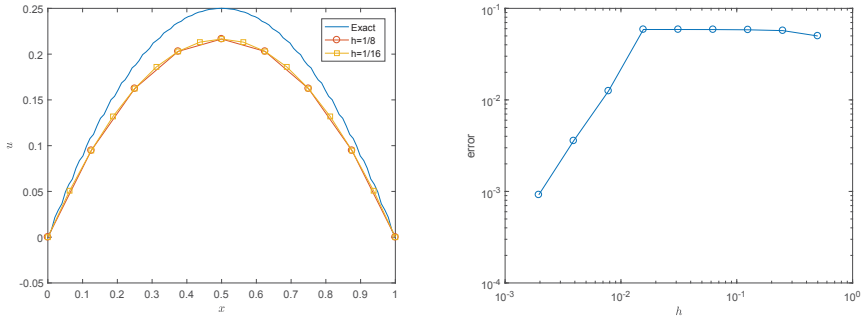
To illustrate this, consider the following one-dimensional problem, suggested in [18], where $\Omega = [0, 1]$, $A = (2 - \cos(2\pi x\epsilon^{-1}))^{-1}$, and $f = 1$, in (1.3.1). The solution is given by

$$u = 4(x - x^2) - \frac{\epsilon}{2\pi}\left(\frac{1}{2}\sin(2\pi x\epsilon^{-1}) - x\sin(2\pi x\epsilon^{-1}) - \frac{\epsilon}{2\pi}\cos(2\pi x\epsilon^{-1}) + \frac{\epsilon}{2\pi}\right).$$

Differentiating twice with respect to x gives $|u''| \sim \epsilon^{-1}$.

For $\epsilon = 2^{-5}$, the true solution and the corresponding FEM solutions for different mesh sizes are plotted in Figure 1 (left). We clearly see that the FEM approximations struggle to approximate u for coarse mesh sizes. The

convergence plot in Figure 1 (right) shows that convergence does not take place until h resolves the data, that is, when $h \leq \epsilon$.



(A) FEM approximations (B) Convergence

FIGURE 1. FEM approximations (left) with relative errors in the H^1-norm (right) of a 1D-problem with rapidly varying data.

1.3.2. Parabolic equations. Consider a parabolic problem on the following weak form; find $u(t) \in V$, such that, $u(0) = u_0$ and

$$(\dot{u}, v) + a(u, v) = (f, v), \quad \forall v \in V, \tag{1.3.4}$$

where $a(u, v) = (A\nabla u, \nabla v)$ as in the elliptic equation (1.3.1). The diffusion coefficient $A : \Omega \to \mathbb{R}^{d\times d}$ is assumed to not depend on time.

Let $\{\mathcal{T}_h\}_{h>0}$ and V_h denote the same triangulation and finite element space as in Subsection 1.3.1. Furthermore, let $0 = t_0 < t_1 < ... < t_N = T$ be a uniform discretization of the time interval such that $t_j - t_{j-1} = \tau > 0$ for $j = 1, ..., N$.

The classical FEM for (1.3.4) with a backward (implicit) Euler discretization in time reads; for $n \in \{1, ..., N\}$ find $u_h^n \in V_h$, such that, $u_h^0 = u_{h,0}$

$$(\bar{\partial}_t u_h^n, v) + a(u_h^n, v) = (f^n, v), \quad \forall v \in V_h, \tag{1.3.5}$$

where $\bar{\partial}_t u_h^n := (u_h^n - u_h^{n-1})/\tau$ and $u_{h,0}$ is a suitable approximation of u_0. The right hand side is evaluated at time t_n, that is, $f^n := f(t_n)$. It is well known, see, e.g., [20], that the following error estimate holds for the parabolic equation

$$\|u_h^n - u(t_n)\|_{H^1} \leq C_\epsilon h + C\tau,$$

where C_ϵ is a constant depending on, among other terms, $\|u(t_n)\|_{H^2}$ and may thus be of size ϵ^{-1} if A varies on scale of size ϵ, see Section 1.3.1. Hence, parabolic problems suffer from the same issues as elliptic problems when using classical finite element.

1.3.3. Linear thermoelasticity. As in the previous sections we define a family of triangulations $\{\mathcal{T}_h\}_{h>0}$ and we let $V_h^1 \subseteq V^1$ and $V_h^2 \subseteq V^2$ denote finite element spaces consisting of continuous piecewise affine functions on this triangulation.

The classical FEM with a backward Euler discretization in time for (1.2.2)-(1.2.3) reads; for $n \in \{1, ..., N\}$ find $u_h^n \in V_h^1$ and $\theta_h^n \in V_h^2$, such that

$$(\sigma(u_h^n) : \varepsilon(v_1)) - (\alpha\theta_h^n, \nabla \cdot v_1) = (f^n, v_1), \quad \forall v_1 \in V_h^1, \tag{1.3.6}$$

$$(\bar{\partial}_t \theta_h^n, v_2) + (\kappa \nabla \theta_h^n, \nabla v_2) + (\alpha \nabla \cdot \bar{\partial}_t u_h^n, v_2) = (g^n, v_2), \quad \forall v_2 \in V_h^2, \tag{1.3.7}$$

with the notation and time discretization as in Section 1.3.2. Here $u_h^0 = u_{h,0}$ and $\theta_h^0 = \theta_{h,0}$, where $u_{h,0} \in V_h^1$ and $\theta_{h,0} \in V_h^2$ denote suitable approximations of the initial conditions.

A priori analysis for the system (1.3.6)-(1.3.7) can be found in [10]. It follows that the error is bounded by

$$\|u_h^n - u^n\|_{H^1} + \|\theta_h^n - \theta^n\| + \left(\sum_{j=1}^{n} \tau \|\theta_h^j - \theta^j\|_{H^1}^2 \right)^{1/2} \leq C_\epsilon h + C\tau,$$

where the constant C_ϵ depends on both $\|u(t_n)\|_{H^2}$ and $\|\theta(t_n)\|_{H^2}$. By arguments similar to the ones used for the elliptic equation in Section 1.3.1, we typically get $\|u(t_n)\|_{H^2} \sim \epsilon^{-1}$ and $\|\theta(t_n)\|_{H^2} \sim \epsilon^{-1}$, if the material has variations on a scale of size ϵ.

1.3.4. Riccati equations. For this equation we also consider a family of triangulations $\{\mathcal{T}_h\}_{h>0}$ and let $V_h \subseteq V$ denote a finite element space consisting of continuous piecewise linear functions on this triangulation. Furthermore, we let $\mathcal{A}_h \colon V_h \to V_h$, $\mathcal{B}_h \colon U \to V_h$, and $\mathcal{C}_h \colon V_h \to Z$, be discretized versions of the operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, defined by

$$(\mathcal{A}_h x, y) = (\mathcal{A}x, y), \quad (\mathcal{B}_h u, y) = (\mathcal{B}u, y), \quad \text{and} \quad (\mathcal{C}_h x, z)_Z = (\mathcal{C}x, z)_Z$$

for all $x, y \in V_h$, $u \in U$ and $z \in Z$.

We can now define the semidiscrete FEM of (1.2.9); find $X_h : V_h \to V_h$ such that

$$\begin{aligned} \left(\dot{X}_h x, y\right) &= (X_h x, \mathcal{A}_h y) + (X_h y, \mathcal{A}_h x) + (\mathcal{C}_h x, \mathcal{C}_h y)_Z \\ &\quad - \left(\mathcal{R}^{-1} \mathcal{B}_h^* X_h x, \mathcal{B}_h^* X_h y\right)_U, \end{aligned} \tag{1.3.8}$$

for all $x, y \in V_h$, where $X_h(0)$ is some suitable initial condition.

Let $\mathrm{Id}_h \colon V_h \to L_2$ be the identity operator from V_h into L_2. Note that its adjoint $\mathrm{Id}_h^* \colon L_2 \to V_h$ is the L_2-orthogonal projection of L_2 onto V_h. In [14] the following error bound is proved for the semidiscrete approximation

$$\|X - X_h \,\mathrm{Id}_h^*\|_{\mathcal{L}(L_2, L_2)} \leq C_\epsilon h^2 (\log h^{-1} + t^{-1}), \tag{1.3.9}$$

under the assumption that $X(0) \in \mathcal{L}(L_2, L_2)$. The constant C_ϵ involves ϵ^{-2}, because the error depends on the error for the parabolic equation, cf. Subsection 1.3.2. In the L_2-norm we generally get ϵ^{-2} in the error and since we also get a factor h^2 this leads to the same condition for convergence; $h < \epsilon$.

1.4. A generalized finite element method

In [17] a GFEM, often referred to as *localized orthogonal decomposition* (LOD), is proposed and analyzed for elliptic equations of the form (1.3.1). In Section 1.4.1 below we describe this method and the main ideas used in the analysis. We then discuss how to extend this framework to parabolic equations. Finally, we describe how this method can be generalized to define a GFEM for linear thermoelasticity and Riccati equations.

1.4.1. Elliptic equations. The method proposed in [17] builds on the ideas from the variational multiscale method [13, 15], where the solution space is decomposed to into a coarse and a fine part. In [17] the nodal basis functions in the coarse space is then modified by adding a *correction* from the fine space.

We begin by assuming that the mesh size h used in the classical FEM in (1.3.2) is fix and sufficiently small, that is $h < \epsilon$, such that the error (1.3.3) is small. In this case, the solution u_h and the space V_h are referred to as the reference solution and the reference space, respectively. Now define V_H similarly to V_h, but with a larger mesh size $H > h$, such that $V_H \subseteq V_h$. Note that the classical FEM solution u_H in the coarse space V_H is not a good approximation to u. It is, however, cheaper to compute than u_h since $\dim(V_H) < \dim(V_h)$. The aim is now to define a new multiscale space V_{ms} with the same dimension as the coarse space V_H, but with better approximations properties.

To define such a space, we need a (quasi-)interpolation operator $I_H : V_h \to V_H$ with the properties $I_H \circ I_H = I_H$ and for $K \in \mathcal{T}_H$

$$H_K^{-1}\|v - I_H v\|_{L_2(K)} + \|\nabla I_H v\|_{L_2(K)} \leq C_I \|\nabla v\|_{L_2(\omega_K)}, \quad v \in V_h, \tag{1.4.1}$$

where $\omega_K := \cup\{\hat{K} \in \mathcal{T}_H : \hat{K} \cap K \neq \emptyset\}$. For a quasi-uniform mesh, the bounds in (1.4.1) can be summed to achieve a global bound

$$H^{-1}\|v - I_H v\| + \|\nabla I_H v\| \leq C\|\nabla v\|, \tag{1.4.2}$$

There are many interpolation operators that satisfy these conditions. In Paper II and Paper III we use an interpolation of the form $I_H = E_H \circ \Pi_H$, where Π_H is the L_2-projection onto $P_1(\mathcal{T}_H)$, the space of functions that are affine on each triangle $K \in \mathcal{T}_H$ and $E_H : P_1(\mathcal{T}_H) \to V_H$ is an averaging operator. We refer to [18, 9] for further details and other possible choices of I_H.

Now let V_{f} denote the kernel to the operator I_H

$$V_{\text{f}} := \ker I_H = \{v \in V_h : I_H v = 0\}.$$

The space V_h can be decomposed as $V_h = V_H \oplus V_{\text{f}}$, meaning that $v_h \in V_h$ can be decomposed into

$$v_h = v_H + v_{\text{f}}, \quad v_H \in V_H,\ v_{\text{f}} \in V_{\text{f}}. \tag{1.4.3}$$

The kernel V_{f} is a fine scale (detail) space in the sense that it captures all features that are not captured by the coarse space V_H. Let $R_{\text{f}} : V_h \to V_{\text{f}}$ denote the Ritz projection onto V_{f}, that is,

$$a(R_{\text{f}} v, w) = a(v, w), \quad \forall w \in V_{\text{f}},\ v \in V_h. \tag{1.4.4}$$

Because of the decomposition (1.4.3) we have the identity

$$v_h - R_{\mathrm{f}} v_h = v_H + v_{\mathrm{f}} - R_{\mathrm{f}}(v_H + v_{\mathrm{f}}) = v_H - R_{\mathrm{f}} v_H,$$

since $v_{\mathrm{f}} \in V_{\mathrm{f}}$. Using this we can define the multiscale space V_{ms}

$$V_{\mathrm{ms}} := V_h - R_{\mathrm{f}} V_h = V_H - R_{\mathrm{f}} V_H. \tag{1.4.5}$$

Note that V_{ms} is the orthogonal complement to V_{f} with respect to the inner product $a(\cdot,\cdot)$ and must have the same dimension as V_H. Indeed, with $\mathcal{N}$ denoting the inner nodes in $\mathcal{T}_H$ and λ_z the basis function at node z, a basis for V_{ms} is given by

$$\{z \in \mathcal{N} : \lambda_z - R_{\mathrm{f}} \lambda_z\}.$$

Hence, the basis functions are the classical nodal basis functions modified by corrections $R_{\mathrm{f}}\lambda_z$ computed in the fine scale space. Note that the correction $R_{\mathrm{f}}\lambda_z$ depends on the choice of interpolation I_H. A different choice leads to a different method, since the space V_{ms} changes.

Replacing V_h with V_{ms} in (1.3.2) we can now define the GFEM; find $u_{\mathrm{ms}} \in V_{\mathrm{ms}}$, such that,

$$a(u_{\mathrm{ms}}, v) = (f, v), \quad \forall v \in V_{\mathrm{ms}}. \tag{1.4.6}$$

The following theorem gives an a priori bound for the GFEM and can be found in [17]. We include the proof here since it is short and highlights the main ideas used in the analysis.

Theorem 1.4.1. *Let u_h be the solution to* (1.3.2) *and u_{ms} the solution to* (1.4.6). *Then*

$$\|u_{\mathrm{ms}} - u_h\|_{H^1} \leq CH\|f\|,$$

where C does not depend on the derivatives of A.

Proof. Define $e := u_{\mathrm{ms}} - u_h$ and note that $e \in V_{\mathrm{f}}$. Hence, $I_H e = 0$. Furthermore we have due to Galerkin orthogonality $a(e, v_{\mathrm{ms}}) = 0$ for $v_{\mathrm{ms}} \in V_{\mathrm{ms}}$. Using this together with the interpolation bound (1.4.2) we have

$$a(e,e) = -a(e,u_h) = -(f,e) \leq \|f\|\|e\| = \|f\|\|e - I_H e\| \leq CH\|f\|\|\nabla e\|,$$

and the bound follows by using equivalence of the energy norm induced by $a(\cdot,\cdot)$ and the H^1-norm. □

From Theorem 1.4.1 we have that the solution given by the GFEM converges to u_h, with optimal order, independently of the derivatives (variations) of A. We emphasize that the total error is bounded by

$$\|u_{\mathrm{ms}} - u\|_{H^1} \leq \|u_{\mathrm{ms}} - u_h\|_{H^1} + \|u_h - u\|_{H^1},$$

where the error in the second term is due to the classical FEM and assumed to be of reasonable size, since h is assumed to be sufficiently small.

Although the a priori analysis seems promising, the GFEM as suggested above suffers from some drawbacks. The problem of finding the corrections $R_{\mathrm{f}}\lambda_z$, which are needed to construct the basis, are posed in the entire fine scale

space V_{f} which has high dimension (of the same order as V_h). Furthermore, the corrections generally have global support, which reduces the sparsity of the resulting discrete system. Both issues are resolved by performing a localization of the corrections. The localization is motivated by the observation that the correction $R_{\mathrm{f}}\lambda_z$ decays exponentially away from node z.

1.4.2. Localization. In [17] it is proved that the corrections decay exponentially and a localization procedure is proposed. However, in [11] a different localization technique is proposed which allows for smaller patches to be used. We describe the procedure in [11] here, which is also the procedure that is used in the appended papers.

We define patches of size k in the following way; for $K \in \mathcal{T}_H$

$$\omega_0(K) := \text{int } K,$$
$$\omega_k(K) := \text{int } \big(\cup \{\hat{K} \in \mathcal{T}_H : \hat{K} \cap \overline{\omega_{k-1}(K)} \neq \emptyset\}\big), \quad k = 1, 2, ...,$$

and let $V_{\mathrm{f}}(\omega_k(K)) := \{v \in V_{\mathrm{f}} : v(z) = 0 \text{ on } \overline{\Omega} \setminus \omega_k(K)\}$ be the restriction of V_{f} to the patch $\omega_k(K)$.

We proceed by noting that R_{f} in (1.4.4) can be written as the sum

$$R_{\mathrm{f}} = \sum_{K \in \mathcal{T}_H} R_{\mathrm{f}}^K,$$

where $R_{\mathrm{f}}^K : V_h \to V_{\mathrm{f}}$ fulfills

$$a(R_{\mathrm{f}}^K v, w) = a(v, w)_K, \quad \forall w \in V_{\mathrm{f}}, \quad v \in V_h, \quad K \in \mathcal{T}_H, \tag{1.4.7}$$

where we define

$$a(v, w)_K := (A\nabla v, \nabla w)_{L_2(K)}, \quad K \in \mathcal{T}_H.$$

The aim is to localize these computations by replacing V_{f} with $V_{\mathrm{f}}(\omega_k(K))$. Define $R_{\mathrm{f},k}^K : V_h \to V_{\mathrm{f}}(\omega_k(K))$ such that

$$a(R_{\mathrm{f},k}^K v, w) = a(v, w)_K, \quad \forall w \in V_{\mathrm{f}}(\omega_k(K)), \quad v \in V_h, \quad K \in \mathcal{T}_H,$$

and set $R_{\mathrm{f},k} := \sum_{K \in \mathcal{T}_H} R_{\mathrm{f},k}^K$. We can now define the localized multiscale space

$$V_{\mathrm{ms},k} = \{v_H - R_{\mathrm{f},k} v_H : v_H \in V_H\}. \tag{1.4.8}$$

By replacing V_{ms} with $V_{\mathrm{ms},k}$ in (1.4.6) a localized GFEM can be defined; find $u_{\mathrm{ms},k} \in V_{\mathrm{ms},k}$ such that

$$a(u_{\mathrm{ms},k}, v) = (f, v), \quad \forall v \in V_{\mathrm{ms},k}. \tag{1.4.9}$$

Since the dimension of $V_{\mathrm{f}}(\omega_k(K))$ can be made significantly smaller than the dimension of V_{f} (depending on k), the problem of finding $R_{\mathrm{f},k}\lambda_z$ is computationally cheaper than finding $R_{\mathrm{f}}\lambda_z$. Moreover, the resulting discrete system is sparse. It should also be noted that the computation of $R_{\mathrm{f},k}\lambda_z$ for all nodes z is suitable for parallelization, since they are independent of each other.

The convergence of the method (1.4.9) depends on the size of the patches. In [17, 11] the following theorem is proved.

Theorem 1.4.2. *Let u_h be the solution to* (1.3.2) *and $u_{\mathrm{ms},k}$ the solution to* (1.4.9)*. Then there exists $\xi \in (0,1)$ such that*

$$\|u_{\mathrm{ms},k} - u_h\|_{H^1} \leq C(H + k^{d/2}\xi^k)\|f\|,$$

where C does not depend on the derivatives of A.

To achieve linear convergence k should be chosen proportional to $\log H^{-1}$, that is, $k = c \log H^{-1}$, for some constant c.

1.4.3. Parabolic equations. A natural first step in generalizing the GFEM to linear thermoelasticity and Riccati equations, is to first extend it to a time dependent problem of parabolic type. Recall that the thermoelastic system (1.2.2)-(1.2.3) is parabolic [19]. This is the subject of Paper I.

For this purpose, we first study the error analysis for the classical finite element method. The error is usually split into the two parts

$$u_h^n - u(t_n) = u_h^n - R_h u(t_n) + R_h u(t_n) - u(t_n) =: \theta^n + \rho^n,$$

where $R_h : V \to V_h$ is the Ritz projection given by

$$a(R_h v, w) = a(v, w), \quad \forall w \in V_h,\ v \in V.$$

The error of the Ritz projection is given by the analysis of the elliptic problem

$$\|R_h v - v\|_{H^1} \leq Ch\|D^2 v\|. \tag{1.4.10}$$

This directly gives the error of ρ^n. Indeed, $\|\rho^n\|_{H^1} \leq Ch\|D^2 u(t_n)\|$, where $\|D^2 u(t_n)\| \leq C_\epsilon \|\nabla \cdot A \nabla u(t_n)\| = C_\epsilon \|f^n - \dot{u}(t_n)\|$ and C_ϵ depends on the derivatives of A. Furthermore, to bound $\|\theta^n\|_{H^1}$ we put θ^n into (1.3.5), which gives

$$\begin{aligned}(\bar{\partial}_t \theta^n, v) + a(\theta^n, v) &= -((R_h - I)\bar{\partial}_t u(t_n) + (\bar{\partial}_t u(t_n) - \dot{u}(t_n)), v) \\ &=: -(\bar{\partial}_t \rho^n + \omega, v),\end{aligned}$$

where the error of $\bar{\partial}_t \rho^n$ follows from (1.4.10). The error of ω only depends on the time discretization and follows from Taylor's formula. In order to bound θ^n in the H^1-norm we can choose $v = \bar{\partial}_t \theta^n$. The key observation from this analysis is that the error depends (apart from ω) on the Ritz projection onto the FEM space V_h, which is given by the error of the elliptic equation.

Motivated by this, we propose the following GFEM for the parabolic problem, where the space V_h in (1.3.5) is simply replaced by the multiscale space V_{ms} defined in Section 1.4.1; for $n \in \{1, ..., N\}$ find $u_{\mathrm{ms}}^n \in V_{\mathrm{ms}}$, such that, $u_{\mathrm{ms}}^0 = u_{\mathrm{ms},0}$

$$(\bar{\partial}_t u_{\mathrm{ms}}^n, v) + a(u_{\mathrm{ms}}^n, v) = (f^n, v), \quad \forall v \in V_{\mathrm{ms}}, \tag{1.4.11}$$

with $u_{\mathrm{ms},0}$ a suitable approximation of $u_{h,0}$. Now, because of the choice of the space V_{ms} we can define a Ritz projection $R_{\mathrm{ms}} : V_h \to V_{\mathrm{ms}}$ by

$$a(R_{\mathrm{ms}} v, w) = a(v, w) = (\mathcal{A}_h v, w), \quad \forall w \in V_{\mathrm{ms}},$$

where $\mathcal{A}_h : V_h \to V_h$ is the operator defined by

$$(\mathcal{A}_h v, w) = a(v, w), \quad \forall w \in V_h.$$

The error analysis for the elliptic problem in [17] gives the bound

$$\|R_{\text{ms}}v - v\|_{H^1} \leq CH\|\mathcal{A}_h v\|, \quad \forall v \in V_h, \tag{1.4.12}$$

where C is independent of the derivatives of A. The assumption that A does not depend on time is crucial here. Otherwise, we would have to define a new space and compute a new set of basis functions at each time step t_n.

As for the elliptic equation we assume that h is sufficiently small to resolve the variations in A. This means that the reference solution u_h given by (1.3.5) approximates u in (1.3.4) sufficiently well. In the error analysis we can thus split

$$\|u^n_{\text{ms}} - u(t_n)\|_{H^1} \leq \|u^n_{\text{ms}} - u^n_h\|_{H^1} + \|u^n_h - u(t_n)\|_{H^1},$$

where the second part is bounded by classical FEM error analysis. For the first part we can use a similar analysis, but with the new Ritz projection R_{ms}. We split the error into the parts

$$u^n_{\text{ms}} - u^n_h = u^n_{\text{ms}} - R_{\text{ms}}u^n_h + R_{\text{ms}}u^n_h - u^n_h =: \theta^n_{\text{ms}} + \rho^n_{\text{ms}},$$

where the error of ρ^n_{ms} is given by (1.4.12) and $\mathcal{A}_h u^n_h = P_h f^n - \bar{\partial}_t u^n_h$ with P_h denoting the L_2-projection onto V_h. For θ^n_{ms} we get

$$(\bar{\partial}_t \theta^n_{\text{ms}}, v) + a(\theta^n_{\text{ms}}, v) = -(\bar{\partial}_t \rho^n_{\text{ms}}, v), \quad \forall v \in V_{\text{ms}}.$$

Naturally, the error bound in this case depends on the regularity of the (discrete) time derivative of the reference solution. Since the initial data is not in H^2 we expect, for instance, $\|\bar{\partial}_t u^n_h\|$ to depend on negative powers of t_n. This is possible since the backward Euler scheme preserves the smoothing effect of parabolic problems. In Paper I this is thoroughly investigated and error bounds involving negative powers of t_n are derived.

To utilize the localization introduced in Section 1.4.1 we can replace V_{ms} by $V_{\text{ms},k}$, define a new Ritz projection $R_{\text{ms},k} : V_h \to V_{\text{ms},k}$, and perform similar splits of the error. The localized GFEM for the parabolic equation reads; for $n \in \{1, ..., N\}$ find $u^n_{\text{ms},k} \in V_{\text{ms},k}$, such that, $u^0_{\text{ms},k} = u_{\text{ms},k,0}$

$$(\bar{\partial}_t u^n_{\text{ms},k}, v) + a(u^n_{\text{ms},k}, v) = (f^n, v), \quad \forall v \in V_{\text{ms},k}, \tag{1.4.13}$$

with $u_{\text{ms},k,0}$ a suitable approximation of $u_{h,0}$.

The main result in Paper I is the following theorem.

THEOREM 1.4.3. *Let* $\{u^n_h\}_{n=1}^N$ *be the solution to* (1.3.5) *and* $\{u^n_{\text{ms},k}\}_{n=1}^N$ *the solution to* (1.4.13). *There exists* $\xi \in (0, 1)$ *such that*

$$\|u^n_h - \tilde{u}^n_{\text{ms},k}\| \leq C(1 + \log n)(H + k^{d/2}\xi^k)^2\big(C_f + t_n^{-1}\|u_{h,0}\|\big),$$

for $n \in \{1, ..., N\}$, *where* C *and* C_f *are constants independent of the variations in* A.

The factor $(1 + \log n)$ can be removed if $f = 0$. In Paper I a more general form of the parabolic problem is studied, but the convergence result remains the same.

1.4.4. A note on the initial data. In the analysis we allow the initial data to be nonsmooth. A nonsmooth initial data means that the error bounds generally blow up close to $t = 0$. This appears as a negative power of t in the error analysis.

One argument for allowing nonsmooth initial data is because the FEM solution that we want to approximate only have initial data $u_{h,0} \in V_h$ and $V_h \not\subset H^2(\Omega)$. A relevant question is if this can be relaxed. Can we impose some other condition on $u_{h,0}$ (and thus u_0) to avoid working in the nonsmooth data regime?

The answer is yes, but it is a restrictive assumption that may not be fulfilled in many applications. To illustrate what this assumption would look like, let us, for simplicity, consider the semidiscrete version that is still continuous in time. Find $u_h \in V_h$ such that

$$(\dot{u}_h, v) + a(u_h, v) = (f, v), \quad \forall v \in V_h, \tag{1.4.14}$$

with initial data $u_h(0) = u_{0,h}$. We note that the error analysis depends on terms like e.g. $\int_0^T \|\dot{u}_h(t)\|_{H^1}^2 \,\mathrm{d}t$, cf. Paper I. By differentiating the parabolic equation once with respect to t we get

$$(\ddot{u}_h, v) + a(\dot{u}_h, v) = (\dot{f}, v),$$

with the initial data $\dot{u}_h(0) = P_h f(0) - \mathcal{A}_h u_{0,h}$. The value of $\dot{u}_h(0)$ is derived by letting t approach zero in (1.4.14). Now, choosing $v = \dot{u}_h$ we may derive an energy estimate for $\int_0^T \|\dot{u}_h(t)\|_{H^1}^2 \,\mathrm{d}t$. However, the initial data is bounded in L_2 only if $\|\mathcal{A}_h u_{0,h}\|_{L_2}$ is bounded. This means that there must exist a $g \in L_2$ such that

$$(\mathcal{A}_h u_{0,h}, v) = a(u_{0,h}, v) = (g, v), \quad \forall v \in V_h.$$

Since $u_{0,h}$ needs to be an approximation of u_0, for instance $u_{0,h} = R_h u_0$, this implies that u_0 must fulfill the same assumption. That is,

$$a(u_0, v) = (g, v), \quad \forall v \in V.$$

Hence, the initial data would have to be the solution to an elliptic equation with right hand side $g \in L_2$. If this is fulfilled, the t^{-1}-factor in the error estimate could be avoided. However, it is restrictive and we have chosen to work with the much more general assumption $u_0 \in L_2$ in Paper I. In Paper III we assume that the initial data is in H_0^1, which is less general, but still nonsmooth.

1.4.5. Linear thermoelasticity. For the parabolic equation we relied on results for the elliptic (stationary) equation. Hence, a natural step is to first analyze the stationary version of the thermoelastic system. By neglecting the terms involving time derivatives we arrive at the following system; find $u_h \in V_h^1$ and $\theta_h \in V_h^2$, such that

$$\begin{aligned}(\sigma(u_h) : \varepsilon(v_1)) - (\alpha\theta_h, \nabla \cdot v_1) &= (f, v_1), \quad \forall v_1 \in V_h^1,\\ (\kappa\nabla\theta_h, \nabla v_2) &= (g, v_2), \quad \forall v_2 \in V_h^2.\end{aligned}$$

To derive a GFEM for this system we need to decompose two different spaces; V_h^1 and V_h^2. One could hope to use the full system as a bilinear form for the split. However, this is not an inner product, since it is not symmetric, and we cannot define a natural orthogonal decomposition of the form $V_h^1 \times V_h^2 = V_{\text{ms}}^1 \times V_{\text{ms}}^2 \oplus V_{\text{f}}^1 \times V_{\text{f}}^2$ in the same fashion as before. Instead, we use $(\sigma(\cdot) : \varepsilon(\cdot))$ to decompose V_h^1 and $(\kappa\nabla\cdot, \nabla\cdot)$ to decompose V_h^2, which are both inner products on their respective space. This is done by mimicking the procedure described in Section 1.4.1. First define two interpolations $I_H^1 : V_h^1 \to V_H^1$ and $I_H^2 : V_h^2 \to V_H^2$ into the coarse finite element spaces $V_H^1 \subseteq V_h^1$ and $V_H^2 \subseteq V_h^2$. Now, the corresponding kernels are $V_{\text{f}}^1 := \ker I_H^1$ and $V_{\text{f}}^2 := \ker I_H^2$, and we can define the Ritz projections onto these spaces $R_{\text{f}}^1 : V_h^1 \to V_{\text{f}}^1$ and $R_{\text{f}}^2 : V_h^2 \to V_{\text{f}}^2$ given by

$$
\begin{aligned}
(\sigma(v_1 - R_{\text{f}}^1 v_1) : \varepsilon(w_1)) &= 0, \quad \forall w_1 \in V_{\text{f}}^1,\ v_1 \in V_h^1 \\
(\kappa\nabla(v_2 - R_{\text{f}}^2 v_2), \nabla w_2) &= 0, \quad \forall w_2 \in V_{\text{f}}^2,\ v_2 \in V_h^2.
\end{aligned}
$$

The multiscale spaces are finally defined as

$$
V_{\text{ms}}^1 := V_H^1 - R_{\text{f}}^1 V_H^1, \quad V_{\text{ms}}^2 := V_H^2 - R_{\text{f}}^2 V_H^2,
$$

as in (1.4.5). With these spaces we can now define a GFEM corresponding to the stationary system. Find $u_{\text{ms}} \in V_{\text{ms}}^1$ and $\theta_{\text{ms}} \in V_{\text{ms}}^2$

$$
\begin{aligned}
(\sigma(u_{\text{ms}}) : \varepsilon(v_1)) - (\alpha\theta_{\text{ms}}, \nabla \cdot v_1) &= (f, v_1), \quad \forall v_1 \in V_{\text{ms}}^1, \\
(\kappa\nabla\theta_{\text{ms}}, \nabla v_2) &= (g, v_2), \quad \forall v_2 \in V_{\text{ms}}^2.
\end{aligned}
$$

The spaces V_{ms}^1 and V_{ms}^2 are designed to handle multiscale behavior in the coefficients μ, λ, and κ respectively. However, α is also material dependent and can be expected to vary at the same scale. For this reason, we shall add an extra correction to the solution u_{ms} inspired by the techniques in [15, 11]. This additional correction is defined as $u_{\text{f}} \in V_{\text{f}}^1$, such that,

$$
(\sigma(u_{\text{f}}) : \varepsilon(w_1)) = (\alpha\theta_{\text{ms}}, \nabla \cdot w_1), \quad \forall w_1 \in V_{\text{f}}^1,
$$

and we define $\tilde{u}_{\text{ms}} = u_{\text{ms}} + u_{\text{f}}$. It can now be proved, cf. Paper III, that the following error bounds hold

$$
\|u_h - \tilde{u}_{\text{ms}}\|_{H^1} \leq CH\|f\| + C\|\theta_h - \theta_{\text{ms}}\|, \tag{1.4.15}
$$

$$
\|\theta_h - \theta_{\text{ms}}\|_{H^1} \leq CH\|g\|, \tag{1.4.16}
$$

where C is independent of the variations in μ, λ, α, and κ.

Inspired by this, we formulate the following GFEM for the time dependent system. For $n \in \{1, ..., N\}$ find $\tilde{u}_{\text{ms}}^n = u_{\text{ms}}^n + u_{\text{f}}^n$, with $u_{\text{ms}}^n \in V_{\text{ms}}^1$ and $u_{\text{f}}^n \in V_{\text{f}}^1$, and $\theta_{\text{ms}}^n \in V_{\text{ms}}^2$, such that

$$
(\sigma(\tilde{u}_{\text{ms}}^n) : \varepsilon(v_1)) - (\alpha\theta_{\text{ms}}^n, \nabla \cdot v_1) = (f^n, v_1) \tag{1.4.17}
$$

$$
(\bar{\partial}_t\theta_{\text{ms}}^n, v_2) + (\kappa\nabla\theta_{\text{ms}}^n, \nabla v_2) + (\alpha\nabla \cdot \bar{\partial}_t\tilde{u}_{\text{ms}}^n, v_2) = (g^n, v_2), \tag{1.4.18}
$$

$$
(\sigma(u_{\text{f}}^n) : \varepsilon(w_1)) - (\alpha\theta_{\text{ms}}^n, \nabla \cdot w_1) = 0, \tag{1.4.19}
$$

for all $v_1 \in V^1_{\text{ms}}$, $v_2 \in V^2_{\text{ms}}$, and $w_1 \in V^1_{\text{f}}$. Here $\tilde{u}^0_{\text{ms}} = \tilde{u}_{\text{ms},0}$ and $\theta^0_{\text{ms}} = \theta_{\text{ms},0}$ are some suitable approximations of $u_{h,0}$ and $\theta_{h,0}$. Here we have added an additional correction, u^n_{f}, on u^n_{ms} in each time step motivated by the correction in the stationary setting. The system now consists of three coupled equations.

For the analysis of the time dependent problem we may now define a Ritz projection corresponding to the stationary system, with the additional correction included, and split the error into two parts, see Paper III.

To proceed we also need to perform a localization of both spaces V^1_{ms} and V^2_{ms}. We use the patches $\omega_k(K)$ defined in Section 1.4.2 to define localized spaces $V^1_{\text{ms},k}$ and $V^2_{\text{ms},k}$, as in (1.4.8). To motivate this we need to show that the corrections $R^1_{\text{f}}\lambda_x$ and $R^2_{\text{f}}\lambda_y$ decay exponentially away from node x and y, where λ_x and λ_y denotes the classical hat functions in V^1_H and V^2_H respectively. The correction $R^2_{\text{f}}\lambda_y$ is based on the inner product $(\kappa\nabla\cdot, \nabla\cdot)$, which is of the same type as $a(\cdot,\cdot)$ in Section 1.4.1. Hence, the decay follows directly from [17, 11]. The correction $R^1_{\text{f}}\lambda_x$ is based on the elasticity form $(\sigma(\cdot) : \varepsilon(\cdot))$ and the decay does *not* follow directly from the earlier results. This is instead proven in Paper II.

The localized GFEM for (1.3.6)-(1.3.7) is now defined as; for $n \in \{1, ..., N\}$ find

$$\tilde{u}^n_{\text{ms},k} = u^n_{\text{ms},k} + \sum_{K\in\mathcal{T}_H} u^{n,K}_{\text{f},k},$$

with $u^n_{\text{ms},k} \in V^1_{\text{ms},k}$, $u^{n,K}_{\text{f},k} \in V^1_{\text{f}}(\omega_k(K))$, and $\theta^n_{\text{ms},k} \in V^2_{\text{ms},k}$, such that

$$(\sigma(\tilde{u}^n_{\text{ms},k}) : \varepsilon(v_1)) - (\alpha\theta^n_{\text{ms},k}, \nabla\cdot v_1) = (f^n, v_1), \tag{1.4.20}$$

$$(\bar{\partial}_t\theta^n_{\text{ms},k}, v_2) + (\kappa\nabla\theta^n_{\text{ms},k}, \nabla v_2) + (\alpha\nabla\cdot\bar{\partial}_t\tilde{u}^n_{\text{ms},k}, v_2) = (g^n, v_2), \tag{1.4.21}$$

$$(\sigma(u^{n,K}_{\text{f},k}) : \varepsilon(w_1)) - (\alpha\theta^n_{\text{ms},k}, \nabla\cdot w_1)_K = 0, \tag{1.4.22}$$

for all $v_1 \in V^1_{\text{ms}.k}$, $v_2 \in V^2_{\text{ms},k}$, and $w_1 \in V^1_{\text{f}}(\omega_k(K))$.

The main theorem in Paper III is Theorem 1.4.4 below which is proved under certain conditions on the size of H. Here $C_{f,g}$ denotes a constant depending on f and g, see Paper III for details.

Theorem 1.4.4. *Let $\{u^n_h\}_{n=1}^N$ and $\{\theta^n_h\}_{n=1}^N$ be the solutions to* (1.3.6)-(1.3.7) *and $\{\tilde{u}^n_{\text{ms},k}\}_{n=1}^N$ and $\{\theta^n_{\text{ms},k}\}_{n=1}^N$ the solutions to* (1.4.20)-(1.4.22)*. There exists $\xi \in (0,1)$, such that*

$$\|u^n_h - \tilde{u}^n_{\text{ms},k}\|_{H^1} + \|\theta^n_h - \theta^n_{\text{ms},k}\|_{H^1} \le C(H + k^{d/2}\xi^k)\big(C_{f,g} + t_n^{-1/2}\|\theta^0_h\|_{H^1}\big),$$

for $n \in \{1, ..., N\}$, where C and $C_{f,g}$ are constants independent of the variations in μ, λ, α, and κ.

1.4.6. Riccati equations. Recall that we are interested in operators $\mathcal{A}$, such that $(\mathcal{A}u, v) = -a(u, v)$ and $a(u, v) = \int \kappa\nabla u\cdot\nabla v$. Hence the state equation is a parabolic equation of the type considered in Paper I. It thus makes sense to use the same GFEM space that we used for the parabolic equation and build upon the results from this paper.

Let us describe the localized GFEM version of this method directly, without taking the step via the global GFEM. The idea is to replace V_h in (1.3.8) with the localized GFEM space $V_{\mathrm{ms},k}$ defined in (1.4.8), . For this purpose, we first define the operators $\mathcal{A}_k^{\mathrm{ms}}\colon V_{\mathrm{ms},k} \to V_{\mathrm{ms},k}$, $\mathcal{B}_k^{\mathrm{ms}}\colon U \to V_{\mathrm{ms},k}$ and $\mathcal{C}_k^{\mathrm{ms}}\colon V_{\mathrm{ms},k} \to Z$ such that

$$(\mathcal{A}_k^{\mathrm{ms}} v, w) = (\mathcal{A} v, w)\,,\ (\mathcal{B}_k^{\mathrm{ms}} u, w) = (\mathcal{B} u, w)\,,\ \text{and}\ (\mathcal{C}_k^{\mathrm{ms}} v, z)_Z = (\mathcal{C} v, z)_Z\,,$$

for all $v, w \in V_{\mathrm{ms},k}$, $u \in U$ and $z \in Z$. With this notation the (semidiscrete) localized GFEM of (1.3.8) is to find $X_k^{\mathrm{ms}}\colon V_{\mathrm{ms},k} \to V_{\mathrm{ms},k}$ satisfying

$$\begin{aligned}(1.4.23)\qquad \left(\dot{X}_k^{\mathrm{ms}} u, v\right) &= (X_k^{\mathrm{ms}} u, \mathcal{A}_k^{\mathrm{ms}} v) + (X_k^{\mathrm{ms}} v, \mathcal{A}_k^{\mathrm{ms}} u) \\ &\quad + (\mathcal{C}_k^{\mathrm{ms}} u, \mathcal{C}_k^{\mathrm{ms}} v)_Z - \left(\mathcal{R}^{-1}(\mathcal{B}_k^{\mathrm{ms}})^* X_k^{\mathrm{ms}} u, (\mathcal{B}_k^{\mathrm{ms}})^* X_k^{\mathrm{ms}} v\right)_U\end{aligned}$$

for all $u, v \in V_{\mathrm{ms},k}$ and with $X_k^{\mathrm{ms}}(0)$ some appropriate approximation of $X(0)$.

In the analysis it is convenient to write the Riccati equations (1.3.8) and (1.4.23) on integral form. For this purpose, we let $E_h(t) : L_2 \to L_2$ denote the solution operator so that $u_h = E_h(t)u_0$ is the solution to the homogeneous parabolic equation

$$(\dot{u}_h, v) + a(u_h, v) = 0, \quad \forall v \in V_h,$$

with $u_h(0) = (\mathrm{Id}_h)^* u_0$. Recall that $(\mathrm{Id}_h)^*$ is the L_2-projection onto V_h. Similarly, we define the solution operator $E_k^{\mathrm{ms}}(t)$ corresponding to the GFEM for the homogeneous parabolic equation in $V_{\mathrm{ms},k}$. Let us also abbreviate

$$(1.4.24)\qquad \tilde{X}(t) = \mathrm{Id}_h\, X_h(t)(\mathrm{Id}_h)^*, \quad \tilde{Y}(t) = \mathrm{Id}_h\, \mathrm{Id}_k^{\mathrm{ms}}\, X_k^{\mathrm{ms}}(t)(\mathrm{Id}_k^{\mathrm{ms}})^*(\mathrm{Id}_h)^*,$$

where $\mathrm{Id}_k^{\mathrm{ms}}\colon V_{\mathrm{ms},k} \to V_h$ is the identity operator such that $\mathrm{Id}_k^{\mathrm{ms}}\, v = v$ for $v \in V_{\mathrm{ms},k}$. Its L_2-adjoint is the L_2-orthogonal projection of V_h onto $V_{\mathrm{ms},k}$. With these definitions, both $\tilde{X}$ and $\tilde{Y}$ are operators from L_2 to L_2, which is an advantage in the analysis. It turns out that on integral form $\tilde{X}$ and $\tilde{Y}$ are equal to the following

$$\begin{aligned}\tilde{X}(t) &= E_h(t)^* \tilde{X}(0) E_h(t) \\ &\quad + \int_0^t E_h(t-s)^* \Big((\mathcal{C}_h P_h)^* \mathcal{C}_h P_h - \tilde{X}(s) S_h \tilde{X}(s) \Big) E_h(t-s)\, \mathrm{d}s,\end{aligned}$$

and

$$\begin{aligned}\tilde{Y}(t) &= E_k^{\mathrm{ms}}(t)^* \tilde{X}(0) E_k^{\mathrm{ms}}(t) \\ &\quad + \int_0^t E_k^{\mathrm{ms}}(t-s)^* \Big((\mathcal{C}_h P_h)^* \mathcal{C}_h P_h - \tilde{Y}(s) S_h \tilde{Y}(s) \Big) E_k^{\mathrm{ms}}(t-s)\, \mathrm{d}s,\end{aligned}$$

where $S_h := \mathrm{Id}_h\, \mathcal{B}_h \mathcal{R}^{-1} \mathcal{B}_h^*\, \mathrm{Id}_h^*$. It is now evident that the error between $\tilde{X}(t)$ and $\tilde{Y}(t)$ depends on the error between the parabolic solution operators $E_h(t)$ and $E_k^{\mathrm{ms}}(t)$, which we analyze in Paper I.

The main result in Paper IV is the following theorem

Theorem 1.4.5. *Let $\tilde{X}$ and $\tilde{Y}$ be localized versions of the operators defined in* (1.4.24)*. Then there exists $\xi \in (0,1)$ such that for*

$$\|\tilde{X}(t) - \tilde{Y}(t)\|_{\mathcal{L}(L_2(\Omega),L_2(\Omega))} \leq C(H + k^{d/2}\xi^k)^2\big(\log(H + k^{d/2}\xi^k)^{-1} + t^{-1}\big),$$

$t \in (0,T]$, where C is a constant independent of the variations in $\mathcal{A}$.

In Paper IV we have chosen, to make the notation less heavy, to suppress the dependency of k (and assume $k \sim \log H^{-1}$) throughout the paper. The error bound then reads $\|\tilde{X}(t) - \tilde{Y}(t)\|_{\mathcal{L}(L_2,L_2)} \leq CH^2\big(\log H^{-1} + t^{-1}\big)$. This is of the same order as for the classical FEM (1.3.9), except that the constant does not depend on ϵ, that is, the constant is independent of the variations in $\mathcal{A}$, in this case.

To perform numerical experiments we need to discretize in time. In Paper IV we split the equation into a parabolic and a nonlinear part and propose a Strang splitting scheme for the time discretization. This is combined with a low-rank formulation to perform fast computations also for large systems.

1.5. Summary of Paper I-IV

Paper I. In Paper I we propose and analyze the GFEM (1.4.11) for parabolic equations with highly varying and oscillating coefficients. We prove convergence of optimal (second) order in the L_2-norm to the reference solution assuming initial data only in L_2. We do not assume any structural conditions on the multiscale coefficient, such as, periodicity or scale separation. Furthermore, we show how to extend this method to semilinear parabolic problems, where the right hand side in (1.3.4) is replaced by $f(u)$.

Paper II. In Paper II we propose a GFEM for linear elasticity equations with applications in heterogeneous materials. In particular, we prove exponential decay of the corrections $R_{\mathrm{f}}^1\lambda_z$ in Section 1.4.5. Furthermore, we prove that the GFEM reduces the locking effect that occurs for materials with large Lamé parameter λ when using classical continuous and piecewise linear finite elements.

Paper III. In Paper III we build on the theory developed in Paper I and Paper II to define a GFEM for linear thermoelasticity with highly varying coefficients describing a heterogeneous material. We prove linear convergence to the reference solution in the H^1-norm independent of the variations in the data, see Theorem 1.4.4 in Section 1.4.5.

Paper IV. In Paper IV we use the results from parabolic equations derived in Paper I to develop a GFEM for differential Riccati equations with multiscale features. We prove second order convergence (except for a logarithmic factor) in the L_2-operator norm for the semidiscrete problem. Furthermore, we show how to derive the fully discrete matrix-valued equations and how to discretize these in time using a splitting scheme of Strang type. In addition, we show how the computations can be performed efficiently in a low-rank setting.

1.6. Future work

In applications involving composite materials there may be uncertainties in the material parameters, such as position or rotation, coming from the assembly procedure. These uncertainties can, for instance, be modeled by letting the coefficients depend on a random variable ω. For a general elliptic problem, the PDE would be of the following form

$$-\nabla \cdot A(x,\omega)\nabla u(x,\omega) = f(x,\omega),$$

where $A(\cdot,\omega)$ is multiscale in space for a fixed ω. However, $A(\cdot,\omega)$ now takes different values for different outcomes ω. This, in turn, means that the space V_{ms} will be different for different outcomes ω. Hence, the main idea to replace the space V_h with V_{ms} fails. This requires new ideas for the construction of an appropriate multiscale space.

While the general stochastic problem is far from a solution, there are many interesting special cases which can be handled within the current LOD framework. For instance, composite materials with defects. This can be modeled as a stochastic problem by letting an inclusion in the material be missing, rotated, or shifted, with a certain (low) probability. This means that for each outcome ω, the coefficient $A(x,\omega)$ may only change its value at some (few) places in the domain. The affected LOD basis functions may then be recomputed locally. In the case of a shift or rotation, the change in A may be described by a mapping with a small derivate, which can be used to derive new corrections efficiently. Moreover, there is a possibility to use error indicators to decide when it is necessary to update the corrections. If the change is very small, the corrections will remain roughly the same and it is unnecessary to compute new ones. This is ongoing work in the community.

Another interesting direction is inverse problems with multiscale features, which can be used to detect defects in composite materials. This is also a very challenging problem numerically, since the solution procedure involves solving a PDE with (different) multiscale coefficients many times. It is possible that one can use the a priori knowledge of where the inclusions are supposed to be, as in the stochastic case in the previous paragraph, to derive a more efficient method.

In viscoelastic theory, the following (strongly) damped wave equation is commonly used

$$\ddot{u} - \nabla \cdot (A\nabla\dot{u} + B\nabla u) = f.$$

If the medium of interest is heterogeneous, both A and B are highly varying coefficients. Thus, it is not enough to only use $b(\cdot,\cdot) = (B\nabla\cdot, \nabla\cdot)$ when defining the multiscale space V_{ms}. The variations in A needs to be accounted for by either considering a time-dependent basis, or adding appropriate additional corrections to the GFEM solution. This is ongoing work in the community.

Furthermore, in Paper IV we focus on diffusion operators of the form $(\mathcal{A}x, v) = \int(\kappa\nabla x \cdot \nabla v)$ for the state equation. However, for many problems in multiphysics, the state equation is a coupled system of equations, such as the

thermoelastic system. It is thus of interest to allow the operator $\mathcal{A}$ in the Riccati equation to represent a system of (multiscale) equations instead. Analysis of problems of this kind should be considered in the future. For this to be possible one first needs to study the LOD method for the corresponding evolution problem without the control input. As for linear thermoelasticity, this involves finding an appropriate multiscale space for the system and, possibly, additional corrections to be added to the solution.

In addition, the convergence analysis of the Riccati equation in Paper IV only applies to the semidiscrete case. A natural next step would be to analyze the convergence of the fully discrete method, by also taking the splitting scheme into account.

References

[1] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank: *Matrix Riccati equations*, Systems & Control: Foundations & Applications, Birkhäuser, Basel, 2003.

[2] R. A. Adams, and J. F. Fournier: *Sobolev spaces*, Pure and Applied Mathematics vol. 140, Elsevier/Academic Press, Amsterdam, Second edition, 2003.

[3] I. Babuška and R. Lipton: *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul. 9 (2011), no. 1, pp. 373–406.

[4] I. Babuška and J. E. Osborn: *Generalized finite element methods: their performance and their relation to mixed methods*, SIAM J. Numer. Anal. 20 (1983), no. 3, 1983.

[5] A. Bensoussan, G. Da Prato, M. C. Delfour, and S. K. Mitter: *Representation and control of infinite dimensional systems*, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, second ed., 2007.

[6] M. A. Biot: *Thermoelasticity and irreversible thermodynamics*, J. Appl. Phys., 27 (1956), pp. 240–253.

[7] B. A. Boley and J. H. Weiner: *Theory of thermal stresses*, John Wiley & Sons, Inc., New York London, 1960.

[8] W. E, and B. Engquist: *The heterogeneous multiscale methods*, Commun. Math. Sci. 1 (2003), no. 1, pp. 87–132.

[9] Ch. Engwer and P. Henning and A. Målqvist and D. Peterseim: *Efficient implementation of the Localized Orthogonal Decomposition method*, Submitted.

[10] A. Ern and S. Meunier: *A posteriori error analysis of Euler-Galerkin approximations to coupled elliptic-parabolic problems*, M2AN Math. Model. Numer. Anal. 43 (2009), no. 2, pp. 353 – 375.

[11] P. Henning and A. Målqvist: *Localized orthogonal decomposition techniques for boundary value problems*, SIAM J. Sci. Comput. 36 (2014), no. 4, pp. A1609–A1634.

[12] T. Y. Hou and X.-H. Wu: *A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media*, J. Comput. Phys. 134 (1997), no. 1, pp. 169–189.

[13] T. J. R. Hughes, G. R. Feijóo, L. Mazzei, and J-B. Quincy: *The variational multiscale method—a paradigm for computational mechanics*, Comput. Methods Appl. Mech. Engrg. 166 (1998), no. 1-2, pp. 3–24.

[14] M. Kroller and K. Kunisch: *Convergence rates for the feedback operators arising in the linear quadratic regulator problem governed by parabolic equations*, SIAM J. Numer. Anal., 28 (1991), no. 5, pp. 1350-1385..

[15] M. G. Larson and A. Målqvist: *Adaptive variational multiscale methods based on a posteriori error estimation: energy norm estimates for elliptic problems*, Comput. Methods Appl. Mech. Engrg. 196 (2007), no. 21-24, pp. 2313–2324.

[16] I. Lasiecka and R. Triggiani: *Control theory for partial differential equations: continuous and approximation theories. I*, vol. 74 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2000.

[17] A. Målqvist and D. Peterseim: *Localization of elliptic multiscale problems*, Math. Comp. 83 (2014), no. 290, pp. 2583–2603.

[18] D. Peterseim: *Variational Multiscale Stabilization and the Exponential Decay of Fine-scale Correctors*, In: G.R. Barrenechea, F. Brezzi, A. Cangiani, E.H. Georgoulis (eds.) Building Bridges: Connections and Challenges in Modern Approaches to Numerical Partial Differential Equations, Lecture Notes in Computational Science and Engineering, vol. 114. Springer (2016). Also available as INS Preprint No. 1509.

[19] R. E. Showalter: *Diffusion in poro-elastic media*, J. Math. Anal. Appl. 251 (2000), no. 1, pp.310–340.

[20] V. Thomée: *Galerkin Finite Element Methods for Parabolic Problems*, Springer Series in Computational Mathematics, Springer-Verlag, Berlin, Second edition, 2006.

[21] A. Ženíšek: *Finite element methods for coupled thermoelasticity and coupled consolidation of clay*, RAIRO Anal. Numér. 18 (1984), no. 2, pp. 183–205.

[22] A. Ženíšek: *The existence and uniqueness theorem in Biot's consolidation theory*, Apl. Mat. 29 (1984), no. 3, pp. 194–211.

CHAPTER 2

The Joule heating problem

2.1. Background

When an electric current is passed through a conductor, heat is produced. A typical example is a light bulb, which becomes warm when in use. Other common applications are thermistors, i.e. resistors whose resistance are temperature dependent. These can be used for temperature sensors, fuses, micro-assembly, etc. In most applications the electrical potential is typically only applied to smaller parts of the boundary of the conductor, for instance through electric pads. To model this properly we need to consider mixed boundary conditions, see, e.g., [7].

The electrical heating effect can be modeled by the following coupled nonlinear system

$$\dot{u} - \Delta u = \sigma(u)|\nabla\varphi|^2, \quad \nabla \cdot \sigma(u)\nabla\varphi = 0, \tag{2.1.1}$$

together with appropriate boundary conditions, where u denotes the temperature and φ the electric potential. It is based on Ohm's and Fourier's law, see, e.g., [4] and references therein. Indeed, if J is the electric current density and q the flow of heat, then

$$J = -\sigma(u)\nabla\varphi, \quad q = -\kappa(u)\nabla u,$$

where σ and κ represents the electrical and thermal conductivity, respectively. The conservation equations are

$$\dot{u} + \nabla \cdot q = J \cdot E, \quad E = -\nabla\varphi, \quad \nabla \cdot J = 0.$$

In this thesis we consider the somewhat simpler version when $\kappa(u) = 1$. Using this, the conservation laws gives (2.1.1).

It turns out that the quadratic source term in (2.1.1) poses difficulties when studying existence, uniqueness, and regularity of the system. Generally, $|\nabla\varphi|^2 \in L_1$ only and in three dimensions this is not enough to guarantee $\sigma(u)|\nabla\varphi|^2 \in H^{-1}$. Thus, results from the classical variational framework, which requires that the right hand side is in H^{-1}, is not available.

In [2] the issue with the source term is avoided by rewriting the term using the equation for φ. With this new formulation they are able to prove existence of a solution in $L_2(H^1)$. However, to prove convergence of numerical solutions to the problem, additional regularity is needed, see, for instance, [6, 1]. Typically,

sufficient regularity in three dimensions cannot be proved, but needs to be assumed.

The purpose of this thesis is to prove strong convergence of finite element approximations of the Joule heating problem with mixed boundary conditions under very mild assumptions on the domain and the data. This is achieved by defining a new variaitonal formulation based on a cut-off functional. The results are applicable to a large class of finite element methods that are conforming in space and piecewise constant in time, satisfying a backward Euler scheme.

Furthermore, under the assumption of creased domains, we prove uniqueness and additional regularity of the solution. In this setting, it is also possible to deduce higher regularity in the interior of the domain. This makes the problem suitable for adaptive mesh refinement.

The rest of this chapter is outlined as follows: In Section 2.2 we introduce the classical variational formulation and the new variational formulation based on the cut-off functional. We also discuss the issue with the source term in more detail. In Section 2.3 we define the finite element approximations and summarize the main results. In Section 2.4 we give an overview of the results on regularity and uniqueness. Finally, in Section 2.5 we summarize Paper V and in Section 2.6 we discuss some possible future work.

2.2. Variational formulation with cut-off

The full Joule heating system, including boundary and initial conditions, is given by

$$D_t u - \Delta u = \sigma(u)|\nabla\varphi|^2, \quad \text{in } \Omega \times (0,T), \tag{2.2.1a}$$

$$\nabla \cdot (\sigma(u)\nabla\varphi) = 0, \quad \text{in } \Omega \times (0,T), \tag{2.2.1b}$$

$$u = g_u, \quad \text{on } \Gamma_D^u \times (0,T), \tag{2.2.1c}$$

$$\varphi = g_\varphi, \quad \text{on } \Gamma_D^\varphi \times (0,T), \tag{2.2.1d}$$

$$n \cdot \nabla u = 0, \quad \text{on } \Gamma_N^u \times (0,T), \tag{2.2.1e}$$

$$n \cdot \nabla\varphi = 0, \quad \text{on } \Gamma_N^\varphi \times (0,T), \tag{2.2.1f}$$

$$u(\cdot,0) = u_0, \quad \text{in } \Omega, \tag{2.2.1g}$$

where D_t denotes the time derivative $\frac{\partial}{\partial t}$ and $\Omega \subseteq \mathbb{R}^3$ describes the body of a conductor. Furthermore, Γ_D^u and Γ_N^u denotes the Dirichlet and Neumann boundary for u, respectively, and $\overline{\Gamma_D^u} \cup \overline{\Gamma_N^u} = \partial\Omega$. Analogously, we define Γ_D^φ and Γ_N^φ for φ. We also assume that g_u and g_φ are defined on the whole domain Ω.

We let $W_p^k(\Omega)$ denote the classical range of Sobolev spaces and define

$$W_p^k(\Omega;\Gamma_D^u) := \{v \in W_p^k(\Omega) : v|_{\Gamma_D^u} = 0\}, \quad \text{for } k > 1/p.$$

The space $W_p^k(\Omega;\Gamma_D^\varphi)$ is defined analogously and H^1 is used to denote W_2^1. We also use V^* for the dual space to V. Furthermore, we use $L_p(0,T;V)$ for the Bochner spaces, see previous chapter, and the notation $v \in H^1(0,T;V)$ is used to denote $v, D_t v \in L_2(0,T;V)$.

From (2.2.1) one can define the following variational formulation; find a pair $(u, \varphi) = (g_u + \tilde{u}, g_\varphi + \tilde{\varphi})$ such that

$$(\tilde{u}, \tilde{\varphi}) \in L_2(0,T; H^1(\Omega; \Gamma_D^u)) \cap H^1(0,T; H^1(\Omega; \Gamma_D^u)^*) \times L_2(0,T; H^1(\Omega, \Gamma_D^\varphi))$$

and for a.e. $t \in (0,T]$

$$\langle D_t u, v\rangle + \langle \nabla u, \nabla v\rangle = \langle \sigma(u)|\nabla\varphi|^2, v\rangle, \tag{2.2.2a}$$

$$\langle \sigma(u)\nabla\varphi, \nabla w\rangle = 0, \tag{2.2.2b}$$

$$\langle u(0), z\rangle = \langle u_0, z\rangle, \tag{2.2.2c}$$

for all $(v, w) \in W_\infty^1(\Omega; \Gamma_D^u) \times H^1(\Omega; \Gamma_D^\varphi)$ and $z \in L_2(\Omega)$, cf. [4].

Since $\tilde{\varphi} \in L_2(0,T; H^1(\Omega, \Gamma_D^\varphi))$, we get $\sigma(u)|\nabla\varphi|^2 \in L_1(0,T; L_1(\Omega))$, if σ is bounded and the boundary data g_φ is smooth enough, see Paper V for precise details.

However, in three spatial dimensions, we cannot guarantee that $L_1(\Omega)$ is in $H^1(\Omega; \Gamma_D^u)^*$. To illustrate this, consider

$$\|f\|_{H^1(\Omega)^*} = \sup_{\substack{v \in H^1(\Omega) \\ v \neq 0}} \frac{\langle f, v\rangle}{\|v\|_{H^1(\Omega)}} \leq \sup_{\substack{v \in H^1(\Omega) \\ v \neq 0}} \frac{\|f\|_{L_1(\Omega)}\|v\|_{L_\infty(\Omega)}}{\|v\|_{H^1(\Omega)}},$$

where we have used Hölder's inequality. This is well defined if we can bound $\|v\|_{L_\infty(\Omega)} \leq C\|v\|_{H^1(\Omega)}$. Using Sobolev's inequality, this holds if $1 > d/2$. Hence, the argument is not valid for $d = 3$ and we may not deduce that $\sigma(u)|\nabla\varphi|^2$ is in $L_2(0,T; H^1(\Omega; \Gamma_D^u)^*)$.

The fact that the right hand side is not in $L_2(0,T; H^1(\Omega; \Gamma_D^u)^*)$ is problematic in this variational framework. The regularity is too low to prove existence of a solution through fixed point theorems.

In [2] another variational formulation is proposed by utilizing the identity

$$\sigma(u)|\nabla\varphi|^2 = \nabla \cdot (\sigma(u)\varphi\nabla\varphi),$$

which follows by employing (2.2.1b). It is also possible to prove a maximum principle for φ, meaning that $\varphi \in L_\infty(0,T; L_\infty(\Omega))$, if g_φ is bounded. Thus,

$$\begin{aligned} |\langle \nabla \cdot (\sigma(u)\varphi\nabla\varphi), v\rangle| &= |-\langle \sigma(u)\varphi\nabla\varphi, \nabla v\rangle| \\ &\leq C\|\sigma(u)\|_{L_\infty(\Omega)}\|\varphi\|_{L_\infty(\Omega)}\|\nabla\varphi\|_{L_2(\Omega)}\|v\|_{L_2(\Omega)}, \end{aligned}$$

and it is clear that $\nabla \cdot (\sigma(u)\varphi\nabla\varphi)$ defines an element in $L_2(0,T; H^1(\Omega; \Gamma_D^u)^*)$. With this approach existence can be proved through Schauder's fixed point theorem, cf. [2].

However, this formulation is not appropriate for finite element approximations. For the right hand side to be well defined for a finite element solution φ_h, we need $\varphi_h \in L_\infty(0,T; L_\infty(\Omega))$. For a fixed h this is true, but it gives a constant that depends inversely on the mesh size h, which causes blow up when $h \to 0$. This poses a demand for a discrete maximum principle, which in turn poses restrictive conditions on the triangulation of the mesh. In this thesis, we

avoid this by introducing the following cut-off functional

$$\lceil f \rceil := \min\{\max\{f + g_\varphi, a\}, b\} - g_\varphi.$$

for some fixed $a, b \in \mathbb{R}$ with $a \leq \min_{\Gamma_D^\varphi \times [0,T]} g_\varphi$ and $b \geq \max_{\Gamma_D^\varphi \times [0,T]} g_\varphi$, inspired by the stationary problem [9]. The cut-off is designed such that $\lceil \tilde{\varphi} \rceil = \tilde{\varphi}$ and for general functions f the bound $a - g_\varphi \leq \lceil f \rceil \leq b - g_\varphi$ holds.

With this cut-off functional we define a new variational formulation using the spaces

$$X := L_2(0,T; H^1(\Omega; \Gamma_D^u)) \cap H^1(0,T; H^1(\Omega; \Gamma_D^u)^*) \times L_2(0,T; H^1(\Omega, \Gamma_D^\varphi)),$$
$$Y := H^1(\Omega; \Gamma_D^u) \times H^1(\Omega; \Gamma_D^\varphi).$$

A weak solution to (2.2.1) is a pair $(u, \varphi) = (g_u + \tilde{u}, g_\varphi + \tilde{\varphi})$, such that $(\tilde{u}, \tilde{\varphi}) \in X$ and for a.e. $t \in (0, T]$

$$\langle D_t u, v\rangle + \langle \nabla u, \nabla v\rangle = -\langle \sigma(u)\lceil \tilde{\varphi} \rceil \nabla \varphi, \nabla v\rangle + \langle \sigma(u)\nabla\varphi \cdot \nabla g_\varphi, v\rangle, \tag{2.2.3a}$$

$$\langle \sigma(u)\nabla\varphi, \nabla w\rangle = 0, \tag{2.2.3b}$$

$$\langle u(0), z\rangle = \langle u_0, z\rangle, \tag{2.2.3c}$$

for all $(v, w) \in Y$ and $z \in L_2(\Omega)$.

Note that for a finite element approximation φ_h we now need $\lceil \tilde{\varphi}_h \rceil \in L_\infty(0,T; L_\infty(\Omega))$ to ensure that the right hand side is in $L_2(0,T; H^1(\Omega; \Gamma_D^u))$. This is fulfilled, independently of h, by definition of the cut-off functional. Hence, a discrete maximum principle is avoided. In Paper V we prove that the variational formulation based on the cut-off (2.2.3) is equivalent to the formulation (2.2.2). This motivates the introduction of the cut-off functional and we can approximate (2.2.3) instead of (2.2.2).

2.3. Finite element approximations

We consider a large class of finite element approximations by allowing $\{V_m^u\}_{m\in\mathbb{N}}$ and $\{V_m^\varphi\}_{m\in\mathbb{N}}$ to be any kind of hierarchical families of subspaces with finite dimension, whose unions are dense in $H^1(\Omega; \Gamma_D^u)$ and $H^1(\Omega; \Gamma_D^\varphi)$, respectively. For a semidiscrete method we define

$$X_m := \{v \in C(0,T; V_m^u) : v|_{[t_i, t_{i+1})} \in C^1(t_i, t_{i+1}; V_m^u)\ \forall\, i\} \times L_\infty(0,T; V_m^\varphi).$$

A semidiscrete Galerkin solution is a pair $(u_m, \varphi_m) = (g_u + \tilde{u}_m, g_\varphi + \tilde{\varphi}_m)$ such that $(\tilde{u}_m, \tilde{\varphi}_m) \in X_m$ and for a.e. $t \in (0, T]$

$$\begin{aligned}\langle D_t u_m, v\rangle + \langle \nabla u_m, \nabla v\rangle = &-\langle \sigma(u_m)\lceil \tilde{\varphi}_m \rceil \nabla \varphi_m, \nabla v\rangle \\ &+ \langle \sigma(u_m)\nabla\varphi_m \cdot \nabla g_\varphi, v\rangle,\end{aligned} \tag{2.3.1a}$$

$$\langle \sigma(u_m)\nabla\varphi_m, \nabla w\rangle = 0, \tag{2.3.1b}$$

$$\langle u_m(0), z\rangle = \langle u_0, z\rangle, \tag{2.3.1c}$$

for all $(v, w) \in V_m^u \times V_m^\varphi$ and $z \in V_m^u$.

For a fully discrete method we need to introduce a time discretization. We let $\{J_l\}_{l\in\mathbb{N}}$ be a family of nested partitions, on the form $0 = t_0 < t_1 < ... < t_N = T$, of the time interval $J = [0, T]$. We denote the subintervals $I_n := (t_{n-1}, t_n]$,

define $f^n := f(t_n)$, and consider a uniform time discretization in the analysis, that is, we assume $t_n - t_{n-1} = \tau_l$.

To define a fully discrete method we define the space

$$\begin{aligned} X_{m,l} =&\{v(x,t) : \forall n\ \exists w \in V_m^u : v(t,\cdot) = w,\ t \in I_n\} \\ &\times \{v(x,t) : \forall n\ \exists w \in V_m^\varphi : v(t,\cdot) = w,\ t \in I_n\}. \end{aligned}$$

to be used with a backward Euler step in time, We seek a pair $(u_{m,l}, \varphi_{m,l}) = (g_u + \tilde{u}_{m,l}, g_\varphi + \tilde{\varphi}_{m,l})$ such that $(\tilde{u}_{m,l}, \tilde{\varphi}_{m,l}) \in X_{m,l}$ and for $n = 1, ..., N$,

$$\left\langle \frac{u_{m,l}^n - u_{m,l}^{n-1}}{\tau_l}, v \right\rangle + \langle \nabla u_{m,l}^n, \nabla v\rangle = -\langle \sigma(u_{m,l}^n)\lceil \tilde{\varphi}_{m,l}^n \rceil \nabla \varphi_{m,l}^n, \nabla v\rangle + \langle \sigma(u_{m,l}^n)\nabla \varphi_{m,l}^n \cdot \nabla g_\varphi^n, v\rangle, \tag{2.3.2a}$$

$$\langle \sigma(u_{m,l}^n)\nabla \varphi_{m,l}^n, \nabla w\rangle = 0, \tag{2.3.2b}$$

$$\langle u_{m,l}^0, z\rangle = \langle u_0, z\rangle, \tag{2.3.2c}$$

for all $(v, w) \in V_m^u \times V_m^\varphi$ and $z \in V_m^u$.

In Paper V we prove the following two theorems for the strong convergence of the semi-discrete and fully discrete methods, respectively.

THEOREM 2.3.1. *A subsequence of solutions $(\tilde{u}_{m_k}, \tilde{\varphi}_{m_k}) \in X_{m_k}$ of (2.3.1) converges strongly in X to a solution $(\tilde{u}, \tilde{\varphi})$ of (2.2.3).*

THEOREM 2.3.2. *A subsequence of solutions $(\tilde{u}_{m_k,l_k}, \tilde{\varphi}_{m_k,l_k}) \in X_{m_k,l_k}$ of (2.3.2) converges strongly in $L_2(0,T; H^1(\Omega;\Gamma_D^u)) \times L_2(0,T; H^1(\Omega;\Gamma_D^\varphi))$ to a solution $(\tilde{u}, \tilde{\varphi})$ of (2.2.3).*

We can only guarantee that there exists a *sub*sequence converging to a weak solution, since uniqueness is not proved for the Joule heating problem. However, if the solution is unique, then the whole sequence converges.

2.4. Regularity and uniqueness of the solution

As discussed in the previous section, the uniqueness of a solution in the space $L_2(0,T; H^1(\Omega))$ is not proved so far and remains an open problem. However, if we make some additional assumptions on the boundary conditions we may improve upon the regularity and prove uniqueness.

The main assumption is that Ω is a creased domain, see [10] for the full definition. In our setting it allows Ω to be a Lipschitz domain with Γ_D^u and Γ_D^φ open and non-empty, but $\partial\Gamma_D^u$ and $\partial\Gamma_D^\varphi$ cannot be re-entrant. This means that the angles between the Dirichlet and Neumann parts of the boundary are strictly less than π. To achieve the result in the following theorem, some additional assumptions on the data are needed, see Paper V for details.

THEOREM 2.4.1. *Let $p > \frac{3}{2}$ and $r > \frac{4p}{2p-3}$. If Ω is a creased domain and the problem data is sufficiently smooth, then there exists a unique solution to (2.2.1) satisfying*

$$\tilde{u} \in W_r^1(0,T_*; L_p(\Omega)) \cap L_r(0,T_*; W_{2p}^1(\Omega;\Gamma_D^u)), \quad \tilde{\varphi} \in L_r(0,T_*; W_{2p}^1(\Omega;\Gamma_D^\varphi)),$$

for some $0 < T_* \leq T$.

This theorem combines results from two papers, [8] and [10]. In Paper V prove that the Joule heating problem fits into the framework presented in [8]. One of the key assumptions is that the Laplacian Δ is a topological isomorphism from $W^1_{2p}(\Omega;\Gamma^u_D)$ to $W^{-1}_{2p}(\Omega;\Gamma^u_D)$ and from $W^1_{2p}(\Omega;\Gamma^\varphi_D)$ to $W^{-1}_{2p}(\Omega;\Gamma^\varphi_D)$. We verify that this is indeed the case if the domain is creased, as proven in [10].

In [2, Section 4] it is proved that if $\nabla\varphi \in L_{2q/(q-3)}(0,T;L_q(\Omega))$, for $q > 3$, then the solution is unique. We emphasize that the proof there is for pure Dirichlet or pure Neumann boundary conditions only, and needs to be adapted to the mixed setting. However, the result is interesting since it coincides with the regularity we get.

Furthermore, assuming that the solution fulfills the regularity in Theorem 2.4.1 we prove, in Paper V, additional regularity in the interior of the domain. Thus, the problem is well suited for *h*- and *hp*-adaptive finite elements.

2.5. Summary of Paper V

Paper V. In Paper V we propose a new variational formulation based on a cut-off functional. With this formulation, we are able to prove strong convergence for a large class of finite element approximations of the Joule heating problem in three spatial dimensions with mixed boundary conditions on Lipschitz domains. The analysis covers both semidiscrete methods on conforming subspaces and fully discrete methods using a backward Euler scheme.

We prove higher regularity and uniqueness of the solution on creased domains. We perform numerical examples to verify the convergence and while all cases do converge, the non-creased domain setting has a significant lower convergence rate.

In addition, we prove higher regularity in the interior of the domain. The difference in regularity throughout the domain implies that the problem is suitable for *h*- and *hp*- adaptive mesh refinements. This is confirmed by considering goal oriented adaptivity for some numerical examples.

2.6. Future work

In future works one should perform a more rigorous a posteriori analysis for the goal oriented adaptivity. In particular, it should be proved that the dual problem is well-posed. Furthermore, the error from linearizing the dual problem and the procedure to approximate the dual solution (extrapolation is used in FEniCS [12]) should be analyzed.

It is still an open problem to show that the solution to (2.2.3), or the version without the cut-off functional, is unique. In the stationary case there are counterexamples proving that the solution is not unique [5]. However, the time-dependent case is different, since we start with the initial data u_0.

The Joule heating equations often appear in more complicated systems, such as the thermoviscoelastic problem [11]. This system models the temperature, electric potential, and deformation of a material and is used to model, for instance, actuators on the micro-scale [7]. A next step could be to analyze these equations.

Finally, one may also combine the two parts in this thesis, by considering the Joule heating problem with multiscale coefficients.

References

[1] G. Akrivis and S. Larsson: *Linearly implicit finite element methods for the time-dependent Joule heating problem*, BIT 45 (2006), no. 3, pp.429–442.

[2] S. N. Antontsev and M. Chipot: *The thermistor problem: existence, smoothness uniqueness, blowup*, SIAM J. Math. Anal. 25 (1994), no. 4, pp. 1128–1156.

[3] M. A. Biot: *General theory of three-dimensional consolidation*, J. Appl. Phys., 18 (1941), no. 2, p. 155–164.

[4] G. Cimatti: *Existence of weak solutions for the nonstationary problem of the joule heating of a conductor*, Ann. Mat. Pura Appl. (4) 162 (1994), pp.33–42.

[5] G. Cimatti: *Stability and multiplicity of solutions for the thermistor problem*, Ann. Mat. Pura Appl. (4) 181 (2002), no. 2, pp.181–212.

[6] C. M. Elliott and S. Larsson: *A finite element model for the time-dependent Joule heating problem*, Math. Comp. 64 (1995), no. 212, pp.1433–1453.

[7] V. A. Henneken, M. Tichem, and P. M. Sarro: *In-package MEMS-based thermal actuators for micro-assembly*, J. Micromech. Microeng. 16 (2006), no. 6, pp.107–115.

[8] M. Hieber and J. Rehberg: *Quasilinear parabolic systems with mixed boundary conditions on nonsmooth domains*, SIAM J. Math. Anal. 40 (2008, no. 1, pp.292–305).

[9] M. Jensen and A. Målqvist: *Finite element convergence for the Joule heating problem with mixed boundary conditions*, BIT 53 (2013), no. 2, pp.475–496.

[10] I. Mitrea and M. Mitrea: *The Poisson problem with mixed boundary conditions in Sobolev and Besov spaces in non-smooth domains*, Trans. Amer. Math. Soc. 359 (2007), no. 9, pp. 4143–4182 (electronic).

[11] A. Målqvist and T. Stillfjord Finite element convergence analysis for the thermoviscoelastic Joule heating problem BIT Numerical Mathematics 57 (2017), no. 3, pp.787–810

[12] M. E. Rognes and A. Logg: *Automated goal-oriented error control I: Stationary variational problems*, SIAM J. Sci. Comput. 35 (2015), no. 3, pp.C173–C193.