



Towards Identification of Effective Parameters in Heterogeneous Media

David Johansson

Faculty of Health, Science and Technology

Mathematics

Master's Thesis, 30 ECTS

Supervisor: Prof.dr.habil. Adrian Muntean

Examiner: Sorina Barza

Date: June 11, 2020

Towards Identification of Effective Parameters in Heterogeneous Media

David Johansson

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics. All material in this thesis which is not my own work has been identified and no material is included for which a degree has previously been conferred.

David Johansson

Approved, June 11, 2020

Advisor: Prof.dr.habil. Adrian Muntean

Examiner: Sorina Barza

Abstract

In this thesis we study a parameter identification problem for a stationary diffusion equation posed in heterogeneous media. This problem is closely related to the Calderón problem with anisotropic conductivities. The anisotropic case is particularly difficult and is ill-posed both in regards to uniqueness of solution and stability on the data. Since the present problem is posed in heterogeneous media, we can take advantage of multiscale modelling and the tools of homogenization theory in the study of the inverse problem, unlike the original Calderón problem. We investigate the possibilities of combining the theory of the Calderón problem with homogenization theory in order to obtain a well-posed parameter identification. We find that homogenization theory indeed can be used to make progress towards a well-posed identification of the diffusion coefficient. The success of the method is, however, dependent both on the precise structure of the heterogeneous media and on the modelling of the measurements in the inverse problem framework.

We have in mind a particular problem formulation which is motivated by an experiment to determine effective coefficients of materials used in food packaging. This experiment comes with a set of requirements on both the heterogeneous media and on the method for making measurements that, unfortunately, are in conflict with the currently available results for well-posedness. We study also an optimization approach to solving the inverse problem under these application specific requirements. Some progress towards well-posedness of the optimization problem is made by proving existence of minimizer, again with homogenization theory playing a key role in obtaining the result. In a proof-of-concept computational study this optimization approach is implemented and compared to two other optimization problems. For the two tested heterogeneous media, the only optimization method that manages to identify reasonably well the diffusion coefficient is the one which makes use of homogenization theory.

Keywords: parameter identification, multiscale modelling, homogenization theory, partial differential equations, inverse modelling

MSC Subject Classification (2010): 35B27, 35R30, 35Q93

Contents

1	Introduction	1
2	Preliminaries	2
2.1	Function spaces and related theorems	2
2.2	Two-scale convergence and homogenization	4
2.3	Necessary optimality conditions as weak formulations	5
3	The forward problem	7
3.1	Weak formulations of the forward problem	8
3.2	Well-posedness of the multiscale forward problem	9
3.3	Homogenization by two-scale convergence	11
3.3.1	Two-scale compactness	12
3.3.2	The two-scale limit	14
3.3.3	The cell problems	15
3.3.4	The upscaled diffusion equation	17
3.4	Differentiability of the parameter-to-solution maps	18
3.4.1	Differentiability with respect to the upscaled coefficient	18
3.4.2	Differentiability with respect to the coefficient of the cell-problems	23
4	The inverse problems	27
4.1	The Neumann-to-Dirichlet and Dirichlet-to-Neumann maps	28
4.2	Formulation of the inverse problems	29
4.3	Ill-posedness of the ε -dependent inverse problem	30
4.4	Towards well-posedness of the upscaled inverse problem	32
4.5	Towards well-posedness by using multiscale modelling	33
4.6	An optimization approach to inversion	34
4.6.1	The upscaled problem - existence of minimizer	35
4.6.2	Two possible approaches to the multiscale problem	37
5	Numerical simulations	38
5.1	The optimization framework	39
5.1.1	Gradient calculations	40
5.1.2	The conjugate gradient algorithm with inexact line search	42
5.1.3	A few remarks on the performance of the inexact line search	44
5.1.4	Choosing a regularization parameter	44
5.2	The simulation setup	45
5.2.1	The measurement	45
5.2.2	Some implementation details	46
5.3	Inversion in the case of fine microstructure	48
5.3.1	The simulated measurement	48
5.3.2	Numerical results	49
5.4	Inversion in the case of coarse microstructure	54

5.4.1	The simulated measurement	54
5.4.2	Numerical results	55
6	Conclusion and outlook	58

1 Introduction

Inverse problems arise in a variety of important applications in science and engineering and has introduced to applied mathematics a wide range of challenging problems. These problems arise when the object of interest is one that cannot be accessed directly. A very well-known example is the so-called computed X-ray tomography, although it is perhaps less known that it is considered as an inverse problem. From measurements of X-rays that are sent through the body, the goal is to reconstruct a cross-sectional image of the body. The problem can be thought of as finding a cause to a measured effect. We measure the effect that the body had on the X-rays and the goal is figure out the cause of the effect, that is, the interior structure of the body that interacted with the X-rays.

We consider in this thesis specifically an inverse problem known as parameter identification for partial differential equations. Particularly important to our setting is the well-known Calderón problem [11]. A physical interpretation of the Calderón problem is to determine the electrical conductivity throughout a domain from measurements of the current and the voltage at the boundary of the domain. The original motivation of Calderón was supposedly applications to subsurface geology and, in particular, to discover oil reservoirs. But the Calderón problem has seen applications also in medical imaging, then known as electrical impedance tomography, for example serving as a tool in breast cancer detection [26]. The application to medical imaging is based on the observation that different body parts have different electrical conductivities.

In this thesis, we consider a problem which is neither motivated by oil prospection nor medical imaging, but still is related to the Calderón problem. Instead we have in mind an application towards material science and the diffusion of gas through a heterogeneous media. Instead of measuring the electrical current and voltage, in this application we measure the particle flux and particle density. From this information, we wish to identify the diffusion coefficient throughout the domain. A complication with this problem is the natural occurrence of a diffusion coefficient which attains different values in different directions, a so-called anisotropic diffusion coefficient. This is therefore related to the anisotropic Calderón problem, which is a version of the Calderón problem which is particularly difficult to solve. The goal is to investigate how to use the additional mathematical structure provided by multiscale modelling [24] to find natural solutions to the issues related to the inverse problem.

This work is motivated by the mathematical modelling research activities currently done within the frame of the Knowledge Foundation project "Multi-Barr" (Multilayer barrier coating technology for fibre-based packaging), project nr. KK20180036. Essentially, small amounts of oxygen crossing a heterogeneous thin layer (called the "barrier") are detected by a checkmate oxygen sensor that is supposed to be robust and stable. The scientific question is: What can we infer on the internal microstructure of the thin layer material, using only information from sensor measurements? This is where this master's thesis contributes via a preliminary study that combines the fine aspects of inverse problems, homogenization theory as well as FEniCS-based numerical simulations.

The thesis is structured as follows. In Chapter 2, we collect some fundamental background that is needed throughout the remainder of the thesis. In Chapter 3, we study the forward problem. We perform the homogenization procedure and obtain some estimates that provide

the key to applying homogenization arguments to the inverse problem. We derive also the continuity and the differentiability of the homogenized forward operators. In Chapter 4, we use the results from the homogenization procedure to present three formulations of the parameter identification problems and we investigate their well-posedness. We present an optimization approach to the inversion and study briefly some issues regarding its well-posedness, again using the results from homogenization as the key to make progress on the well-posedness. In Chapter 5, we investigate numerically the inversion of a particular anisotropic coefficient using the optimization approach presented in Chapter 4. In Chapter 6, we conclude the thesis with some thoughts on the obtained results and discuss a few ways in which the work can be continued.

2 Preliminaries

2.1 Function spaces and related theorems

Let $\Omega \subset \mathbb{R}^2$ be an open bounded subset. By $C^k(\Omega)$ we denote the space of functions $f: \Omega \rightarrow \mathbb{R}$ for which the derivatives $f^{(l)}$ exist and are continuous for $l \leq k$. The space of infinitely differentiable functions is defined as the intersection $C^\infty(\Omega) := \bigcap_{k \in \mathbb{N}} C^k(\Omega)$. The space $C_0^\infty(\Omega)$ is the subset of functions of $C^\infty(\Omega)$ that have compact support in Ω , that is, the functions $f \in C^\infty(\Omega)$ for which there exists a compact subset $K \subset \Omega$ such that $f = 0$ on $\Omega \setminus K$.

Rather than continuously differentiable functions, we most often use functions belonging to Lebesgue spaces or Sobolev spaces. Unless otherwise stated, measurability and integrability is considered with respect to the Lebesgue measure. By $L^2(\Omega)$ we denote the space of functions $f: \Omega \rightarrow \mathbb{R}$ that are measurable and satisfy $\int_\Omega f(x)^2 dx < \infty$ and where functions that differ on at most a set of measure 0 are identified. We equip $L^2(\Omega)$ with the inner product $\langle f, g \rangle_{L^2(\Omega)} := \int_\Omega f(x)g(x) dx$ and denote by $\|f\|_{L^2(\Omega)}$ the norm induced by this inner product. By replacing the Lebesgue measure with the surface measure on $\partial\Omega$, we define $L^2(\partial\Omega)$, the inner product $\langle \cdot, \cdot \rangle_{L^2(\partial\Omega)}$, and the induced norm $\|\cdot\|_{L^2(\partial\Omega)}$ just as for $L^2(\Omega)$. We denote by $L_\diamond^2(\partial\Omega)$ the space of functions $f \in L^2(\partial\Omega)$ such that $\int_{\partial\Omega} f(x) d\sigma(x) = 0$, where σ is the surface measure. By $L^\infty(\Omega)$ we denote the space of functions $f: \Omega \rightarrow \mathbb{R}$ that are measurable and for which $\text{esssup}_{x \in \Omega} |f(x)| < \infty$ and where functions that differ on at most a set of measure 0 are identified. $L^\infty(\Omega)$ is equipped with the norm $\|f\|_{L^\infty(\Omega)} := \text{esssup}_{x \in \Omega} |f(x)|$.

We define the Sobolev space $H^1(\Omega)$ to contain the functions $f \in L^2(\Omega)$ for which there exist $\partial_{x_1} f, \partial_{x_2} f \in L^2(\Omega)$ that satisfy $\int_\Omega f(x) \partial_{x_i} \varphi(x) dx = - \int_\Omega \partial_{x_i} f(x) \varphi(x) dx$ for every $\varphi \in C_0^\infty(\Omega)$ and for $i \in \{1, 2\}$. $H^1(\Omega)$ is equipped with the inner product $\langle f, g \rangle_{H^1(\Omega)} := \langle f, g \rangle_{L^2(\Omega)} + \sum_{i=1}^2 \langle \partial_{x_i} f, \partial_{x_i} g \rangle_{L^2(\Omega)}$. Rather than using the norm induced by this inner product, we use the equivalent norm $\|f\|_{H^1(\Omega)} := \|f\|_{L^2(\Omega)} + \|\nabla f\|_{L^2(\Omega)}$. We often use the subspace of $H^1(\Omega)$ whose integrals over Ω have a fixed value. Let $m \in \mathbb{R}$. Then $H_\diamond^1(\Omega; m)$ is defined to be the space of function $f \in H^1(\Omega)$ for which $\int_\Omega f(x) dx = m$. We define the fractional Sobolev space $H^{1/2}(\partial\Omega)$ intuitively to be the image of $H^1(\Omega)$ under the trace operator, and $H^{-1/2}(\partial\Omega)$ to be its dual space. We never use the inner product or norm of fractional Sobolev spaces and therefore omit their definition. We sometimes need integral constraints also on the elements of $H^{-1/2}(\partial\Omega)$. We denote by $H_\diamond^{-1/2}(\partial\Omega)$ the set of functionals $f \in H^{-1/2}(\partial\Omega)$ such that

$\langle f, 1 \rangle_{L^2(\partial\Omega)} = 0$.¹

If $H(\Omega)$ is one of the above function spaces then we denote by $H_{\#}(\Omega)$ the set of Ω -periodic functions whose restriction to Ω belong to $H(\Omega)$. If H has an inner product or norm, then we equip $H_{\#}$ with the same inner product and norm. We use in particular $H_{\#}^1(\Omega)$ but also $C_{\#}^{\infty}(\Omega)$ and $L_{\#}^{\infty}(\Omega)$.

In addition to Ω , now consider an open subset $Y \subset \mathbb{R}^2$. We sometimes need to use functions $f(x, y)$ for $x \in \Omega$ and $y \in Y$ where $f(x, \cdot)$ belongs to one type of function space and $f(\cdot, y)$ belongs to another. This is accomplished by using Bochner spaces. Let $H_1(\Omega), H_2(Y)$ be two of the above function space. Then we define $H_1(\Omega; H_2(Y))$ to contain the functions $f(x, y)$ such that $f(x, \cdot) \in H_2(Y)$ for each $x \in \Omega$ and $f(\cdot, y) \in H_1(\Omega)$ for each $y \in Y$. We use, in particular, $L^2(\Omega; C_{\#}(Y)), L^2(\Omega; H_{\#}^1(Y))$ and $C_0^{\infty}(\Omega; C_{\#}^{\infty}(Y))$.

We often work with product spaces of the above defined spaces. If V is a vector space then $[V]^n$ is the space of n -tuples with components in V and $[V]^{m \times n}$ is the space of $m \times n$ matrices with components in V . If V has a norm $\|\cdot\|_V$ then we define for $f \in [V]^n$ and $g \in [V]^{m \times n}$ the product space norms $\|f\|_{[V]^n} := \sum_{i=1}^n \|f_i\|_V$ and $\|g\|_{[V]^{m \times n}} := \sum_{i=1}^m \sum_{j=1}^n \|g_{ij}\|_V$. We use most commonly the product space $[L^{\infty}(\Omega)]^{2 \times 2}$, which is defined in this way.

Let us end this section by recalling a few important theorems, but first, a caveat. The following three results are stated here as they are stated in [12], but neither proofs nor reference to proofs are provided in [12]. Proofs of very similar results can be found in [2]. The main difference between the following results and those in [2] is that instead of $H^{1/2}(\partial\Omega)$ and $L^2(\partial\Omega)$ the results in [2] apply to $H^{1/2}(\Omega_k)$ and $L^2(\Omega_k)$, where Ω_k is the intersection of Ω and a k -dimensional plane. It is claimed in [2] that the results generalize to the intersection with other smooth manifolds.

Theorem 2.1 ([12]). *Let $\Omega \subset \mathbb{R}^n$ be an open bounded set with Lipschitz continuous boundary $\partial\Omega$. Then $W^{1,n}(\Omega)$ is compactly embedded in $L^q(\Omega)$ for any $q \in [1, \infty)$.*

Theorem 2.2 ([12]). *Let $\Omega \subset \mathbb{R}^n$ be an open bounded set with Lipschitz continuous boundary $\partial\Omega$. Let $\gamma: H^1(\Omega) \rightarrow L^2(\partial\Omega)$ be the trace operator. Then there exists a constant $c > 0$ depending on γ and Ω such that $\|\gamma(u)\|_{H^{1/2}(\partial\Omega)} \leq c\|u\|_{H^1(\Omega)}$ for all $u \in H^1(\Omega)$.*

Theorem 2.3 ([12]). *Let $\Omega \subset \mathbb{R}^n$ be an open bounded set with Lipschitz continuous boundary $\partial\Omega$. Then $H^{1/2}(\partial\Omega)$ is compactly embedded in $L^2(\partial\Omega)$.*

The significance of theorems 2.2 and 2.3 in the present context is that, as a result, the trace operator $\gamma: H^1(\Omega) \rightarrow L^2(\partial\Omega)$ is compact.

Theorem 2.4 ([13]). *Let $B: H \times H \rightarrow \mathbb{R}$ be a bilinear form and let $L: H \rightarrow \mathbb{R}$ be a bounded linear functional. If there exist constants $\alpha, \beta > 0$ such that*

$$|B(u, v)| \leq \alpha\|u\|_H\|v\|_H \quad \forall u, v \in H$$

and

$$B(u, u) \geq \beta\|u\|_H^2 \quad \forall u \in H,$$

¹It may look like a typo to use the L^2 -pairing for a functional in $H^{-1/2}$, but it is intended to be the L^2 -pairing.

then there exists an element $u \in H$ such that

$$B(u, v) = L(v) \quad \forall v \in H.$$

Theorem 2.5 ([13]). *Let Ω be a bounded, connected, open subset of \mathbb{R}^n , with C^1 boundary $\partial\Omega$. Assume $1 \leq p \leq \infty$. Then there exists a constant C , depending only on n, p and U , such that*

$$\|u - (u)_\Omega\|_{L^p(\Omega)} \leq C \|\nabla u\|_{L^p(\Omega)},$$

where $(u)_\Omega = \frac{1}{|\Omega|} \int_\Omega u(x) dx$.

Theorem 2.6 ([20]). *Let X, Y be normed spaces and $T: X \rightarrow Y$ a compact linear operator. Suppose that x_n is a weakly convergent sequence in X with weak limit x . Then Tx_n is a strongly convergent sequence in Y with limit Tx .*

2.2 Two-scale convergence and homogenization

In this section we define the concept of two-scale convergence, present two important results regarding the existence of two-scale limits and briefly discuss the application to homogenization of linear elliptic partial differential equations. In Chapter 3, we go through the homogenization procedure in some detail, following closely [24].

Definition 2.1. Let (u_ε) be a sequence in $L^2(\Omega)$. We say that (u_ε) converges two-scale to a unique function $u_0(x, y) \in L^2(\Omega \times Y)$ if and only if for any $v \in C_0^\infty(\Omega; C_\#^\infty(Y))$, we have

$$\lim_{\varepsilon \rightarrow 0} \int_\Omega u_\varepsilon(x) v\left(x, \frac{x}{\varepsilon}\right) dx = \frac{1}{|Y|} \int_\Omega \int_Y u_0(x, y) v(x, y) dy dx.$$

We need two existence results, the proofs of which can be found in [22].

Theorem 2.7. *Let $(u_\varepsilon) \subset L^2(\Omega)$ be a bounded sequence. Then there exist a subsequence which two-scale converges to $u_0(x, y) \in L^2(\Omega \times Y)$.*

Theorem 2.8. *Let $(u_\varepsilon) \subset H^1(\Omega)$ be a bounded sequence which converges weakly to $u \in H^1(\Omega)$. Then there exists $u_0 \in H^1(\Omega \times Y)$ and $u_1 \in L^2(\Omega; H_\#^1(Y))$ such that $\nabla_x u_\varepsilon$ converges two-scale to $\nabla_x u_0 + \nabla_y u_1$. Furthermore, u_0 is independent of y and $u_0(x, y) = u(x)$.*

When applying two-scale convergence to the homogenization of multiscale partial differential equations, we think of Ω as the macroscopic domain and Y as the microscopic domain. Consider specifically $Y = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. Let $k \in \mathbb{Z}^2$ and let $Y_k = Y + \sum_{i=1}^2 k_i e_i$, where $e_i \in \mathbb{R}^2$ is the i th standard unit vector.² Then Y_k is simply the set Y translated in a periodic manner. In order to interpret Y as the periodic microstructure in the domain Ω , we use the relation

$$\Omega = \bigcup_{k \in \mathbb{Z}^2} \{\varepsilon Y_k : \varepsilon Y_k \subset \Omega\}. \quad (2.1)$$

²We want a Y -periodic microstructure. We use the shifts e_i since, with Y being the unit square, the microstructure is 1-periodic along each coordinate axis. If Y is not the unit square then we pick different vectors e_i .

By defining a periodic function on Y , (2.1) gives us an oscillating function on Ω . The smaller ε is, the faster the oscillations. In the context of a differential equation, the oscillations occur to the various data of the equation. Consider the equation

$$\operatorname{div}(-k_\varepsilon(x)\nabla u(x)) = f_\varepsilon(x) \quad x \in \Omega \quad (2.2)$$

with ε -dependent k_ε and f_ε . Since the data depends on ε , also the unknown u depends on ε , $u = u_\varepsilon$. Let $k: \Omega \times Y \rightarrow \mathbb{R}$, $f: Y \rightarrow \mathbb{R}$ and take $k_\varepsilon(x) = k(x, \frac{x}{\varepsilon})$, $f_\varepsilon(x) = f(\frac{x}{\varepsilon})$. Then we have obtained a function f_ε whose variations are entirely due to the microstructure but a coefficient k_ε for which the microstructure may have different effect at different points in Ω , due to the first argument x of k . The goal of homogenization is to obtain an equation which describes the limiting behaviour of (2.2) as $\varepsilon \rightarrow 0$. We do this by applying two-scale convergence to the weak formulation of (2.2) and Theorem 2.7 and Theorem 2.8 help us to ensure the existence of the limit.

2.3 Necessary optimality conditions as weak formulations

Obtaining a weak formulation of a single linear elliptic partial differential equation with Neumann and Dirichlet boundary conditions is done by a straightforward procedure of multiplying the equation by a test function and integrating by parts. A reoccurring complication in this thesis is an integral constraint to the equation. In such a situation it is not clear how to obtain a weak formulation with the standard approach of multiplying by a test function and integrating. This is resolved by making a connection between weak formulations of partial differential equations and necessary optimality conditions of a related constrained functional. Then Lagrange multipliers handle the unusual constraints to the partial differential equation, such as integral constraints. To this end, we need the following theorem.

Definition 2.2. Let T be a continuously Fréchet differentiable transformation from an open set D in a Banach space X into a Banach space Y . If $x_0 \in D$ is such that $T'(x_0)$ maps X onto Y , then the point x_0 is said to be a *regular point* of the transformation T .

Theorem 2.9 ([21]). *Let X, Y be Banach spaces, $f: X \rightarrow \mathbb{R}$ be a continuously Fréchet differentiable functional and $H: X \rightarrow Y$ a continuously Fréchet differentiable operator. If f has a local extremum under the constraint $H(x) = 0$ at the regular point x_0 , then there exists an element $y_0^* \in Y^*$ such that the Lagrangian functional*

$$L(x) = f(x) + y_0^* H(x)$$

is stationary at x_0 , that is,

$$f'(x_0) + y_0^* H'(x_0) = 0.$$

Throughout this section we use the notation $H_g^1(\Omega; \Gamma_D)$ where $\Gamma_D \subset \partial\Omega$ and $g \in H^{1/2}(\partial\Omega)$ to denote the set of functions $u \in H^1(\Omega)$ such that $u|_{\Gamma_D} = g$. Now define the functional $J: H_g^1(\Omega; \Gamma_D) \rightarrow \mathbb{R}$ by

$$J(u) = \int_{\Omega} \frac{1}{2} |\nabla u|^2 - f u \, dx + \int_{\Gamma_N} h u \, ds, \quad (2.3)$$

where $\Gamma_N = \partial\Omega \setminus \Gamma_D$. Consider the following minimization problem

$$\min_{u \in H_g^1(\Omega; \Gamma_D)} J(u).$$

A necessary condition for optimality is that Fréchet derivative vanishes,

$$J'(u)v = 0, \tag{2.4}$$

for all admissible variations v . A function $v \in H^1(\Omega)$ is an admissible variation of u if $u + v \in H_g^1(\Omega; \Gamma_D)$. It follows that the admissible variations are in $H_0^1(\Omega; \Gamma_D)$. Then the necessary condition (2.4) reads

$$\int_{\Omega} \nabla u \cdot \nabla v - f v \, dx + \int_{\Gamma_N} h v \, ds = 0 \tag{2.5}$$

for all $v \in H_0^1(\Omega; \Gamma_D)$. But (2.5) is just the weak form of

$$\begin{aligned} -\operatorname{div} \nabla u(x) &= f & x \in \Omega \\ u(x) &= g(x) & x \in \Gamma_D \\ -\nabla u(x) \cdot n(x) &= h(x) & x \in \Gamma_N. \end{aligned} \tag{2.6}$$

So there is a close relation between the weak(or variational) formulation of a linear elliptic partial differential equation and the necessary optimality condition that the Fréchet derivative vanishes for a certain functional.

If instead of (2.6), we have the partial differential equation

$$\begin{aligned} -\operatorname{div} \nabla u(x) &= f & x \in \Omega \\ -\nabla u(x) \cdot n(x) &= h(x) & x \in \partial\Omega \\ \int_{\Omega} u(x) \, dx &= 0 \end{aligned} \tag{2.7}$$

then it is not so clear how to obtain a weak formulation by the standard approach of multiplying by a test function and integrating. However, in the optimization framework just described, it makes sense to define the weak form of (2.7) to be the first order optimality condition of the minimization problem

$$\begin{aligned} \min_{u \in H^1(\Omega)} & J(u) \\ \text{subject to} & \int_{\Omega} u(x) \, dx = 0, \end{aligned}$$

where the functional J is defined by

$$J(u) = \int_{\Omega} \frac{1}{2} |\nabla u|^2 - f u \, dx + \int_{\partial\Omega} h u \, ds$$

In the context of theorem 2.9, $H(u) = \int_{\Omega} u \, dx$ and $f(u) = J(u)$. To see that $H'(u)$ is surjective, take a constant increment $v = c|\Omega|^{-1}$ for any $c \in \mathbb{R}$. Then $H'(u)v = \int_{\Omega} c|\Omega|^{-1} \, dx = c$. Now introduce the Lagrangian $L: H^1(\Omega) \times \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$L(u, \lambda) = J(u) + \lambda H(u).$$

Then theorem 2.9 states that

$$\begin{aligned} 0 &= \frac{\partial L(u, \lambda)}{\partial u} v = J'(u)v + \lambda H'(u)v \\ &= \int_{\Omega} \nabla u(x) \cdot \nabla v(x) - f(x)v(x) + \lambda v(x) dx + \int_{\partial\Omega} h(x)v(x) ds(x). \end{aligned} \quad (2.8)$$

In addition, since u is an admissible function,

$$0 = \frac{\partial L(u, \lambda)}{\partial \lambda} \mu = \mu \int_{\Omega} u(x) dx, \quad (2.9)$$

for any $\mu \in \mathbb{R}$. Adding (2.9) to (2.8) yields the combined necessary condition

$$0 = \int_{\Omega} \nabla u(x) \cdot \nabla v(x) - f(x)v(x) + \lambda v(x) + \mu u(x) dx + \int_{\partial\Omega} h(x)v(x) ds(x).$$

This is now an equation in the space $H^1(\Omega) \times \mathbb{R}$ in which (v, μ) take the role of test functions and (u, λ) are the unknowns.

3 The forward problem

Consider a connected open set $\Omega \subset \mathbb{R}^2$ with two open subsets Ω_0 and Ω_1 such that $\Omega_0 = \Omega \setminus \bar{\Omega}_1$ and Ω_0 is a disconnected open set. Let Ω_l and Ω_r be two disjoint open sets such that $\Omega_0 = \Omega_l \cup \Omega_r$.

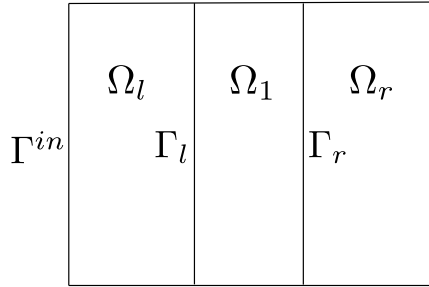


Figure 3.1: A simple example geometry.

The diffusion coefficient k_ε in Ω is given by

$$k_\varepsilon(x) = \begin{cases} k_0(x), & x \in \Omega_0 \\ k_1\left(\frac{x}{\varepsilon}\right), & x \in \Omega_1. \end{cases}$$

Oxygen enters the domain Ω through a subset $\Gamma^{in} \subset \partial\Omega_l \setminus \partial\Omega_1$. The oxygen spreads by diffusion through Ω_l, Ω_1 , and Ω_r . Apart from Γ^{in} , the boundary of Ω is impenetrable. Let $\Gamma_0 = \partial\Omega \setminus \Gamma^{in}$ be the impenetrable part of the boundary. However, the oxygen can pass

through the boundary Γ_l between Ω_l and Ω_1 , and the boundary Γ_r , between Ω_1 and Ω_r . The stationary diffusion equation for this system is

$$(P_\varepsilon) \quad \begin{cases} \operatorname{div}(-k_\varepsilon(x)\nabla u_\varepsilon(x)) = 0 & \text{for } x \in \Omega \\ -k_0(x)\nabla u_\varepsilon(x) \cdot n(x) = g(x) & \text{for } x \in \Gamma^{in} \\ -k_\varepsilon(x)\nabla u_\varepsilon(x) \cdot n(x) = 0 & \text{for } x \in \Gamma_0 \\ \int_\Omega u_\varepsilon(x) dx = m. \end{cases} \quad (3.1)$$

Since (3.1) is a Neumann problem, the flux boundary conditions are not sufficient for uniquely determining the solution. The integral constraint in (3.1) is included to ensure the uniqueness of a solution, the physical interpretation of the integral constraint being that the total mass of oxygen throughout Ω be m .

Now the forward problem is: Find the density u_ε of oxygen throughout the domain Ω , given the diffusion coefficient $k_\varepsilon(\cdot)$, the oxygen flux g through Γ^{in} and given that the total oxygen mass in Ω is m .

3.1 Weak formulations of the forward problem

There are at least two ways to obtain reasonable weak formulations of (P_ε) . One way is to incorporate the integral constraint into the function space definition, $H_\diamond^1(\Omega; m)$, and then take the weak formulation in $H_\diamond^1(\Omega; m)$. This formulation is suitable for theoretical investigations of (P_ε) . It is for example straightforward to prove well-posedness of the weak formulation in $H_\diamond^1(\Omega; m)$ by applying Lax-Milgram's Lemma and it is in the space $H_\diamond^1(\Omega; m)$ that we apply two-scale convergence. But the weak formulation in $H_\diamond^1(\Omega; m)$ has the disadvantage that it is not obvious how to formulate it in FEniCS. Another way to obtain a weak formulation of (P_ε) is to incorporate the integral constraint into the weak form by using Lagrange multipliers. This is advantageous from an implementation point of view and it is easily implemented in FEniCS. But a disadvantage is that the use of Lagrange multipliers to enforce the constraint removes important structure from the function space. One implication of this is that the generalized Poincaré inequality cannot be used to obtain an equivalent norm. Without the equivalent norm it is not that easy to apply Lax-Milgram's Lemma to obtain well-posedness, if it is even possible. So the Lagrange multiplier approach is very good from a practical point of view, but it severely limits the available tools to use in theoretical investigations. For this reason we will find both weak formulations useful, one for theoretical investigations and one for implementation.

Having already seen in Section 2.3 how to obtain weak formulations from necessary optimality conditions for constrained optimization problems, it is straightforward to obtain the following weak formulation with Lagrange multipliers.

Definition 3.1 (Weak solution with Lagrange multipliers). A pair $(u_\varepsilon, \lambda) \in H^1(\Omega) \times \mathbb{R}$ is said to be a weak solution of (P_ε) if it satisfies

$$\int_\Omega k_\varepsilon(x)\nabla u_\varepsilon(x) \cdot \nabla v(x) + \lambda v(x) + \mu u_\varepsilon(x) dx = \mu m - \int_{\Gamma^{in}} g(x)v(x) d\sigma(x) \quad \forall (v, \mu) \in H^1(\Omega) \times \mathbb{R}. \quad (3.2)$$

Let us now turn to the weak formulation in $H_\diamond^1(\Omega; m)$. There is one important thing to keep in mind here, and it is the fact that $H_\diamond^1(\Omega; m)$ is not a Hilbert space, because it is not even a vector space. Then in order to apply results such as Lax-Milgram's Lemma, the weak formulation needs to be shifted so that it is formulated in the Hilbert space $H_\diamond^1(\Omega; 0)$. Then we need to decide when to do this shift. The shift could for example be made already at the level of (P_ε) . But in order to keep the connection to the application, it seems reasonable to keep the equations as close to the physical model as possible, that is, not to shift the model (P_ε) into a 0 integral constraint. Whenever the structure of $H_\diamond^1(\Omega; 0)$ is required we can still temporarily shift the weak formulation of (P_ε) . So let us formulate the weak formulation in $H_\diamond^1(\Omega; m)$. We seek u_ε in $H_\diamond^1(\Omega; m)$ and then the test functions v should be functions in $H^1(\Omega)$ such that $u_\varepsilon + v \in H_\diamond^1(\Omega; m)$. It follows that $v \in H_\diamond^1(\Omega; 0)$. By multiplying the PDE in (P_ε) by v and integrating by parts we arrive at the following weak formulation.

Definition 3.2 (Weak solution). A function $u_\varepsilon \in H_\diamond^1(\Omega; m)$ is said to be a weak solution of (P_ε) if it satisfies

$$\int_{\Omega} k_\varepsilon(x) \nabla u_\varepsilon(x) \cdot \nabla v(x) dx = - \int_{\Gamma^{in}} g(x) v(x) d\sigma(x) \quad \forall v \in H_\diamond^1(\Omega; 0), \quad (3.3)$$

where $g \in H_\diamond^{-1/2}(\partial\Omega)$ with $\text{supp}(g) \in \Gamma^{in}$ and $k_\varepsilon \in [L^\infty(\Omega)]^{2 \times 2}$.

3.2 Well-posedness of the multiscale forward problem

We show the well-posedness of (P_ε) via its weak formulation (3.3) in $H_\diamond^1(\Omega; m)$. Lax-Milgram's Lemma requires the use of Hilbert spaces and we therefore shift the weak formulation (3.3) into a weak formulation in $H_\diamond^1(\Omega; 0)$. Let $f \in H_\diamond^1(\Omega; m)$ be arbitrary. Then u_ε solves (3.3) if and only if $w_\varepsilon = u_\varepsilon - f$ solves

$$\int_{\Omega} (k_\varepsilon \nabla w_\varepsilon) \cdot \nabla v dx = - \int_{\Omega} (k_\varepsilon \nabla f) \cdot \nabla v dx - \int_{\Gamma^{in}} g v d\sigma(x) \quad \forall v \in H_\diamond^1(\Omega; 0). \quad (3.4)$$

It is important to note that since any solution of either (3.3) or (3.4) corresponds to a solution of the other, both existence or uniqueness of either equation translates into existence or uniqueness of the other. Let us make the following assumptions:

- A1. Let $k_0 \in [L^\infty(\Omega_0)]^{2 \times 2}$ and $k_1 \in [L_{\#}^\infty(Y)]^{2 \times 2}$ such that for some $\alpha_0, \alpha_1 > 0$ the coercivity conditions $k_0(x)\xi \cdot \xi \geq \alpha_0|\xi|^2$ and $k_1(y)\xi \cdot \xi \geq \alpha_1|\xi|^2$ hold for a.e. $x \in \Omega_0, y \in Y$ and all $\xi \in \mathbb{R}^2$.
- A2. Let k_ε be defined by $k_\varepsilon(x) = k_0(x)$ for $x \in \Omega_0$ and $k_\varepsilon(x) = k_1(\frac{x}{\varepsilon})$ for $x \in \Omega_1$. Let $\alpha = \min\{\alpha_0, \alpha_1\}$. Then $k_\varepsilon \in [L^\infty(\Omega)]^{2 \times 2}$ and $k_\varepsilon(x)\xi \cdot \xi \geq \alpha|\xi|^2$ for a.e. $x \in \Omega$ and all $\xi \in \mathbb{R}^2$.
- A3. $f \in H_\diamond^1(\Omega; m)$
- A4. $g \in L^2(\Gamma^{in})$

A5. $\partial\Omega = \Gamma_0 \cup \Gamma^{in}$ is C^1 .

The bilinear form $B: H_\diamond^1(\Omega) \times H_\diamond^1(\Omega) \rightarrow \mathbb{R}$ and the linear form $L: H_\diamond^1(\Omega) \rightarrow \mathbb{R}$ that corresponds to the weak formulation (3.4) are defined by

$$\begin{aligned} B(w, v) &:= \int_{\Omega} k_\varepsilon(x) \nabla w(x) \cdot \nabla v(x) \\ L(v) &:= - \int_{\Omega} k_\varepsilon(x) \nabla f(x) \cdot \nabla v(x) dx - \int_{\Gamma^{in}} g(x) v(x) dx. \end{aligned}$$

The key step in making Lax-Milgram work is to use the equivalence of the norms $\|\nabla(\cdot)\|_{L^2(\Omega)}$ and $\|\cdot\|_{H^1(\Omega)}$ in $H_\diamond^1(\Omega; 0)$ that is obtained by Theorem 2.5. This is where A5. is needed, as Theorem 2.5 requires a C^1 boundary.

Using the coercivity of k_ε , we have

$$\begin{aligned} B(w, w) &= \int_{\Omega} k_\varepsilon(x) \nabla w(x) \cdot \nabla w(x) dx \geq \int_{\Omega} \alpha |\nabla w(x)|^2 dx \\ &= \alpha \|\nabla w\|_{L^2(\Omega)}^2 \geq \alpha C_N^2 \|w\|_{H^1(\Omega)}^2, \end{aligned} \quad (3.5)$$

where C_N is a constant used for the equivalence of norms. So B is coercive. Let us show the boundedness of B . First, we have by Cauchy-Schwarz inequality

$$|B(w, v)| \leq \int_{\Omega} |k_\varepsilon(x) \nabla w(x) \cdot \nabla v(x)| dx \leq \|k_\varepsilon \nabla w\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)}. \quad (3.6)$$

We now derive an estimate for $\|k_\varepsilon \nabla w\|_{L^2(\Omega)}$, which we will use many times throughout the thesis but we only go through the calculations in detail here. By using the inequality $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$, we have

$$\begin{aligned} \|k_\varepsilon \nabla w\|_{L^2(\Omega)}^2 &= \int_{\Omega} \sum_{i=1}^2 \left(k_\varepsilon^{i1}(x) \partial_{x_1} w(x) + k_\varepsilon^{i2}(x) \partial_{x_2} w(x) \right)^2 dx \\ &\leq \int_{\Omega} 2 \sum_{i=1}^2 \left((k_\varepsilon^{i1}(x) \partial_{x_1} w(x))^2 + (k_\varepsilon^{i2}(x) \partial_{x_2} w(x))^2 \right) dx \\ &\leq \|k_\varepsilon\|_{[L^\infty(\Omega)]^{2 \times 2}}^2 \int_{\Omega} 2 \sum_{i=1}^2 \left((\partial_{x_1} w(x))^2 + (\partial_{x_2} w(x))^2 \right) dx \\ &= 4 \|k_\varepsilon\|_{[L^\infty(\Omega)]^{2 \times 2}}^2 \|\nabla w\|_{L^2(\Omega)}^2. \end{aligned} \quad (3.7)$$

Using (3.7) in (3.6), followed by equivalence of norms, yields

$$\begin{aligned} |B(w, v)| &\leq 2 \|k_\varepsilon\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla w\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \\ &\leq 2 \|k_\varepsilon\|_{[L^\infty(\Omega)]^{2 \times 2}} \|w\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}. \end{aligned} \quad (3.8)$$

To prove the boundedness of the functional L , we first use Cauchy-Schwarz inequality on both terms and then the trace inequality on the second term. This yields

$$\begin{aligned} |L(v)| &\leq \|k_\varepsilon \nabla f\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} + \|g\|_{L^2(\Gamma^{in})} \|v\|_{L^2(\Gamma^{in})} \\ &\leq (2 \|k_\varepsilon\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla f\|_{L^2(\Omega)} + \|g\|_{L^2(\Gamma^{in})}) \|v\|_{H^1(\Omega)}. \end{aligned} \quad (3.9)$$

Now (3.5),(3.8),(3.9) allow the application of Lax-Milgram's Lemma. Hence, it follows that there is a unique $w \in H_\diamond^1(\Omega; 0)$ such that $B(w, v) = L(v)$ for all $v \in H_\diamond^1(\Omega; 0)$. By shifting $u_\varepsilon = w + f$ we obtain the existence and uniqueness of a weak solution in $H_\diamond^1(\Omega; m)$ satisfying (3.3).

Obtaining the continuity of the weak solution of (3.3) on the data g follows easily from the observation that if $u_\varepsilon^1, u_\varepsilon^2$ are solutions of (3.3) corresponding to data $g_1, g_2 \in L^2(\Gamma^{in}), g_1 \neq g_2$, then $u_\varepsilon^1 - u_\varepsilon^2$ is a valid test function. By taking $v = u_\varepsilon^1 - u_\varepsilon^2$ and subtracting the weak formulation of u_ε from the weak formulation of u_ε^1 , we have

$$\int_{\Omega} k_\varepsilon(x) \nabla(u_\varepsilon^1(x) - u_\varepsilon^2(x)) \cdot \nabla(u_\varepsilon^1(x) - u_\varepsilon^2(x)) dx = - \int_{\Gamma^{in}} (g_1(x) - g_2(x)) (u_\varepsilon^1(x) - u_\varepsilon^2(x)) d\sigma(x).$$

Since $u_\varepsilon^1 - u_\varepsilon^2 \in H_\diamond^1(\Omega; 0)$, we can apply the equivalence of norms. Hence, using also the coercivity of k_ε and the trace inequality, we have

$$\begin{aligned} \|u_\varepsilon^1 - u_\varepsilon^2\|_{H^1(\Omega)}^2 &\leq \frac{1}{C_N^2} \|\nabla(u_\varepsilon^1 - u_\varepsilon^2)\|_{L^2(\Omega)}^2 \\ &\leq \frac{1}{\alpha C_N^2} \int_{\Omega} k_\varepsilon(x) \nabla(u_\varepsilon^1(x) - u_\varepsilon^2(x)) \cdot \nabla(u_\varepsilon^1(x) - u_\varepsilon^2(x)) dx \\ &= \frac{1}{\alpha C_n^2} - \int_{\Gamma^{in}} (g_1(x) - g_2(x)) (u_\varepsilon^1(x) - u_\varepsilon^2(x)) d\sigma(x) \\ &\leq \frac{C_T}{\alpha C_N^2} \|g_1 - g_2\|_{L^2(\Gamma^{in})} \|u_\varepsilon^1 - u_\varepsilon^2\|_{H^1(\Omega)}. \end{aligned}$$

It follows that

$$\|u_\varepsilon^1 - u_\varepsilon^2\|_{H^1(\Omega)} \leq \frac{C_T}{\alpha C_N^2} \|g_1 - g_2\|_{L^2(\Gamma^{in})}. \quad (3.10)$$

Hence that the solution to (3.3) indeed depends continuously on the boundary data g . We could also prove in a similar manner the continuity of u_ε on the diffusion coefficient k_ε . But while we need to use (3.10) for the homogenization procedure in the next subsection, we will not actually need the continuity of u_ε on k_ε . However, we will later need such a continuity property with respect to the effective diffusion coefficient for the solution of the weak formulation of the upscaled equation. We therefore postpone this proof until the upscaled equation is derived.

3.3 Homogenization by two-scale convergence

In this section, we apply the concept of two-scale convergence to obtain the homogenized equation, the effective diffusion coefficient and the cell problems associated to (3.1). We break down the process in two major steps. First, we justify the two-scale limit by ensuring the existence of the two-scale limit via compactness results and ensuring mass conservation of the two-scale limit process. Then, knowing that the two-scale limit exists, we pass to the homogenization limit and derive the effective diffusion coefficient, the upscaled equation and the cell problems. To summarize the end product in one place, we will derive the following.

The cell problems are to find w_i such that

$$(P_{cell}) \quad \begin{cases} \operatorname{div}_y \left(-k_1(y) \nabla_y w_i(x, y) \right) = \operatorname{div}_y (k_1^i(y)) & y \in Y, i \in \{1, 2\} \\ w_i \text{ is } Y\text{-periodic, } i \in \{1, 2\}. \end{cases} \quad (3.11)$$

The upscaled equation reads

$$(P_{eff}) \quad \begin{cases} \operatorname{div}_x \left(-D(x, k) \nabla_x u_0(x) \right) = 0 & x \in \Omega \\ -D(x, k) \nabla_x u_0(x) \cdot n(x) = g(x) & x \in \Gamma^{in} \\ -D(x, k) \nabla_x u_0(x) \cdot n(x) = 0 & x \in \Gamma_0 \\ \int_{\Omega} u(x) dx = m, \end{cases}$$

where effective diffusion coefficient is defined by

$$D(x, k_0, k_1) = \begin{cases} k_0(x) & x \in \Omega_0 \\ \frac{1}{|Y|} \int_Y k_1(y) (I + [\nabla_y w_1(y) \quad \nabla_y w_2(y)]) dy & x \in \Omega_1, \end{cases} \quad (3.12)$$

where w_i are the solutions to (P_{cell}) . Just as for the (P_{ε}) , the equations we actually use are the weak formulations of the above equations. By the standard procedure we arrive at the following two definitions of weak solutions.

Definition 3.3 (Weak solution of cell-problems). Let $k \in [L^{\infty}(Y)]^{2 \times 2}$. A function $w_i \in H_{\#}^1(Y) \cap H_{\diamond}^1(Y; 0)$ for $i \in \{1, 2\}$ is said to be a weak solution of (3.11) if it satisfies

$$\int_Y (k(y) \nabla w_i(y)) \cdot \nabla v(y) dy = \int_Y \operatorname{div} (k_i(y)) v(y) dy \quad \forall v \in H_{\#}^1(Y),$$

where k_i is the i th column of k .

Definition 3.4 (Weak solution of upscaled equation). A function $u \in H_{\diamond}^1(\Omega; m)$ is said to be a weak solution of (P_{eff}) if it satisfies

$$\int_{\Omega} (D(x, k_0, k_1) \nabla u(x)) \cdot \nabla v(x) dx = - \int_{\Gamma^{in}} g(x) v(x) d\sigma(x) \quad \forall v \in H^1(\Omega), \quad (3.13)$$

where $g \in H_{\diamond}^{-1/2}(\partial\Omega; 0)$ with $\operatorname{supp}(g) \in \Gamma^{in}$, the effective diffusion coefficient D is defined by (3.12).

3.3.1 Two-scale compactness

The first step towards obtaining the necessary compactness results is to obtain an ε -independent upper bound on the sequence of weak solutions u_{ε} to (P_{ε}) in the sense of Definition 3.2. If each u_{ε} was a valid test function, that is, if $u_{\varepsilon} \in H_{\diamond}^1(\Omega; 0)$, then such a bound is easily obtained from the coercivity and the boundedness of the bilinear form B and the linear form L . It is

not certain that $u_\varepsilon \in H_\diamond^1(\Omega; 0)$ but we can obtain useful information by considering the shifted weak formulation (3.4) with $f = \frac{m}{|\Omega|}$ and $w_\varepsilon = u_\varepsilon - f \in H_\diamond^1(\Omega; 0)$. From (3.5) and (3.9) we have

$$\|w_\varepsilon\|_{H^1(\Omega)}^2 \leq \frac{1}{\alpha C_N^2} B(w_\varepsilon, w_\varepsilon) = \frac{1}{\alpha C_N^2} L(w_\varepsilon) \leq \frac{1}{\alpha C_N^2} \|g\|_{L^2(\Gamma^{in})} \|w_\varepsilon\|_{H^1(\Omega)}.$$

We conclude that $\|w_\varepsilon\|_{H^1(\Omega)} \leq \frac{1}{\alpha C_N} \|g\|_{L^2(\Gamma^{in})}$. Since $u_\varepsilon = w_\varepsilon + \frac{m}{|\Omega|}$, we have

$$\|u_\varepsilon\|_{H^1(\Omega)} \leq \|w_\varepsilon\|_{H^1(\Omega)} + \left\| \frac{m}{|\Omega|} \right\|_{H^1(\Omega)} \leq \frac{1}{\alpha C_N} \|g\|_{L^2(\Gamma^{in})} + \frac{m}{\sqrt{|\Omega|}}. \quad (3.14)$$

From (3.14) we get ε -independent bounds under the assumptions that

- A6. The flux g and the boundary Γ^{in} are both ε -independent.
- A7. The coercivity constant α of k_ε is ε -independent.
- A8. The size of the domain Ω is ε -independent.

It can be shown that C_N is ε -independent so that no additional assumption is necessary for that constant. The mass m is ε -independent. The application we have in mind confines the ε -dependence to functions defined on Ω_1 , so that assuming g to be ε -independent is natural. The domain Ω being ε -independent is also natural since we have no pores. The only assumption that is questionable is the ε -independence of the coercivity constant α . But this is a result of A1. and A2. combined.

Consequently, by accepting the above assumptions, we have an ε -independent bound on $\|u_\varepsilon\|_{H^1(\Omega)}$ and by the Eberlein-Smuljan Theorem there exists a subsequence (also indexed by ε) u_ε which converges weakly to some $u_0 \in H^1(\Omega)$. Then u_ε also two-scale converges to the same limit u_0 and there exists a function $u_1 \in L^2(\Omega; H_\#^1(Y))$ and a subsequence (also indexed by ε) of u_ε such that ∇u_ε two-scale converges to $\nabla_x u_0 + \nabla_y u_1$; see Theorem 20 in [22].

We note at this stage that each function u_ε belongs to $H_\diamond^1(\Omega; m)$ but we currently know nothing about the integrals of u_0 or u_1 . Since the integrals represent the mass of $O_2(g)$ in the system it is physically important that the mass is preserved after passing to the two-scale limit. Indeed, this turns out to be true. We know that $u_\varepsilon \rightharpoonup u_0$ weakly in $H^1(\Omega)$. Since $H^1(\Omega)$ is compactly embedded into $L^2(\Omega)$, the identity operator $H^1(\Omega) \ni u \rightarrow u \in L^2(\Omega)$ is compact. Since $u_\varepsilon \rightharpoonup u_0$ in $H^1(\Omega)$ we have by Theorem 2.6 that $u_\varepsilon \rightarrow u_0$ strongly in $L^2(\Omega)$. Now we have

$$\left| \int_\Omega u_\varepsilon(x) dx - \int_\Omega u_0(x) dx \right| \leq \sqrt{|\Omega|} \|u_\varepsilon - u_0\|_{L^2(\Omega)},$$

and it follows that

$$\int_\Omega u_0(x) dx = \int_\Omega u_\varepsilon(x) dx = m.$$

Consequently, the two-scale limit u_0 does indeed belong to $H_\diamond^1(\Omega; m)$.

Remark. *As far as results that are necessary for the application of two-scale convergence, we are done. But when turning to the inverse problem we will need one more convergence result*

for u_ε , namely, convergence in $L^2(\partial\Omega)$. Recall that u_0 is obtained as the weak limit in $H^1(\Omega)$ of u_ε . Since the trace operator $\text{Tr}: H^1(\Omega) \rightarrow L^2(\partial\Omega)$ is compact, then Theorem 2.6 ensures that $\text{Tr}(u_\varepsilon) \rightarrow \text{Tr}(u_0)$ strongly in $L^2(\partial\Omega)$. To summarize the important results of this subsection, we have

$$\begin{aligned} u_0 &\in H_\diamond^1(\Omega; m), u_1 \in L^2(\Omega; H_\#^1(Y)), \\ u_\varepsilon &\xrightarrow{2} u_0, \\ u_\varepsilon &\rightarrow u_0 \text{ strongly in } L^2(\Omega), \\ \text{Tr}(u_\varepsilon) &\rightarrow \text{Tr}(u_0) \text{ strongly in } L^2(\partial\Omega), \\ \nabla u_\varepsilon &\xrightarrow{2} \nabla_x u_0 + \nabla_y u_1. \end{aligned} \tag{3.15}$$

3.3.2 The two-scale limit

With the convergence results (3.15) at hand, we are now ready to apply two-scale convergence to (3.1). The two-scale limits are applied to the weak formulation (3.3). The bilinear form B and linear form L for the weak formulation are defined by

$$\begin{aligned} B(u_\varepsilon, v) &:= \int_\Omega k_\varepsilon(x) \nabla u_\varepsilon(x) \cdot \nabla v(x) dx \\ L(v) &:= - \int_{\Gamma^{in}} g(x) v(x) dx. \end{aligned}$$

Now take test functions of the form $v(x) = v_0(x) + \varepsilon v_1(x, \frac{x}{\varepsilon})$ with $v_0 \in H_\diamond^1(\Omega; 0)$, $v_1 \in L^2(\Omega; H_\#^1(Y)/\mathbb{R})$. Then we wish to calculate the two-scale limit, as $\varepsilon \rightarrow 0$, for

$$0 = \lim_{\varepsilon \rightarrow 0} B(u_\varepsilon, v_0 + \varepsilon v_1) - L(v_0 + \varepsilon v_1). \tag{3.16}$$

The goal is to find an expression for the limit, rather than to find the value of the limit. But the value of the limit is easily seen to be 0 since for each $\varepsilon > 0$, the right-hand side is equal to 0 since each u_ε is a weak solution in the sense of Definition 3.2. Term by term, we have

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} B(u_\varepsilon, v_0) &= \lim_{\varepsilon \rightarrow 0} \int_\Omega k_\varepsilon(x) \nabla u_\varepsilon(x) \cdot \nabla v_0(x) dx \\ &= \frac{1}{|Y|} \int_\Omega \int_Y k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \cdot \nabla v_0(x) dx dy, \end{aligned}$$

where

$$k(x, y) = \begin{cases} k_0(x), & x \in \Omega_0 \\ k_1(y), & x \in \Omega_1. \end{cases}$$

Next,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} B(u_\varepsilon, \varepsilon v_1) &= \lim_{\varepsilon \rightarrow 0} \int_\Omega k_\varepsilon(x) \nabla u_\varepsilon(x) \cdot \varepsilon \nabla v_1\left(x, \frac{x}{\varepsilon}\right) dx \\ &= \lim_{\varepsilon \rightarrow 0} \int_\Omega k_\varepsilon(x) \nabla u_\varepsilon(x) \cdot \left(\varepsilon \nabla_x v_1\left(x, \frac{x}{\varepsilon}\right) + \nabla_y v_1\left(x, \frac{x}{\varepsilon}\right)\right) dx \\ &= \frac{1}{|Y|} \int_\Omega \int_Y k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \cdot \nabla_y v_1(x, y) dx dy. \end{aligned}$$

Here we assume that we have an ε -independent upper bound on the term involving $\nabla_x v_1$, so that the limit of that term is 0 due to the factor ε . For the linear form, we have

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} L(v_0 + \varepsilon v_1) &= \lim_{\varepsilon \rightarrow 0} \int_{\Gamma^{in}} g(x) \left(v_0(x) + \varepsilon v_1 \left(x, \frac{x}{\varepsilon} \right) \right) d\sigma(x) \\ &= \int_{\Gamma^{in}} g(x) v_0(x) d\sigma(x) + \lim_{\varepsilon \rightarrow 0} \int_{\Gamma^{in}} g(x) \varepsilon v_1 \left(x, \frac{x}{\varepsilon} \right) d\sigma(x) \\ &= \int_{\Gamma^{in}} g(x) v_0(x) d\sigma(x). \end{aligned}$$

Here the v_1 term is evaluated to 0 by using Cauchy-Schwarz inequality followed by the trace inequality and an ε -independent bound of the H_1 norm of v_1 . Combining the above, we find that (3.16) yields

$$\begin{aligned} &\frac{1}{|Y|} \int_{\Omega} \int_Y k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \cdot (\nabla_x v_0(x) + \nabla_y v_1(x, y)) dx dy \\ &= - \int_{\Gamma^{in}} g(x) v_0(x) d\sigma(x). \end{aligned} \tag{3.17}$$

To obtain the strong form of the effective equation and cell-problems, we first apply the divergence theorem to (3.17) and write

$$\begin{aligned} &\frac{1}{|Y|} \int_{\Omega} \int_Y \operatorname{div}_x \left(-k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) v_0(x) dx dy \\ &+ \frac{1}{|Y|} \int_{\partial\Omega} \int_Y \left(-k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) \cdot n(x) v_0(x) d\sigma(x) dy \\ &+ \frac{1}{|Y|} \int_{\Omega} \int_Y \operatorname{div}_y \left(-k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) v_1(x, y) dx dy \\ &= - \int_{\Gamma^{in}} g(x) v_0(x) d\sigma(x) \quad \forall (v_0, v_1) \in H_{\diamond}^1(\Omega; 0) \times L^2(\Omega; H_{\#}^1(Y)/\mathbb{R}) \end{aligned} \tag{3.18}$$

3.3.3 The cell problems

Taking $v_0 = 0$ in (3.18) yields

$$\frac{1}{|Y|} \int_{\Omega} \int_Y \operatorname{div}_y \left(-k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) v_1(x, y) dx dy = 0 \quad \forall v_1 \in L^2(\Omega; H_{\#}^1(Y)/\mathbb{R}).$$

Hence we conclude that

$$\operatorname{div}_y \left(-k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) = 0 \quad \text{a.e. } (x, y) \in \Omega \times Y. \tag{3.19}$$

Now assume that $u_1(x, y) = \nabla_x u_0(x) \cdot w(x, y)$, where $w(x, y) := (w_1(x, y), w_2(x, y))$ and $w_i \in L^2(\Omega; H_{\#}^1(Y))$, $i \in \{1, 2\}$ are the cell functions. Substituting the cell functions into

(3.19) yields for a.e. $(x, y) \in \Omega \times Y$ that

$$\begin{aligned}
0 &= \operatorname{div}_y \left(-k(x, y)(\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) \\
&= \operatorname{div}_y \left(-k(x, y)(\nabla_x u_0(x) + \nabla_y(\nabla_x u_0(x) \cdot w(x, y))) \right) \\
&= \operatorname{div}_y \left(-k(x, y)(I + \nabla_y W(x, y)) \right) \nabla_x u_0(x),
\end{aligned} \tag{3.20}$$

where $\nabla_y W(x, y)$ is the matrix whose i th column is $\nabla_y w_i(x, y)$, $i \in \{1, 2\}$. Now using the identity $\frac{\partial_{x_i} u_0(x)}{\partial_{x_j} u_0(x)} = \frac{\partial u_0(x)}{\partial u_0(x)} \frac{\partial x_j}{\partial x_i} = \delta_{ij}$, where $i, j \in \{1, 2\}$ and δ_{ij} is the Kronecker delta, we have $\frac{1}{\partial_{x_i} u_0(x)} \nabla_x u_0(x) = e_i$ where e_i is the unit vector with 1 in the i th position and 0 otherwise. Now dividing (3.20) by $\partial_{x_i} u_0(x)$ yields

$$\operatorname{div}_y \left(-k(x, y)(e_i + \nabla_y w_i(x, y)) \right) = 0 \quad (x, y) \in \Omega \times Y, i \in \{1, 2\}. \tag{3.21}$$

Let us now show that the assumption that $k|_{\Omega_0} = k_0$ is independent of y implies that for $x \in \Omega_0$ we have $\nabla_y w_i(x, \cdot) = 0$ and $w_i(x, \cdot) = 0$. This is important for showing that the effective coefficient reflects the assumption that no oscillations occur in Ω_0 , which we get back to once we have derived the effective coefficient. Let $x \in \Omega_0$ and consider the weak formulation of (3.21),

$$\int_Y k_0(x) \nabla_y w_i(x, y) \cdot v(y) dy = 0$$

with test function $v \in H_{\#}^1(Y)$. For any particular x , we have $w_i(x, \cdot) \in H_{\#}^1(Y)$, so that w_i is a valid test function. Then we have by coercivity that

$$\int_Y |\nabla_y w_i(x, y)|^2 \leq \frac{1}{c} \int_Y k_0(x) \nabla_y w_i(x, y) \cdot w_i(x, y) dy = 0$$

so that $\nabla_y w_i(x, y) = 0$ and $w_i(x, \cdot)$ is constant. For the sake of well-posedness, we supply the cell problem (3.21) with an integral constraint $\int_Y w_i(x, y) dy = 0$. Then we get $w_i(x, y) = 0$ for a.e. $(x, y) \in \Omega_0 \times Y$ and $i \in \{1, 2\}$.

For $x \in \Omega_1$, we have the cell problems

$$-\operatorname{div}_y (k_1^i(y)) + \operatorname{div}_y \left(-K_1(y_2) \nabla_y w_i(x, y) \right) = 0 \quad (x, y) \in \Omega \times Y, i \in \{1, 2\},$$

where k_1^i is the i th column of k_1 . Note that these equations have no x -dependent data, so we obtain cell problems depending only on y . Adding an integral constraint $\int_Y w_i(y) dy = 0$ for the sake of well-posedness, we have the cell problems to find $w_i \in H_{\#}^1(Y) \cap \dot{H}_{\diamond}^1(Y; 0)$ such that

$$\operatorname{div}_y \left(-k_1(y) \nabla_y w_i(y) \right) = \operatorname{div}_y (k_1^i(y)) \quad y \in Y, i \in \{1, 2\}. \tag{3.22}$$

3.3.4 The upscaled diffusion equation

Taking $v_1 = 0$ in (3.18) yields

$$\begin{aligned} & \frac{1}{|Y|} \int_{\Omega} \int_Y \operatorname{div}_x \left(-k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) v_0(x) dx dy \\ & + \frac{1}{|Y|} \int_{\partial\Omega} \int_Y \left(-k(x, y) (\nabla_x u_0(x) + \nabla_y u_1(x, y)) \right) \cdot n(x) v_0(x) d\sigma(x) dy \\ & = - \int_{\Gamma^{in}} g(x) v_0(x) d\sigma(x) \quad \forall v_0 \in H_{\diamond}^1(\Omega; 0). \end{aligned}$$

By substituting the cell functions $u_1(x, y) = \nabla_x u_0(x) w(x, y)$ and bringing the x -derivatives out of the Y -integrals, we have

$$\begin{aligned} & \frac{1}{|Y|} \int_{\Omega} \operatorname{div}_x \left(- \int_Y k(x, y) (I + \nabla_y W(x, y)) dy \nabla_x u_0(x) \right) v_0(x) dx \\ & + \frac{1}{|Y|} \int_{\partial\Omega} \left(- \int_Y k(x, y) (I + \nabla_y W(x, y)) dy \nabla_x u_0(x) \right) \cdot n(x) v_0(x) d\sigma(x) dy \\ & = - \int_{\Gamma^{in}} g(x) v_0(x) d\sigma(x). \end{aligned}$$

Now, define the effective diffusion coefficient $D(x, k)$ by

$$D(x, k) = \frac{1}{|Y|} \int_Y k(x, y) (I + \nabla_y W(x, y)) dy \quad x \in \Omega. \quad (3.23)$$

Since $\nabla_y w_i(x, y) = 0$ for $x \in \Omega_0$, we have $\nabla_y W(x, y) = 0$ for $x \in \Omega_0$. Also, taking into account the subdomain specific x and y dependence of k , (3.23) simplifies to

$$D(x, k) = \begin{cases} k_0(x) & x \in \Omega_0 \\ \frac{1}{|Y|} \int_Y k_1(y) (I + \nabla_y W(y)) dy & x \in \Omega_1. \end{cases}$$

Now, we can write the upscaled equation more compactly as

$$\begin{aligned} & \int_{\Omega} \operatorname{div}_x \left(-D(x, k) \nabla_x u_0(x) \right) v_0(x) dx \\ & + \int_{\partial\Omega} \left(-D(x, k) \nabla_x u_0(x) \right) \cdot n(x) v_0(x) d\sigma(x) dy \\ & = - \int_{\Gamma^{in}} g(x) v_0(x) d\sigma(x). \end{aligned}$$

Since this equation hold for any test function $v_0 \in H_{\diamond}^1(\Omega; 0)$ we obtain the following strong form of the upscaled model

$$\begin{aligned} \operatorname{div} \left(-D(x, k) \nabla u_0(x) \right) &= 0 & x \in \Omega \\ -D(x, k) \nabla u_0(x) \cdot n(x) &= g(x) & x \in \Gamma^{in} \\ -D(x, k) \nabla u_0(x) \cdot n(x) &= 0 & x \in \Gamma_0 \\ \int_{\Omega} u_0(x) dx &= m. \end{aligned}$$

3.4 Differentiability of the parameter-to-solution maps

In this section, we derive rigorously the continuity and differentiability of the solution u_0 of (P_{eff}) with respect to the effective diffusion coefficient, $D(x, k_0, k_1)$. We consider the continuity and differentiability with respect to both the full coefficient $D(x, k_0, k_1)$ and with respect to the underlying parameter k_1 in the cell-problems.

3.4.1 Differentiability with respect to the upscaled coefficient

Let $\alpha > 0$ be some fixed real number and consider the subset $\mathcal{U} \subset [L^\infty(\Omega)]^{2 \times 2}$ defined by $\mathcal{U} = \{k \in [L^\infty(\Omega)]^{2 \times 2} : y^T k(x) y \geq \alpha |y|^2 \text{ for a.e } x \in \Omega, \forall y \in \mathbb{R}^2\}$. In this section, we study the continuity and differentiability properties of the parameter-to-state map $\mathcal{G}_{eff} : \mathcal{U} \rightarrow H_\diamond^1(\Omega; m)$. This map is defined by $\mathcal{G}_{eff}(k) = u$, where u solves the equation

$$\int_{\Omega} (k \nabla u) \cdot \nabla v \, dx = \int_{\Gamma^{in}} g v \, d\sigma(x) \quad \forall v \in H_\diamond^1(\Omega; 0).$$

The proofs rely on the fact that the coercivity constant α is independent of the coefficients, which is why we consider the map \mathcal{G}_{eff} only on the subset \mathcal{U} . The proofs of Proposition 3.1 and Proposition 3.2 are straightforward adaptations of the proofs from chapter 5 of [18] to the case of matrix coefficients and in the space $H_\diamond^1(\Omega; m)$ for arbitrary m . Then we extend the argument to also prove twice differentiability.

Proposition 3.1. *The parameter-to-solution map $\mathcal{G}_{eff} : \mathcal{U} \rightarrow H_\diamond^1(\Omega; 0)$ is Lipschitz continuous.*

Proof. Let $k_1, k_2 \in \mathcal{U}$ and let $u_1 = \mathcal{G}_{eff}(k_1)$ and $u_2 = \mathcal{G}_{eff}(k_2)$. The goal is to show that there exists a constant $c > 0$ such that $\|u_1 - u_2\|_{H^1(\Omega)} \leq c \|k_1 - k_2\|_{[L^\infty(\Omega)]^{2 \times 2}}$.

The argument relies heavily on the use of equivalent norms in $H_\diamond^1(\Omega; 0)$, and therefore we need to shift the solutions u_1, u_2 into $H_\diamond^1(\Omega; 0)$. At one point in the argument, it is very important that the function used in the shift has a 0 gradient, otherwise we can still obtain an upper bound on u_1, u_2 in terms of the coefficients k_1, k_2 but it will not be a Lipschitz-type bound. Therefore, we need the shift to be constant. Let $h = \frac{m}{|\Omega|}$ where $|\cdot|$ denotes the 2 dimensional Lebesgue measure. Then $w_1 = u_1 - h, w_2 = u_2 - h \in H_\diamond^1(\Omega; 0)$ satisfy the weak formulations

$$\int_{\Omega} (k_1 \nabla w_1) \cdot \nabla v \, dx = - \int_{\Gamma^{in}} g v \, d\sigma(x) \tag{3.24}$$

$$\int_{\Omega} (k_2 \nabla w_2) \cdot \nabla v \, dx = - \int_{\Gamma^{in}} g v \, d\sigma(x), \tag{3.25}$$

for every $v \in H_\diamond^1(\Omega; 0)$. The key ingredient in the argument is to use these two weak forms to obtain an equation with two terms where one term has a factor k_1^T and the other has a factor

$k_1^T - k_2^T$. Subtracting (3.25) from (3.24) and rewriting a bit yields

$$\begin{aligned}
0 &= \int_{\Omega} (k_1 \nabla w_1) \cdot \nabla v - (k_2 \nabla w_2) \cdot \nabla v \, dx \\
&= \int_{\Omega} (k_1 \nabla (w_1 - w_2)) \cdot \nabla v - ((k_2 - k_1) \nabla w_2) \cdot \nabla v \, dx \\
&= \int_{\Omega} \nabla (w_1 - w_2) \cdot (k_1^T \nabla v) - \nabla w_2 \cdot ((k_2^T - k_1^T) \nabla v) \, dx.
\end{aligned} \tag{3.26}$$

By using coercivity on k_1 followed by using (3.26) with $v = w_1 - w_2$ and then using Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\alpha \|\nabla (w_1 - w_2)\|_{L^2(\Omega)}^2 &\leq \int_{\Omega} (k_1 \nabla (w_1 - w_2)) \cdot \nabla (w_1 - w_2) \, dx \\
&= \int_{\Omega} \nabla (w_1 - w_2) \cdot (k_1^T \nabla (w_1 - w_2)) \, dx \\
&= \int_{\Omega} \nabla w_2 \cdot ((k_2^T - k_1^T) \nabla (w_1 - w_2)) \, dx \\
&\leq \|\nabla w_2\|_{L^2(\Omega)} \|(k_2^T - k_1^T) \cdot \nabla (w_1 - w_2)\|_{L^2(\Omega)}.
\end{aligned}$$

Using the inequality $\|(k_2^T - k_1^T) \cdot \nabla (w_1 - w_2)\|_{L^2(\Omega)} \leq 2\|k_2^T - k_1^T\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla (w_1 - w_2)\|_{L^2(\Omega)}$, we have

$$\|\nabla (w_1 - w_2)\|_{L^2(\Omega)} \leq \frac{2}{\alpha} \|\nabla w_2\|_{L^2(\Omega)} \|k_1 - k_2\|_{[L^\infty(\Omega)]^{2 \times 2}} \tag{3.27}$$

We can now use the trace inequality and Poincaré's inequality to estimate $\|\nabla w_2\|_{L^2(\Omega)}$ independently of k_1, k_2 . We have

$$\begin{aligned}
\alpha \|\nabla w_2\|_{L^2(\Omega)}^2 &\leq \int_{\Omega} (k_2 \nabla w_2) \cdot \nabla w_2 \, dx \\
&= \int_{\Gamma^{in}} g w_2 \, d\sigma(x) \\
&\leq C_T \|g\|_{L^2(\Gamma^{in})} \|w_2\|_{H^1(\Omega)} \\
&\leq C_T (C_P + 1) \|g\|_{L^2(\Gamma^{in})} \|\nabla w_2\|_{L^2(\Omega)}.
\end{aligned} \tag{3.28}$$

This is where it is important that the shift h is constant. If we did not immediately pick h to be constant then the second row in (3.28) would have a term involving k_2 and ∇h . By taking h to be constant, this term involving k_2 vanishes. Using (3.28) in (3.27), we have

$$\|\nabla (w_1 - w_2)\|_{L^2(\Omega)} \leq \frac{2C_T(C_P + 1)}{\alpha^2} \|g\|_{L^2(\Gamma^{in})} \|k_1 - k_2\|_{L^\infty(\Omega)}.$$

Writing $u_1 - u_2 = (w_1 + h) - (w_2 + h)$ and using Poincaré's inequality yields

$$\|u_1 - u_2\|_{L^2(\Omega)} \leq \|u_1 - u_2\|_{H^1(\Omega)} \leq \frac{2C_T(C_P + 1)^2}{\alpha^2} \|g\|_{L^2(\Gamma^{in})} \|k_1 - k_2\|_{L^\infty(\Omega)} \tag{3.29}$$

By writing $u_1 = \mathcal{G}_{eff}(k_1), u_2 = \mathcal{G}_{eff}(k_2)$ we obtain the Lipschitz continuity of the parameter-to-state operator \mathcal{G}_{eff} . \square

Proposition 3.2. *The operator $k \rightarrow \mathcal{G}_{eff}(k)$ for $k \in \mathcal{U}$ is Fréchet differentiable, and the directional derivative $\mathcal{G}'_{eff}(k)h = w \in H^1_\diamond(\Omega; 0)$ is the unique solution of the equation*

$$-\int_{\Omega} (k\nabla w) \cdot \nabla v \, dx = \int_{\Omega} (h\nabla u_k) \cdot \nabla v \, dx \quad \forall v \in H^1_\diamond(\Omega; 0), \quad (3.30)$$

where $u_k = \mathcal{G}_{eff}(k)$.

Proof. The proof is by a verification of the definition of Fréchet differentiability, that is, to show that

$$\lim_{\|h\| \rightarrow 0} \frac{\|\mathcal{G}_{eff}(k+h) - \mathcal{G}_{eff}(k) - \mathcal{G}'_{eff}(k)h\|_{H^1(\Omega)}}{\|h\|_{[L^\infty(\Omega)]^{2 \times 2}}} = 0. \quad (3.31)$$

For ease of notation, let $u_k = \mathcal{G}_{eff}(k)$ and $u_h = \mathcal{G}_{eff}(k+h)$ and suppose for the moment that $w = \mathcal{G}'_{eff}(k)h$ exists. Since $w, u_k - u_h \in H^1_\diamond(\Omega; 0)$ we have by equivalence of norm and coercivity of k that

$$\begin{aligned} \|u_h - u_k - w\|_{H^1(\Omega)}^2 &\leq C_N \|\nabla(u_h - u_k - w)\|_{L^2(\Omega)}^2 \\ &\leq \frac{C_N}{\alpha} \int_{\Omega} (k\nabla(u_h - u_k - w)) \cdot \nabla(u_h - u_k - w) \, dx. \end{aligned} \quad (3.32)$$

By subtracting the weak formulation (3.24) with $k_1 = k$ from the weak formulation with $k_1 = k+h$ and rearranging slightly, we have

$$\int_{\Omega} (k\nabla(u_h - u_k)) \cdot \nabla v \, dx = - \int_{\Omega} (h\nabla u_h) \cdot \nabla v(x) \, dx. \quad (3.33)$$

Using (3.33) with $v = u_h - u_k - w$ followed by using (3.30), we have

$$\begin{aligned} \int_{\Omega} (k\nabla(u_h - u_k - w)) \cdot \nabla(u_h - u_k - w) \, dx &= \int_{\Omega} (-h\nabla u_h - k\nabla w) \cdot \nabla(u_h - u_k - w) \, dx \\ &= \int_{\Omega} (h\nabla(u_k - u_h)) \cdot \nabla(u_h - u_k - w) \, dx. \end{aligned} \quad (3.34)$$

Inserting (3.34) into (3.32) and then using Cauchy-Schwarz' inequality we have

$$\begin{aligned} \|u_h - u_k - w\|_{H^1(\Omega)}^2 &\leq \frac{C_N}{\alpha} \|h\nabla(u_k - u_h)\|_{L^2(\Omega)} \|u_h - u_k - w\|_{H^1(\Omega)} \\ &\leq 2 \frac{C_N}{\alpha} \|h\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla(u_k - u_h)\|_{L^2(\Omega)} \|u_h - u_k - w\|_{H^1(\Omega)} \end{aligned}$$

By the Lipschitz continuity of \mathcal{G}_{eff} we have

$$\begin{aligned} \|u_h - u_k - w\|_{H^1(\Omega)} &\leq \frac{2C_N}{\alpha} \|h\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla(u_k - u_h)\|_{L^2(\Omega)} \\ &\leq \frac{2C_N C}{\alpha} \|h\|_{[L^\infty(\Omega)]^{2 \times 2}}^2, \end{aligned}$$

and now (3.31) follows. Furthermore, it is easy to verify by the definition of the operator $h \rightarrow \mathcal{G}'_{eff}(k)h$ as a solution of (3.30) yields a linear operator and, as a consequence of Lax-Milgram's Lemma, it is a bounded operator. Therefore, it is a Fréchet derivative. \square

Proposition 3.3. *The operator $k \rightarrow \mathcal{G}'_{eff}(k)h$ for $k \in \mathcal{U}$ is Lipschitz continuous.*

Proof. Denote $w_1 = \mathcal{G}'_{eff}(k_1)h$ and $w_2 = \mathcal{G}'_{eff}(k_2)h$. Subtracting (3.30) with w_2 from the same equation with w_1 yields

$$\int_{\Omega} (k_2 \nabla w_2 - k_1 \nabla w_1) \cdot \nabla v \, dx = \int_{\Omega} (h \nabla(u_1 - u_2)) \cdot \nabla v \, dx, \quad (3.35)$$

where $u_1 = \mathcal{G}_{eff}(k_1)$ and $u_2 = \mathcal{G}_{eff}(k_2)$. Adding and subtracting $k_2 \nabla w_1$ in the integral on the left-hand side of (3.35) and rearranging slightly in the resulting expression, yields

$$\int_{\Omega} (k_2 \nabla(w_2 - w_1)) \cdot \nabla v \, dx = \int_{\Omega} ((k_1 - k_2) \nabla w_1) \cdot \nabla v + (h \nabla(u_1 - u_2)) \cdot \nabla v \, dx \quad (3.36)$$

Using the coercivity of k_2 and (3.36) followed by Cauchy-Schwarz inequality, we have

$$\begin{aligned} \|\nabla(w_2 - w_1)\|_{L^2(\Omega)}^2 &\leq \frac{1}{\alpha} \int_{\Omega} (k_2 \nabla(w_2 - w_1)) \nabla(w_2 - w_1) \, dx \\ &= \frac{1}{\alpha} \int_{\Omega} \left((k_1 - k_2) \nabla w_1 + (h \nabla(u_1 - u_2)) \right) \cdot \nabla(w_2 - w_1) \, dx \\ &\leq \left(\|(k_1 - k_2) \nabla w_1\|_{L^2(\Omega)} + \|h \nabla(u_1 - u_2)\|_{L^2(\Omega)} \right) \|\nabla(w_2 - w_1)\|_{L^2(\Omega)} \\ &\leq 2 \left(\|k_1 - k_2\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla w_1\|_{L^2(\Omega)} \right. \\ &\quad \left. + \|h\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla(u_1 - u_2)\|_{L^2(\Omega)} \right) \|\nabla(w_2 - w_1)\|_{L^2(\Omega)} \end{aligned}$$

Dividing the last expression by $\|\nabla(w_2 - w_1)\|_{L^2(\Omega)}$ and using the Lipschitz continuity of \mathcal{G}_{eff} to estimate the norm $\|\nabla(u_1 - u_2)\|_{L^2(\Omega)}$ yields

$$\|\nabla(w_2 - w_1)\|_{L^2(\Omega)} \leq 2 \|k_1 - k_2\|_{[L^\infty(\Omega)]^{2 \times 2}} \left(\|\nabla w_1\|_{L^2(\Omega)} + C \|h\|_{[L^\infty(\Omega)]^{2 \times 2}} \right) \quad (3.37)$$

By applying the coercivity property to (3.30) we have

$$\|\nabla w_1\|_{L^2(\Omega)} \leq \frac{1}{\alpha} \|h \nabla u_2\|_{L^2(\Omega)} \leq 2 \|h\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla u_2\|_{L^2(\Omega)}. \quad (3.38)$$

Now taking u_2 in place of w_2 in (3.28), we have that

$$\|\nabla u_2\|_{L^2(\Omega)} \leq \frac{C_T(C_P + 1)}{\alpha} \|g\|_{L^2(\Gamma^{in})}. \quad (3.39)$$

Inserting (3.38) and (3.39) into (3.37) yields

$$\|w_2 - w_1\|_{L^2(\Omega)} \leq 2 \|h\|_{[L^\infty(\Omega)]^{2 \times 2}} \left(\frac{2C_T(C_P + 1)}{\alpha} \|g\|_{L^2(\Gamma^{in})} + C \right) \|k_1 - k_2\|_{[L^\infty(\Omega)]^{2 \times 2}},$$

which proves the desired Lipschitz continuity. \square

Proposition 3.4. *The parameter-to-solution map \mathcal{G}_{eff} is twice Fréchet differentiable and the derivative $(\mathcal{G}''_{eff}(k)h)f = m \in H^1_{\diamond}(\Omega; 0)$ is the unique solution of the equation*

$$- \int_{\Omega} (k \nabla m) \cdot \nabla v \, dx = \int_{\Omega} (h \nabla w) \cdot \nabla v + (f \nabla w_k) \cdot \nabla v \, dx, \quad (3.40)$$

where $w_k = \mathcal{G}'_{eff}(k)h$ and $w = \mathcal{G}'_{eff}(k)f$.

Proof. The idea of this proof is very similar to that of Lemma 3.2 and we will verify that

$$\lim_{\|f\| \rightarrow 0} \frac{\|\mathcal{G}'_{eff}(k+f)h - \mathcal{G}'_{eff}(k)h - (\mathcal{G}''_{eff}(k)h)f\|_{H^1(\Omega)}}{\|f\|_{[L^\infty(\Omega)]^{2 \times 2}}} = 0. \quad (3.41)$$

Let $w_f = \mathcal{G}'_{eff}(k+f)h$ and $w_k = \mathcal{G}'_{eff}(k)h$ and let m be the solution of (3.40). We will show that $m = (\mathcal{G}'_{eff}(k)h)f$.

By equivalence of norm in $H^1_\diamond(\Omega; 0)$ and by coercivity of k , we have

$$\begin{aligned} \|w_f - w_k - m\|_{H^1(\Omega)}^2 &\leq C \|\nabla(w_f - w_k - m)\|_{L^2(\Omega)}^2 \\ &\leq \frac{C}{\alpha} \int_{\Omega} (k \nabla(w_f - w_k - m)) \cdot \nabla(w_f - w_k - m) dx. \end{aligned} \quad (3.42)$$

By definition, w_f and w_k satisfy (3.30) with coefficients $k+f$ and k , respectively. Subtracting (3.30) with coefficient k and w_k form the same equation with coefficient $k+f$ and w_f yields, after slight rearrangement,

$$\int_{\Omega} (k \nabla(w_f - w_k)) \cdot \nabla v dx = - \int_{\Omega} (h \nabla(u_f - u_k)) \cdot \nabla v + (f \nabla w_f) \cdot \nabla v dx, \quad (3.43)$$

where $u_f = \mathcal{G}_{eff}(k+f)$ and $u_k = \mathcal{G}_{eff}(k)$. Using (3.43) we have

$$\int_{\Omega} (k \nabla(w_f - w_k - m)) \cdot \nabla v dx = - \int_{\Omega} (k \nabla m) \cdot \nabla v + (h \nabla(u_f - u_k)) \cdot \nabla v + (f \nabla w_f) \cdot \nabla v dx. \quad (3.44)$$

Inserting (3.40) in (3.44) and estimating in terms of $\|f\|_{[L^\infty(\Omega)]^{2 \times 2}}$ we have

$$\begin{aligned} \int_{\Omega} (k \nabla(w_f - w_k - m)) \cdot \nabla v &= - \int_{\Omega} (h \nabla(u_f - u_k - w) + f \nabla(w_f - w_k)) \cdot \nabla v dx \\ &\leq (\|h \nabla(u_f - u_k - w)\|_{L^2(\Omega)} + \|f \nabla(w_f - w_k)\|_{L^2(\Omega)}) \|\nabla v\|_{L^2(\Omega)} \\ &\leq 2 \|h\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla(u_f - u_k - w)\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \\ &\quad + 2 \|f\|_{[L^\infty(\Omega)]^{2 \times 2}} \|\nabla(w_f - w_k)\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \\ &\leq 2 (\|h\|_{[L^\infty(\Omega)]^{2 \times 2}} C_1 \|f\|_{[L^\infty(\Omega)]^{2 \times 2}}^2 + C_2 \|f\|_{[L^\infty(\Omega)]^{2 \times 2}}^2) \|\nabla v\|_{L^2(\Omega)} \\ &\leq 2 (\|h\|_{[L^\infty(\Omega)]^{2 \times 2}} C_1 + C_2) \|f\|_{[L^\infty(\Omega)]^{2 \times 2}}^2 \|\nabla v\|_{L^2(\Omega)}. \end{aligned} \quad (3.45)$$

The leftmost $\|f\|_{[L^\infty(\Omega)]^{2 \times 2}}^2$ on the second to last row of (3.45) comes from the definition $w = \mathcal{G}'_{eff}(k)f$, because then it holds that $\|\nabla(u_f - u_k - w)\|_{L^2(Y)} \leq C \|f\|_{[L^\infty(\Omega)]^{2 \times 2}}^2$. Taking $v = w_f - w_k - m$ in (3.45) and inserting into (3.42) yields

$$\|w_f - w_k - m\|_{H^1(\Omega)} \leq \frac{2C (\|h\|_{[L^\infty(\Omega)]^{2 \times 2}} C_1 + C_2)}{\alpha} \|f\|_{[L^\infty(\Omega)]^{2 \times 2}}^2,$$

from which (3.41) follows. \square

3.4.2 Differentiability with respect to the coefficient of the cell-problems

While in the previous subsection we obtained the differentiability of the solution u_0 of (P_{eff}) with respect to the effective diffusion coefficient, we are also interested in such differentiability properties of u_0 with respect to the diffusion coefficient of the cell-problems. Using the composition $\mathcal{G}_{eff}(D(k))$, where $D(k)$ is the effective coefficient, we can obtain the derivative of $u_0 = \mathcal{G}_{eff}(D(k))$ with respect to k by the chain rule

$$\frac{d\mathcal{G}_{eff}(D(k, \mathcal{G}_{cell}(k)))}{dk} = \mathcal{G}'_{eff}(D(k, \mathcal{G}_{cell}(k))) \frac{dD(k, \mathcal{G}_{cell}(k))}{dk} h. \quad (3.46)$$

Hidden in the expression $\frac{dD}{dk}$ is also the Fréchet derivative \mathcal{G}'_{cell} . So in order to compute the derivative of \mathcal{G}_{eff} on the microstructure k , we first study the derivative $\mathcal{G}'_{cell}(k)$. Then the results regarding $\mathcal{G}'_{cell}(k)$ help us derive $\frac{dD}{dk}$. This can then be combined with the results of Section 3.4.1 to compute the full Fréchet derivative in (3.46). In summary, we have the operators

$$\begin{aligned} \mathcal{G}_{eff}: [L^\infty(\Omega)]^{2 \times 2} &\rightarrow L^2(\Omega) \\ \mathcal{G}_{cell}: [L^\infty(Y)]^{2 \times 2} &\rightarrow [L^2(Y)]^2 \\ D: [L^\infty(Y)]^{2 \times 2} \times [L^2(Y)]^{2 \times 2} &\rightarrow [L^\infty(\Omega)]^{2 \times 2} \end{aligned}$$

and we need to prove Fréchet differentiability of these operators. From Section 3.4.1, we already have existence of \mathcal{G}'_{eff} . The proof of existence of \mathcal{G}'_{cell} is very similar to the proof for \mathcal{G}'_{eff} except that the weak form that defines \mathcal{G}_{cell} has a right-hand side that depends on the parameter k . But the main novelty to the differentiability proofs now that we consider the microstructure is the derivative of the effective coefficient $D(k, \mathcal{G}_{cell}(k))$ with respect to the coefficient k .

Just as in Section 3.4.1, the proofs in this section rely on an assumption that the coercivity constant for the coefficients is independent of the coefficients. Therefore, let $\alpha > 0$ be fixed and define $\mathcal{U} = \{k \in [L^\infty(Y)]^{2 \times 2} : x^T k(y) x \geq |x|^2 \text{ for a.e. } y \in Y, \forall x \in \mathbb{R}^2\}$.

Proposition 3.5. *The operator $k \rightarrow \mathcal{G}_{cell}(k)$ for $k \in \mathcal{U}$ satisfies*

$$\|\mathcal{G}_{cell}(k) - \mathcal{G}_{cell}(h)\|_{\mathcal{H}} \leq (C_1 + C_2 \|h\|_{[L^\infty(Y)]^{2 \times 2}}) \|k - h\|_{[L^\infty(Y)]^{2 \times 2}}$$

where $C_1, C_2 > 0$ are independent of k, h and \mathcal{H} denotes either $[L^2(Y)]^2$ or $[H^1(Y)]^2$.

Proof. Let us start by obtaining the inequality when considering only one component in the range of \mathcal{G}_{cell} , that is, we consider one cell function w_i at the time. This result is then easily extended into the product space, since the equations that define the components of \mathcal{G}_{cell} are very similar. Let w_i be the i th component with $i \in \{1, 2\}$. Then w_i solves the equation

$$\int_Y (k(y) \nabla w_i(y)) \cdot \nabla v(y) dy = - \int_Y k_i(y) \cdot \nabla v(y) dy, \quad (3.47)$$

where k_i is the i th column of k . Consider $k, h \in \mathcal{U}$ and the corresponding solutions w_i^k, w_i^h . By subtracting the weak form corresponding to the coefficient h from the weak form with coefficient

k and then adding and subtracting the term $k\nabla w_i^h$, we have

$$\int_Y (k\nabla(w_i^k - w_i^h)) \cdot \nabla v \, dy = - \int_Y (k_i - h_i) \cdot \nabla v + ((k - h) \cdot \nabla w_i^h) \cdot \nabla v \, dy. \quad (3.48)$$

Using the coercivity on k and Cauchy-Schwarz' inequality, we have

$$\begin{aligned} \|\nabla(w_i^k - w_i^h)\|_{L^2(Y)}^2 &\leq -\frac{1}{\alpha} \int_Y ((k_i - h_i) + (k - h)\nabla w_i^h) \cdot \nabla(w_i^k - w_i^h) \, dy \\ &\leq \frac{1}{\alpha} (\|k_i - h_i\|_{L^2(Y)} + 2\|k - h\|_{[L^\infty(Y)]^{2 \times 2}} \|\nabla w_i^h\|_{L^2(Y)}) \|\nabla(w_i^k - w_i^h)\|_{L^2(Y)} \\ &\leq \frac{1}{\alpha} (\sqrt{|Y|} + 2\|\nabla w_i^h\|_{L^2(Y)}) \|k - h\|_{[L^\infty(Y)]^{2 \times 2}} \|\nabla(w_i^k - w_i^h)\|_{L^2(Y)}. \end{aligned} \quad (3.49)$$

By Lax-Milgram's Lemma applied to (3.47) we get

$$\|\nabla w_i^h\|_{L^2(Y)} \leq \frac{\sqrt{|Y|}}{\alpha} \|h\|_{[L^\infty(Y)]^{2 \times 2}}. \quad (3.50)$$

Combining (3.49) and (3.50), we get

$$\|\nabla(w_i^k - w_i^h)\|_{L^2(Y)} \leq \left(\frac{\sqrt{|Y|}}{\alpha} + \frac{2\sqrt{|Y|}}{\alpha} \|h\|_{[L^\infty(Y)]^{2 \times 2}} \right) \|k - h\|_{[L^\infty(Y)]^{2 \times 2}}.$$

This inequality holds for each $i \in \{1, 2\}$. Adding them up yields

$$\|\nabla(w_1^k - w_1^h)\|_{L^2(Y)} + \|\nabla(w_2^k - w_2^h)\|_{L^2(Y)} \leq \frac{2\sqrt{|Y|}}{\alpha} (1 + 2\|h\|_{[L^\infty(Y)]^{2 \times 2}}) \|k - h\|_{[L^\infty(Y)]^{2 \times 2}},$$

which proves the Proposition for $\mathcal{H} = [L^2(Y)]^2$. By the equivalence of norms, the result holds also for the choice $\mathcal{H} = [H^1(Y)]^2$. \square

Proposition 3.6. *The mapping $k \rightarrow (w_1, w_2) = \mathcal{G}_{\text{cell}}(k)$ for $k \in \mathcal{U}$ is Fréchet differentiable, viewed as a mapping into $[L^2(Y)]^2$ or $[H^1(Y)]^2$, with directional derivative $\mathcal{G}'_{\text{cell}}(k)h = (m_1, m_2) \in [H_\diamond^1(Y; 0)]^2$ satisfying*

$$\int_Y (k\nabla m_i) \cdot \nabla v \, dy = \int_Y (h\nabla w_i) \nabla v + h_i \cdot \nabla v \, dy \quad \forall v \in H_\diamond^1(\Omega; 0). \quad (3.51)$$

Furthermore, the mapping $k \rightarrow (\nabla w_1, \nabla w_2) \in [L^2(Y)]^{2 \times 2}$ is Fréchet differentiable, with directional derivative in direction h given by ∇m .

Proof. Let $k, h \in \mathcal{U}$ and let $(w_1^h, w_2^h) = \mathcal{G}_{\text{cell}}(k + h)$, $(w_1^k, w_2^k) = \mathcal{G}_{\text{cell}}(k)$ and let $(m_1, m_2) \in [H_\diamond^1(Y; 0)]^2$ solve (3.51). We will show that $(m_1, m_2) = \mathcal{G}'_{\text{cell}}(k)h$. Let us first consider an arbitrary component $i \in \{1, 2\}$ and then extend the result from a single component to the product space. By using the coercivity of k followed by Cauchy-Schwarz' inequality, we have

$$\|\nabla(w_i^h - w_i^k - m_i)\|_{L^2(Y)} \leq \frac{1}{\alpha} \int_Y (k\nabla(w_i^h - w_i^k - m_i)) \cdot \nabla(w_i^h - w_i^k - m_i) \, dy. \quad (3.52)$$

Substituting $k + h$ for h in (3.48) and $w_i^h - w_i^k - m_i$ for v , we have

$$\int_Y (k \nabla(w_i^h - w_i^k)) \cdot \nabla(w_i^h - w_i^k - m_i) dy = \int_Y h_i \cdot \nabla(w_i^h - w_i^k - m_i) + (h \nabla w_i^h) \cdot \nabla(w_i^h - w_i^k - m_i) dy. \quad (3.53)$$

Inserting (3.53) on the right-hand side of (3.52) yields

$$\|\nabla(w_i^h - w_i^k - m_i)\|_{L^2(Y)}^2 \leq \frac{1}{\alpha} \int_Y (h_i + h \nabla w_i^h - k \nabla m_i) \cdot \nabla(w_i^h - w_i^k - m_i). \quad (3.54)$$

By substituting (3.51) with $v = w_i^h - w_i^k - m_i$ back into the right-hand side of (3.54), we have

$$\|\nabla(w_i^h - w_i^k - m_i)\|_{L^2(Y)}^2 \leq \frac{1}{\alpha} \int_Y (h \nabla(w_i^h - w_i^k)) \cdot \nabla(w_i^h - w_i^k - m_i) dy.$$

By using Cauchy-Schwarz' inequality and Proposition 3.5, we now have

$$\|\nabla(w_i^h - w_i^k - m_i)\|_{L^2(Y)} \leq \frac{2}{\alpha} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) \|h\|_{[L^\infty(Y)]^{2 \times 2}}. \quad (3.55)$$

Adding the result for both components i , we have

$$\sum_{i=1}^2 \|\nabla(w_i^h - w_i^k - m_i)\|_{L^2(Y)} \leq \frac{4}{\alpha} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) \|h\|_{[L^\infty(Y)]^{2 \times 2}}. \quad (3.56)$$

(3.56) proves that the Fréchet derivative of the map $k \rightarrow (\nabla w_1, \nabla w_2)$ is $(\nabla m_1, \nabla m_2)$. By the equivalence of norms in $H_\diamond^1(Y; 0)$, we have

$$\sum_{i=1}^2 \|w_i^h - w_i^k - m_i\|_{H^1(Y)} \leq \frac{4}{\alpha} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) \|h\|_{[L^\infty(Y)]^{2 \times 2}},$$

which shows that $\mathcal{G}'_{cell}(k)h = (m_1, m_2)$. □

Proposition 3.7. *The map $k \rightarrow D(k, \mathcal{G}_{cell}(k))$ for $k \in \mathcal{U}$ is Fréchet differentiable and the derivative is given by*

$$\frac{dD(k, \mathcal{G}_{cell}(k))}{dk} h = \int_Y k(y) \nabla M(y) + h(y) (I + \nabla W_k(y)) dy \quad (3.57)$$

where $\nabla M(y) = [\nabla m_1(y) \quad \nabla m_2(y)]$, $\nabla W_k(y) = [\nabla w_1^k(y) \quad \nabla w_2^k(y)]$ and $(m_1, m_2) = \mathcal{G}'_{cell}(k)h$ and $(w_1^k, w_2^k) = \mathcal{G}_{cell}(k)$.

Proof. We want to verify that

$$\lim_{\|h\| \rightarrow 0} \frac{\|D(k+h, \mathcal{G}_{cell}(k+h)) - D(k, \mathcal{G}_{cell}(k)) - \frac{dD(k, \mathcal{G}_{cell}(k))}{dk} h\|_{[L^\infty(\Omega)]^{2 \times 2}}}{\|h\|_{[L^\infty(Y)]^{2 \times 2}}} = 0.$$

Let $\nabla W_h = [\nabla w_1^h \ \nabla w_2^h]$ where $(w_1^h, w_2^h) = \mathcal{G}_{cell}(k+h)$. By using (3.57), we have

$$\begin{aligned}
& \left\| D(k+h, \mathcal{G}_{cell}(k+h)) - D(k, \mathcal{G}_{cell}(k)) - \frac{dD(k, \mathcal{G}_{cell}(k))}{dk} h \right\|_{[L^\infty(\Omega)]^{2 \times 2}} \\
&= \sup_{x \in \Omega} \left| \int_Y (k+h)(I + \nabla W_h) - k(I + \nabla W_k) - k \nabla M - h(I + \nabla W_k) \right| \\
&= \left| \int_Y (k+h)(I + \nabla W_h) - k(I + \nabla W_k) - k \nabla M - h(I + \nabla W_k) \right| \\
&= \left| \int_Y k(\nabla W_h - \nabla W_k - \nabla M) + h(\nabla W_h - \nabla W_k) \right|,
\end{aligned} \tag{3.58}$$

where the second equality in (3.58) is due to the fact that all functions in the integral are independent of $x \in \Omega$. Let us turn to studying the matrix by its components. Let k_i, h_i denote the i th row of k and h , respectively. Then we have on row i and column j of the matrix (3.58) the element

$$\int_Y k_i \cdot \nabla(w_j^h - w_j^k - m_j) + h_i \cdot \nabla(w_j^h - w_j^k) dy.$$

By using Cauchy-Schwarz' inequality followed by (3.55) and Proposition 3.5, we have

$$\begin{aligned}
& \left| \int_Y k_i \cdot \nabla(w_j^h - w_j^k - m_j) + h_i \cdot \nabla(w_j^h - w_j^k) dy \right| \\
&\leq \sqrt{|Y|} \|k\|_{[L^\infty(Y)]^{2 \times 2}} \|\nabla(w_j^h - w_j^k - m_j)\|_{L^2(Y)} + \sqrt{|Y|} \|h\|_{[L^\infty(Y)]^{2 \times 2}} \|\nabla(w_j^h - w_j^k)\|_{L^2(Y)} \\
&\leq \sqrt{|Y|} \|k\|_{[L^\infty(Y)]^{2 \times 2}} \frac{2}{\alpha} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) \|h\|_{[L^\infty(Y)]^{2 \times 2}}^2 + \sqrt{|Y|} \|h\|_{[L^\infty(Y)]^{2 \times 2}} \|\nabla(w_j^h - w_j^k)\|_{L^2(Y)} \\
&\leq \sqrt{|Y|} \|k\|_{[L^\infty(Y)]^{2 \times 2}} \frac{2}{\alpha} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) \|h\|_{[L^\infty(Y)]^{2 \times 2}}^2 + \sqrt{|Y|} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) \|h\|_{[L^\infty(Y)]^{2 \times 2}}^2 \\
&\leq \sqrt{|Y|} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) (\|k\|_{[L^\infty(Y)]^{2 \times 2}} \frac{2}{\alpha} + 1) \|h\|_{[L^\infty(Y)]^{2 \times 2}}^2
\end{aligned}$$

Combining the above, we now have

$$\begin{aligned}
& \lim_{\|h\| \rightarrow 0} \frac{\|D(k+h, \mathcal{G}_{cell}(k+h)) - D(k, \mathcal{G}_{cell}(k)) - \frac{dD(k, \mathcal{G}_{cell}(k))}{dk} h\|_{[L^\infty(\Omega)]^{2 \times 2}}}{\|h\|_{[L^\infty(Y)]^{2 \times 2}}} \\
&\leq \lim_{\|h\| \rightarrow 0} \frac{\sqrt{|Y|} (C_1 + C_2 \|k\|_{[L^\infty(Y)]^{2 \times 2}}) (\|k\|_{[L^\infty(Y)]^{2 \times 2}} \frac{2}{\alpha} + 1) \|h\|_{[L^\infty(Y)]^{2 \times 2}}^2}{\|h\|_{[L^\infty(Y)]^{2 \times 2}}} = 0.
\end{aligned}$$

□

Corollary 3.1. *The map $k \rightarrow \mathcal{G}_{eff}(D(k, \mathcal{G}_{cell}(k)))$ for $k \in \mathcal{U}$ is Fréchet differentiable. The directional derivative $u_h = \frac{d}{dk} [\mathcal{G}_{eff}(D(k, \mathcal{G}_{cell}(k)))] h$ can be computed by the following algorithm.*

1. Compute $w = (w_1, w_2) = \mathcal{G}_{cell}(k)$.
2. Compute $(m_1, m_2) = \mathcal{G}'_{cell}(k)h$.

3. Construct $D_k = D(k, w)$ and compute $u_k = \mathcal{G}_{eff}(D_k)$.
4. Construct $D_h = \frac{dD(k, \mathcal{G}_{cell}(k))}{dk} h$ from (w_1, w_2) and (m_1, m_2) via (3.57).
5. Compute $u_h = \mathcal{G}'_{eff}(D_k) D_h$.

4 The inverse problems

Now we turn the studying an inverse problem to the forward problem that we have studied so far. We have the equation

$$\begin{aligned} \operatorname{div}(-k\nabla u) &= 0 & \text{in } \Omega \\ -k\nabla u \cdot n &= g & \text{on } \partial\Omega \\ \int_{\Omega} u \, dx &= m \end{aligned} \tag{4.1}$$

where $\operatorname{supp}(g) \subseteq \Gamma^{in}$ and

$$k(x) = \begin{cases} k_0(x), & x \in \Omega_l \cup \Omega_r \\ k_1(x), & x \in \Omega_1. \end{cases} \tag{4.2}$$

Depending on whether the problem is posed before or after the two-scale limit and whether or not we consider both the upscaled equation and the cell-problem or only the upscaled equation, we get the following 3 different choices for k_1 .

- (P_ε) : $k_1(x) = k_\varepsilon(x) = k(\frac{x}{\varepsilon}), k \in [L^\infty_{\#}(Y)]^{2 \times 2}$,
- (P_{eff}) : $k_1(x) = k, k \in \mathbb{R}^{2 \times 2}$,
- $(P_{eff}) \& (P_{cell})$: $k_1(x) = \int_Y k(I + \nabla W) \, dy, w = \mathcal{G}_{cell}(k), k \in [L^\infty_{\#}(Y)]^{2 \times 2}$.

In the next sections, we will see that the inverse problem of (P_ε) has some serious problems and that some of those problems are resolved by the use of homogenization theory and that some problems are not yet resolved but that there is reason to be hopeful.

Solving the inverse problem can be formulated as solving the equation

$$y = \mathcal{F}(k)$$

where $\mathcal{F} = \mathcal{C} \circ \mathcal{G}$, \mathcal{G} is a parameter-to-solution operator such as \mathcal{G}_ε or \mathcal{G}_{eff} , and \mathcal{C} is an observation operator which represents a physical measurement of the quantity $u = \mathcal{G}(k)$, and y represents a known observation. That is, we wish to find the parameter k that corresponds to the real-world observation y . This raises the question of well-posedness of this equation. Of particular interest is the uniqueness of the solution k , which is often approached by studying the injectivity of the operator \mathcal{F} , and the continuous dependence of the solution k on the observation y . Existence of solution is not that common to study, since that is more a question of how well the operator \mathcal{F} models reality.

Often it turns out that this equation is not well-posed. In particular continuity of \mathcal{F} is often violated in inverse problems, but we will see that even uniqueness is violated in our case. One approach to resolving ill-posedness is to not solve the equation directly, but instead solve it by a least-squares approach. For a so-called regularization parameter $\alpha > 0$ we solve

$$\begin{aligned} T_\alpha(k; y) &= \frac{1}{2} \|\mathcal{F}(k) - y\|_y^2 + \alpha J(k), \\ \mathcal{R}_\alpha(y) &= \operatorname{argmin}_k \{T_\alpha(k; y)\}, \end{aligned} \tag{4.3}$$

where the term $\alpha J(k)$ is added to resolve the ill-posedness. Precisely which norm $\|\cdot\|_y$ is chosen to be depends on the application, but we will pick an L^2 -norm on some domain. J can be a quite general functional, but we will consider L^2 -norms or Sobolev space norms or a total variation functional. In the case that \mathcal{F} is a linear operator, it can be shown under quite generous assumptions that the inverse \mathcal{R}_α^{-1} is continuous with respect to y for each $\alpha > 0$; see for example [25]. Unfortunately, this continuity does not translate well into our setting since, as mentioned briefly above, \mathcal{F} is defined from parameter-to-solution maps such as \mathcal{G}_{eff} which are nonlinear. If continuous dependence is of great importance, it is possible to linearize the problem using the Fréchet derivatives derived in Section 3.4. We will, however, instead investigate ways to justify solving the full nonlinear problems in a well-posed way.

4.1 The Neumann-to-Dirichlet and Dirichlet-to-Neumann maps

Modelling of the physical measurement is a crucial part of any inverse problem. In our situation we consider the measurements to be pressure or density measurements at the boundary $\Gamma_M \subseteq \partial\Omega_r \cap \partial\Omega$ of the gas whose density is given by u , the solution of the partial differential equation.

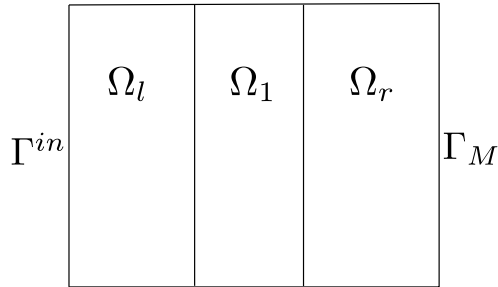


Figure 4.1: An example showing, in particular, the placement of Γ^{in} , Γ_M and Ω_1 .

This measurement is natural to model as pointwise measurements at a finite number of points along Γ_M . However, by instead modelling the measurement as the trace of the solution u onto Γ_M , we get access to mathematical tools that are much more well-suited for a partial differential equation setting than those of pointwise evaluation. Hence, we will use trace-based measurement modelling. More specifically, we use the so-called Neumann-to-Dirichlet operator, which maps the right-hand side Neumann data g to the trace $u|_{\partial\Omega}$ of the solution u onto the boundary. In the literature, the related Dirichlet-to-Neumann map is often studied instead, which maps $u|_{\partial\Omega}$ to g . While these may at first sight appear to be the inverse of each

other, there are some subtle details that prevent this from being the case. With that said, it is important to properly define the Neumann-to-Dirichlet map. But since a lot of the related research is done for the Dirichlet-to-Neumann map we need it as well.

Let $k \in [L^\infty(\Omega)]^{2 \times 2}$ be symmetric and such that there exists an $\alpha > 0$ such that $y^T k(x) y \geq \alpha |y|^2$, and consider the equation

$$\operatorname{div}(-k(x)\nabla u(x)) = 0 \quad x \in \Omega. \quad (4.4)$$

We define the Neumann-to-Dirichlet map $\mathcal{N}_k: H_\diamond^{-1/2}(\partial\Omega) \rightarrow H^{1/2}(\partial\Omega)$ by $\mathcal{N}_k(g) = u|_{\partial\Omega}$, where $u \in H_\diamond^1(\Omega; 0)$ is the unique weak solution to (4.4) when supplied with the Neumann boundary condition $(k\nabla u) \cdot n = g$ on $\partial\Omega$. Then we can define \mathcal{N}_k as the self-adjoint operator on $H_\diamond^{-1/2}(\partial\Omega)$ satisfying

$$\langle g, \mathcal{N}_k(f) \rangle = \int_\Omega (k(x)\nabla u(x)) \cdot \nabla v(x) dx \quad (4.5)$$

where $u, v \in H_\diamond^1(\Omega; 0)$ are the unique weak solutions with Neumann data g and f , respectively.

Similarly, we define the Dirichlet-to-Neumann map $\Lambda_k: H^{1/2}(\partial\Omega) \rightarrow H^{-1/2}(\partial\Omega)$ to be $\Lambda_k(f) = (k\nabla u) \cdot n|_{\partial\Omega}$, or by the weak formulation

$$\langle \Lambda_k(f), g \rangle = \int_\Omega (k(x)\nabla u(x)) \cdot \nabla v(x) dx \quad (4.6)$$

where $u, v \in H^1(\Omega)$ are the unique weak solutions with Dirichlet data f and g , respectively.

We note that \mathcal{N}_k is usually defined to map into the space $\{u \in H^{1/2}(\partial\Omega): \int_{\partial\Omega} u d\sigma = 0\}$. As defined above, \mathcal{N}_k instead maps into the space $\{u \in H^{1/2}(\partial\Omega): \exists v \in H^1(\Omega), v|_{\partial\Omega} = u, \int_\Omega v dx = 0\}$. Moving the integral constraint from the boundary $\partial\Omega$ to the interior Ω in this way likely has little impact on the related theory, as the integral constraint appears to mainly serve as a way to obtain a well-defined operator. Then it makes sense to use the integral constraint over all of Ω since that constraint is required to hold by the forward problem.

4.2 Formulation of the inverse problems

We can express the inverse problems as follows. Suppose that we know the particle flux through the boundary of a domain and that we know the particle density at a subset of the boundary. Determine the diffusion coefficient throughout the domain.

With reference to (4.1) and (4.2), the particle flux g is considered to be a parameter that we can modify as part of the design of the experiment. The diffusion coefficient k_0 in the subdomains Ω_l, Ω_r is set to a constant matrix, $k_0(x) = k_0$, and it is not a parameter that we modify between experiments. The particle density is $u|_{\Gamma_M}$. This is a quantity that has to be measured for each prescribed flux g . We wish to determine the diffusion coefficient k_1 in the subdomain Ω_1 . Let us from now on denote by \mathcal{N}_h the Neumann-to-Dirichlet map \mathcal{N}_k for the coefficient k defined in (4.2) with $k_1 = h$. For N different experiments with different fluxes g_1, \dots, g_N the data for the inversion is contained in the set of Neumann-Dirichlet pairs $\{(g_i, \mathcal{N}_{k_1}(g_i))\}_{i=1}^N$. When studying whether or not the data is sufficient for determining the

coefficient, it is usually assumed that we have all the data $\{(g, \mathcal{N}_{k_1}(g))\}_{g \in H^{-1/2}(\Omega)}$ for every possible particle flux g , but this is of course not possible in practice. Therefore, we choose the realistic finite data model here. We can now define 3 inverse problems, corresponding to the 3 choices of k_1 described in the introduction to this chapter.

Definition 4.1 (ε -dependent inverse problem). Let $k_\varepsilon(x)$ be an unknown function, assumed to be of the form $k_\varepsilon(x) = \bar{k}(\frac{x}{\varepsilon})$ where $\bar{k} \in [L^\infty(Y)]^{2 \times 2}$. Let $\{(g_i, \mathcal{N}_{k_\varepsilon}(g_i)|_{\Gamma_M})\}_{i=1}^N$ be known data from N experiments with particle fluxes $g_i \in H^{-1/2}(\partial\Omega)$ with $\text{supp}(g_i) \in \Gamma^{in}$. Determine the diffusion coefficient k_ε .

Recall that from the limits in (3.15) we have $u_\varepsilon \rightarrow u_0$ in $L^2(\partial\Omega)$, as $\varepsilon \rightarrow 0$, where u_ε is the weak solution in the sense of Definition 3.2 and u_0 be a weak solution in the sense of Definition 3.4. Now let $\mathcal{N}_D^{eff}(g)$ be the Neumann-to-Dirichlet operator for the upscaled equation, that is, if D is an effective diffusion coefficient then $\mathcal{N}_D^{eff}(g) = u|_{\partial\Omega}$, where u is the weak solution in the sense of Definition 3.4. Let $\eta_\varepsilon = (u_\varepsilon - u)|_{\Gamma_M} := (\mathcal{N}_{k_\varepsilon}(g) - \mathcal{N}_D^{eff}(g))|_{\Gamma_M}$ be the approximation error. Then

$$\mathcal{N}_{k_\varepsilon}(g)|_{\Gamma_M} = \mathcal{N}_D^{eff}(g)|_{\Gamma_M} + \eta_\varepsilon, \quad (4.7)$$

where $\|\eta_\varepsilon\|_{L^2(\Gamma_M)} \rightarrow 0$ as $\varepsilon \rightarrow 0$. That is, if we try to find an effective diffusion coefficient D that explains the observations $\mathcal{N}_{k_\varepsilon}(g)$, then we only introduce the small error η_ε whose L^2 -norm tends to 0. Based on this observation we can formulate a new inverse problem, which aims at reconstructing the constant effective diffusion coefficient rather than the rapidly oscillating diffusion coefficient.

Definition 4.2 (Upscaled inverse problem). Let $D \in \mathbb{R}^{2 \times 2}$ be an unknown effective diffusion coefficient and let $\{(g_i, \mathcal{N}_D^{eff}(g_i)|_{\Gamma_M})\}_{i=1}^N$ be known data from N experiments with particle fluxes $g_i \in H^{-1/2}(\partial\Omega)$ with $\text{supp}(g_i) \in \Gamma^{in}$. Determine the effective diffusion coefficient D .

One last way in which we can formulate the inverse problem is to still use the effective coefficient, but now use its parametrization on the cell-problems. The same observation as above show that this only introduces a small error into the inversion.

Definition 4.3 (Multiscale inverse problem). Let $k \in [L^\infty(Y)]^{2 \times 2}$ be an unknown coefficient for the cell-problems and let $D(k)$ be the corresponding effective coefficient. Let $\{(g_i, \mathcal{N}_{D(k)}^{eff}(g_i)|_{\Gamma_M})\}_{i=1}^N$ be known data from N experiments with particle fluxes $g_i \in H^{-1/2}(\partial\Omega)$ with $\text{supp}(g_i) \in \Gamma^{in}$. Determine the effective diffusion coefficient $D(k)$ and/or the cell-problem coefficient k .

4.3 Ill-posedness of the ε -dependent inverse problem

Our inverse problems are identical in their mathematical structure to the famous Calderón problem or the inverse conductivity problem in electrical impedance tomography[11]. These two problems have been heavily studied in the last four decades and we can use a lot of this research to conclude that the present ε -dependent problem is ill-posed. There is one caveat, however, namely that most of this research considers the Dirichlet-to-Neumann map rather

than the Neumann-to-Dirichlet map. But by the strong similarity between the two maps, it seems quite likely that ill-posedness results for one translates into similar results for the other.

Let us start with the isotropic case, that is, the case where the diffusion coefficient k_ε is a real-valued function rather than a matrix-valued function, or if we wish to interpret it as matrix valued then it is of the form $f(x)I$, where f is real-valued and I is the identity matrix. The ε -dependent problem may in some cases correspond to this isotropic case. The uniqueness problem has been solved in a very general setting in [7], requiring no regularity on the boundary $\partial\Omega$ and allowing to identify coefficients k in the subset of functions in $L^\infty(\Omega)$ that satisfy $c^{-1} \leq k \leq c$ for some fixed $c > 0$. Note that we identify the entire coefficient k , and not only $k_1 = k_\varepsilon$ on Ω_1 . But since the identification is unique, the identified coefficient agrees with the known coefficient k_0 on Ω_l and Ω_r . This means that, at least in principle, it is possible to identify the diffusion coefficient k_ε in (3.1) from observations of $(k_\varepsilon \nabla u_\varepsilon) \cdot n$ on the boundary $\partial\Omega$.

A serious issue is that of stability of the solution of the inverse problem. It is shown with a counterexample in chapter 5 of [18] that even a piecewise constant diffusion coefficient cannot be reconstructed in a continuous manner. More specifically, the example presents a sequence of piecewise constant coefficients k_r depending on a parameter $r > 0$ and a fixed coefficient k_0 such that $\|k_0 - k_r\|_{L^\infty(\Omega)}$ is constant while $\|\mathcal{N}_{k_0} - \mathcal{N}_{k_r}\| \rightarrow 0$ as $r \rightarrow 0$. This disproves the continuity of the inverse of $k \rightarrow \mathcal{N}_k$. This is a well-known example in the inverse problems community and the original source seems to be [3]³. Furthermore, [3] is often credited with obtaining a logarithmic stability for the inverse problem, after restricting attention to coefficients in the subset $H^{s+2}(\Omega)$ of $L^\infty(\Omega)$ for $s > n/2$ and $n \geq 3$. Specifically, the stability supposedly is $\|k_1 - k_2\|_{L^\infty(\Omega)} \leq w(\|\mathcal{N}_{k_1} - \mathcal{N}_{k_2}\|)$, where w is a function satisfying $w(t) \leq C|\ln t|^{-\alpha}$. For $n = 2$, a similar logarithmic bound is obtained in [9]. So, it appears that by requiring the coefficients to be differentiable, we can obtain some form of stability, albeit a weak logarithmic one. It is shown in [23] that such a logarithmic stability is actually the strongest stability that can be obtained if all we do in order to obtain stronger stability is to require the coefficients to be differentiable up to some order. This is bad news for the ε -dependent inverse problem, because a small improvement in $\|k_1 - k_2\|_{L^\infty(\Omega)}$ requires a much larger improvement in $\|\mathcal{N}_{k_1} - \mathcal{N}_{k_2}\|$. In practice, this means that we need incredibly accurate measurements to capture the oscillations in k_ε . But due to the oscillations, such accurate measurements might not be possible. Therefore, reconstructing the true oscillating coefficient k_ε may not be practically possible.

Since the isotropic case can be considered as a special case of the anisotropic case, it seems unlikely that the anisotropic case would work out any better. Actually, it turns out that the anisotropic case is even worse in regards to uniqueness. An anisotropic diffusion coefficients cannot be determined uniquely in the same way that isotropic coefficients can. It was first observed in [19] that a change of variables $y = F(x)$ in (4.1) by a diffeomorphism $F: \Omega \rightarrow \Omega$ such that $F|_{\partial\Omega}$ is the identity forms a new partial differential equation of the same linear elliptic structure and with a different diffusion coefficient but with the exact same

³But I have been unable to get access to this article and therefore cannot verify the source. I suspect that the original counterexample presents it for the Dirichlet-to-Neumann map. So the fact that it is given for the Neumann-to-Dirichlet map in [18] is good.

Dirichlet-to-Neumann map. In [8] the converse to this statement was proved, namely that if any two diffusion coefficients correspond to the same boundary measurements then those two coefficients are related by such a diffeomorphism. That is, in the very general setting where we seek to identify a symmetric diffusion coefficient in a bounded subset of $[L^\infty(\Omega)]^{2 \times 2}$, the best identification we have is that of an equivalence class of coefficients related in a certain way by diffeomorphisms.

In conclusion, if we wish to identify an isotropic coefficient k_ε then it is possible to uniquely determine the coefficient, but it depends sensitively on the observations. If, instead, we wish to identify an anisotropic coefficient k_ε then we have non-uniqueness in addition to the sensitivity on observations.

4.4 Towards well-posedness of the upscaled inverse problem

Despite the issues of non-uniqueness for anisotropic case mentioned in Section 4.3, there is one thing that provides a natural solution, namely, the homogenization theory. While one line of research for the Calderón problem has been concerned with the problem of determining uniqueness of the coefficient up to diffeomorphisms, another line of research has attempted to actually obtain uniqueness by restricting attention to certain subsets of coefficients. There is a very recent result that obtains uniqueness of piecewise constant anisotropic coefficients of the form

$$k(x) = \sum_{i=1}^l k_i \chi_{D_i}(x), \quad (4.8)$$

where k_i are symmetric matrices and D_i are fixed and nonoverlapping known open subsets of Ω such that $\bar{\Omega} = \bigcup_{i=1}^l \bar{D}_i$, along with a few more technical requirements. Then it is shown in [4] that such a piecewise constant diffusion coefficient can be uniquely identified from measurements made on a subset $\Gamma \subseteq \partial\Omega$. This last point is important, since in our setting we do measure on a subset of the boundary. However, in our setting the measurement is done on a subset Γ_M which is disjoint from the subset Γ^{in} where the nonzero part of Neumann boundary is prescribed. This does not quite agree with the result in [4], which requires $\Gamma_M = \Gamma^{in}$. But the result in [4] is still a big step towards well-posedness. It also shows that homogenization theory may play a crucial role in obtaining a well-posed parameter identification of diffusion coefficients from boundary measurements, since it is homogenization theory that provide us with the piecewise constant structure of the coefficient. But some work still needs to be done to obtain a partial data result for disjoint sets Γ_M and Γ^{in} . However, we should note that the case with disjoint Γ_M and Γ^{in} is still an open problem even for the isotropic case, according to the relatively recent survey [17] on uniqueness with partial data.

The next step towards well-posedness is stability with respect to the observations. Also here has progress been made, but the current results does not quite agree with our setup. There are two results in particular that are close to solving the stability problem for our problem with a piecewise constant coefficient. First, in [6], it is shown for the isotropic situation that if the coefficient is of the form (4.8) with $k_i \in \mathbb{R}$ there holds a Lipschitz bound $\|k_1 - k_2\|_{L^\infty(\Omega)} \leq C \|\Lambda_{k_1} - \Lambda_{k_2}\|$. Furthermore, this result holds for the partial data situation where $\Gamma_M = \Gamma^{in}$. Note, however, that this is a bound for the Dirichlet-to-Neumann map. In

[15] it is shown that if the coefficient is of the form

$$k(x) = \gamma(x)A(x) \tag{4.9}$$

where $A(x)$ is a Lipschitz continuous 2×2 matrix-valued function and γ is an isotropic coefficient of the form (4.8) then we again have Lipschitz stability, and with the partial data situation corresponding to $\Gamma_M = \Gamma^{in}$. However, instead of (4.9) we would need something like

$$k(x) = \sum_{i=1}^l \gamma_i(x)A_i(x). \tag{4.10}$$

and then we could pick each $A_i(x)$ to be a symmetric constant matrix. The proofs presented in [4] and [15] are very similar, despite one being isotropic and one anisotropic. It appears that the main part of the proof presented in [15] may also apply to (4.10), assuming that a few technical lemmas generalize in an expected way. These technical lemmas are, however, very technical and it is beyond the scope of this thesis to prove them. But at the very least, it seems likely that a diffusion coefficient of the form (4.10) enjoys the same Lipschitz stability with partial data as was derived in [4] and [15].

4.5 Towards well-posedness by using multiscale modelling

Rather than using homogenization theory as the main tool to obtaining well-posedness, there is a result on the Calderón problem that may allow multiscale modelling to be used to obtain well-posedness. This result can then be applied to both the ε -dependent inverse problem in the sense of Definition 4.1 or the multiscale inverse problem in the sense of Definition 4.3. However, it is likely computationally advantageous to use the multiscale inverse problem, since the ε -dependent inverse problem most likely requires a much finer mesh to be used in the discretization.

In [5], is presented a uniqueness result for the identification of diffusion coefficients with a particular parametrization of the form $A(x) = A(x, a(x))$, where A is a matrix-valued function in $W^{1,p}$ and a is a real-valued function in $W^{1,p}$. Different results are obtained for finite p and for infinite p . If $p = \infty$ then it is proved that the full coefficient $A(\cdot, a(\cdot))$ can be recovered from the Dirichlet-to-Neumann map. For finite p then it is shown that $A(\cdot, a(\cdot))|_{\partial\Omega}$ can be identified uniquely from the Dirichlet-to-Neumann map. The identification on the boundary $\partial\Omega$ is shown to be Lipschitz and even Hölder stable.

It is tempting to use this result on the effective coefficient by taking $A = D$ and $a = k$ where $D(x, k)$ is defined as in (3.23). Then k is a function on $\Omega \times Y$ but a should only be a function on Ω and therefore this does not work. It is also required that a is real-valued, but k may be matrix-valued.

Instead of applying the parametrization result directly to the effective coefficient $D(x, k)$, it could be applied to the microstructure given by k . Say, for example, that it is known that the microstructure has a constant base value k_b and that the microstructure deviates from k_b according to some function $\bar{k}(y)$, but that it is unknown how large effect the microstructure \bar{k} has. That is, the amplitude of \bar{k} is unknown. If we let $h(x)$ be the amplitude of the

microstructure at the point $x \in \Omega$ then we could assume that k is of the form

$$k(h(x), y) = k_b + h(x)\bar{k}(y). \quad (4.11)$$

Then we could consider the effective coefficient as being parametrized by h by inserting (4.11) into (3.23) and denote the resulting coefficient by $D(x, h(x))$. After verifying that several inequalities involving D and h hold, we could apply the result in [5] to $D(x, h(x))$. This type of parametrization of the microstructure is studied in [14]. The overall framework is somewhat different in [14] than in this thesis, however, as the authors of [14] appears to study the ε -dependent inverse problem rather than the homogenized one. In [14], they use a parametrization such as (4.11), among several other parametrizations that correspond to various microstructures.

This approach has several drawbacks when considering its application to the present problem. Two major issues with the result in [5] are that it uses the Dirichlet-to-Neumann map and that it is not a partial data result. Then the Neumann and Dirichlet data has to be known on the entire boundary $\partial\Omega$. Using a parametrization such as (4.11) also requires quite a lot of prior knowledge of the microstructure due to the assumed knowledge of \bar{k} , which we may not have in our application. Furthermore, for unique identification of $A(\cdot, a(\cdot))$ inside all of Ω , the result in [5] requires the function a to be Lipschitz continuous. But in the present problem, it is the most natural to consider a to be discontinuous. We could for example have the parametrization

$$k(x, y) = h(x)k_0(x) + (1 - h(x))k_1(y). \quad (4.12)$$

Taking $h = \chi_{\Omega_0}$ in (4.12), where χ_{Ω_0} is the characteristic function for the set Ω_0 , and inserting (4.12) into (3.23) gives exactly the effective coefficient in (3.12). So a parametrization such as (4.12) describes our problem, if the parameter h is allowed to be discontinuous. But in order to use the result in [5] for identification inside all of Ω , h has to take the place of a and h must therefore be Lipschitz continuous. Another issue is that a in the results in [5] is not allowed to attain the value 0 but in order for (4.12) to properly describe our problem, h needs to attain the value 0 in Ω_0 . So while (4.12) can potentially describe our problem, there are some issues that prevent h from taking the place of a in [5].

It seems that the results in [5] can be used to make progress towards well-posedness in a multiscale setting, as demonstrated in [14], but the success of such an approach seems to rely on the microstructure changing continuously throughout Ω . Since this is not the case for the inverse problem considered in this thesis, the result in [5] cannot be applied.

4.6 An optimization approach to inversion

In this section we consider the actual inversion process, which we do by solving an optimization problem of the form (4.3). Since the homogenized inverse problem of reconstructing the piecewise constant effective coefficient is the closest to being proved well-posed, that is the problem we focus on now. We will prove the existence of minimizers, which is enabled by the fact that the homogenized problem is finite dimensional. We also briefly discuss why this argument does not apply to the ε -dependent problem, which demonstrates that homogenization theory plays a crucial role also in the optimization-based inversion process.

Recall that we consider the observation $y = \mathcal{F}(k_\varepsilon)$ to be an observation of the ε -dependent forward problem and that inverting with respect to the homogenized coefficient introduces the error $\eta_\varepsilon = (u_\varepsilon - u_{eff})|_{\Gamma_M}$. Then the equation we wish to solve is $y = \mathcal{F}(k) + \eta_\varepsilon$, for an effective coefficient k . Since we now have an error term, we cannot conclude from the uniqueness of the homogenized inverse problem that the equation has a unique solution, or even a solution at all. Therefore, instead of solving the equation directly, we take the least-squares optimization approach. But it is not certain that the optimization problem is well-posed. In this section we will give partial solution this well-posedness question by proving existence of minimizer. Actually, even if we solved the ε -dependent inverse problem by the optimization approach a solution may not exist, since the data is generally assumed to be affected by measurement error and therefore we would still have an error term even for the ε -dependent problem.

4.6.1 The upscaled problem - existence of minimizer

For the proof, it is important that we optimize in reflexive Banach spaces, because we want to obtain weakly convergent subsequences from bounded sequences. The solution u of the partial differential equation belongs to a Hilbert space and therefore belongs to a reflexive space. The coefficient k , however, belongs to L^∞ which is not reflexive. But this issue is resolved by the fact that the effective coefficient belongs to the finite dimensional subspace

$$\mathcal{V} = \{k \in [L^\infty(\Omega)]^{2 \times 2} : k_{ij}|_{\Omega_l}, k_{ij}|_{\Omega_1}, k_{ij}|_{\Omega_r} \text{ are constant}\}$$

and finite dimensional spaces are reflexive. Furthermore, since the functions are constant in each of the 3 subdomains, this space \mathcal{V} contains precisely the same functions as the space

$$\{k \in [L^2(\Omega)]^{2 \times 2} : k_{ij}|_{\Omega_l}, k_{ij}|_{\Omega_1}, k_{ij}|_{\Omega_r} \text{ are constant}\}.$$

Due to the finite dimensions, the norm induced by the L^2 -inner product is equivalent to the L^∞ -norm. Therefore, we will instead work with the space \mathcal{V} as an inner product space, with the L^2 -inner product and its induced norm. This is simply for the convenience of having access to an inner product, which becomes even more convenient when we discuss the algorithmic side of actually solving the optimization problem.

For ease of notation, define for some given $\alpha_0 > 0$ the following spaces and subsets

$$\begin{aligned} \mathcal{V} &= \{k \in [L^2(\Omega)]^{2 \times 2} : k_{ij}|_{\Omega_0}, k_{ij}|_{\Omega_1} \text{ are constant}\}, & \mathcal{U} &= H^1(\Omega) \\ \mathcal{V}_{ad} &= \{k \in \mathcal{V} : y^T k(x) y \geq \alpha_0 |y|^2 \text{ for all } y \in \mathbb{R}^2, \text{ a.e. } x \in \Omega\}, & \mathcal{U}_{ad} &= H_\diamond^1(\Omega; m). \end{aligned}$$

The sets $\mathcal{U}_{ad}, \mathcal{V}_{ad}$ are used as constraints in the optimization problem, while \mathcal{U}, \mathcal{V} are the Hilbert spaces in which the optimization problem is formulated. For N different fluxes g_i , the partial differential equation constraint is defined by the mapping $e_i: \mathcal{U} \times \mathcal{V} \rightarrow \mathcal{U}^*$, which is given by

$$e_i(u, k)v = \int_{\Omega} (k \nabla u) \nabla v \, dx + \int_{\Gamma_{in}} g_i v \, d\sigma, \quad i \in \{1, \dots, N\}. \quad (4.13)$$

Now, the constraints are given as equations in \mathcal{U}^* as $e_i(u_i, k) = 0$, where 0 is interpreted as the functional $v \rightarrow 0$ for all $v \in \mathcal{U}$. Each flux g_i corresponds to one measurement, where the

measured data is $u_{M,i}$. Finally, for some given $k_0 \in [L^2(\Omega_0)]^{2 \times 2}$ with $\text{supp}(k_{0,ij}) \subseteq \Omega_0$ define the functional

$$T_\alpha(u, k; u_M, k_0) = \frac{1}{2} \sum_{i=1}^N \|u_i - u_{M,i}\|_{L^2(\Gamma_M)}^2 + \frac{\alpha}{2} \|k - k_0\|_{[L^2(\Omega)]^{2 \times 2}}^2,$$

where $u = (u_1, \dots, u_N)$, $u_M = (u_{M,1}, \dots, u_{M,N})$. The function k_0 is the diffusion coefficient in the subdomain Ω_0 , which is assumed to be known. Adding k_0 to T_α in this way can be thought of as adding our prior knowledge of the diffusion coefficient into the functional. With these definitions, we consider the optimization problem

$$\begin{aligned} \min_{u_i \in \mathcal{U}, k \in \mathcal{V}} \quad & T_\alpha(u, k; u_M, k_0) \\ \text{subject to} \quad & e_i(u_i, k) = 0 \quad i \in \{1, \dots, N\} \\ & u_i \in \mathcal{U}_{ad}, k \in \mathcal{V}_{ad}. \end{aligned} \tag{4.14}$$

Proposition 4.1. *The optimization problem (4.14) has a minimizer for each $\alpha > 0$.*

Proof. First note that for each $k \in \mathcal{V}_{ad}$ and each $i \in \{1, \dots, N\}$ there exists a unique $u_i \in \mathcal{U}_{ad}$ such that $e_i(u_i, k) = 0$. Therefore, the set \mathcal{W}_{ad} of feasible points of (4.14) is nonempty. Since $T_\alpha \geq 0$, the infimum $T^* = \inf_{(u,k) \in \mathcal{W}_{ad}} T_\alpha(u, k; u_M, k_0)$ exists and is finite. Then there exists a minimizing sequence, that is, a sequence $(u_n, k_n) \in \mathcal{W}_{ad}$ such that

$$\lim_{n \rightarrow \infty} T_\alpha(u_n, k_n; u_M, k_0) = T^*.$$

We will show that this limit is attained in the set \mathcal{W}_{ad} , which shows existence of a minimizer. The sequence $T_\alpha(u_n, k_n; u_M, k_0)$ is bounded and hence also $\|k_n\|_{[L^2(\Omega)]^{2 \times 2}}$ is bounded, as seen from

$$\begin{aligned} \|k_n\|_{[L^2(\Omega)]^{2 \times 2}}^2 &\leq 2(\|k_n - k_0\|_{[L^2(\Omega)]^{2 \times 2}}^2 + \|k_0\|_{[L^2(\Omega)]^{2 \times 2}}^2) \\ &\leq \frac{4}{\alpha} T_\alpha(u_n, k_n; u_M, k_0) + 2\|k_0\|_{[L^2(\Omega)]^{2 \times 2}}^2. \end{aligned}$$

Then k_n has a weakly convergent subsequence, also denoted by k_n , in $[L^2(\Omega)]^{2 \times 2}$, which is also a strongly convergent subsequence due to the finite dimension of \mathcal{V} . Since \mathcal{V}_{ad} is a closed subset of \mathcal{V} and $k_n \in \mathcal{V}_{ad}$ then $k_n \rightarrow k^* \in \mathcal{V}_{ad}$. There exists a unique $u_i^* \in \mathcal{U}_{ad}$ such that $e_i(u_i^*, k^*) = 0$. By the Lipschitz continuity of the forward map and by equivalence of norms in finite dimensional spaces,

$$\|u_{i,n} - u_i^*\|_{H^1(\Omega)} \leq C \|k_n - k^*\|_{[L^\infty(\Omega)]^{2 \times 2}} \leq C \|k_n - k^*\|_{[L^2(\Omega)]^{2 \times 2}},$$

and hence $u_{i,n} \rightarrow u_i^*$ strongly in \mathcal{U} . Now we have

$$T^* = \lim_{n \rightarrow \infty} T_\alpha(u_n, k_n; u_M, k_0) = T_\alpha(u^*, k^*; u_M, k_0) \geq T^*,$$

where the inequality follows from the fact that T^* is defined as an infimum over \mathcal{W}_{ad} and $(u^*, k^*) \in \mathcal{W}_{ad}$. We conclude that $T_\alpha(u^*, k^*; u_M, k_0) = T^*$ so that (u^*, k^*) indeed is a minimizer. \square

If we tried to apply the same argument to an optimization problem for the ε -dependent inverse problem with $\mathcal{V} = [L^\infty(\Omega)]^{2 \times 2}$, then we encounter some issues. The first issue is the \mathcal{V} is not a reflexive space, but that could still be resolved by taking instead $V = [H^s(\Omega)]^{2 \times 2}$ for $s > n/2$, because then H^s is embedded in L^∞ . Choosing a smaller space with nicer properties such as H^s is essentially a result of using regularization, and is therefore a quite natural thing to do in inverse problems. Next, we would still obtain a weakly convergent subsequence of the minimizing sequence k_n , and we would obtain weakly convergent subsequences $u_{i,n}$ from the Lipschitz continuity of the forward map. The issue is then to prove that the weak limit belongs to the feasibility set. This follows automatically in the finite dimensional case due to the strong convergence. When we only obtain weak convergence of $k_n, u_{i,n}$, we instead have to prove that the map e_i is closed under weak convergence.⁴

In addition to the existence result just presented, it would also be desirable to obtain uniqueness and continuity of the minimizer on the data u_M . If the constraints $e_i(u, k)$ were linear, then the convexity of T_α in u and strict convexity in k could be used to obtain uniqueness. But since $e_i(u, k)$ is nonlinear, such an argument for uniqueness fails. Regarding the continuity of minimizers on the data u_M , the nonlinearity of $e_i(u, k)$ is again problematic. We would want to obtain that if u_M^n is any sequence of observations that converge to u_M^0 then the minimizers u_n, k_n that correspond to u_M^n converge to the minimizers u_0, k_0 that correspond to u_M^0 . As shown in [16] this can at least be obtained for a subsequence and for weak convergence of u_n, k_n to u^*, k^* respectively, where u^*, k^* are minimizers not necessarily equal to u_0, k_0 . In our finite dimensional setting, the weak convergence of the subsequence turns into strong convergence. But we still cannot ensure that the limit u^*, k^* is u_0, k_0 .

4.6.2 Two possible approaches to the multiscale problem

While we will not put much effort into studying the possibility of reconstructing the coefficient of the cell-problems, let us at least mention two ways in which it could possibly be done. Both approaches are based on the parametrization of the effective coefficient on the cell-problems given by (3.12). The first idea is to first find an effective diffusion coefficient D by solving the optimization problem (4.14). Then we can try to find a corresponding microstructure by minimizing $\|D - D(k)\|$ in a suitable norm, and with the cell-problems as constraints. Then we obtain the following optimization problem

$$\begin{aligned} \min_{k, M} \quad & \frac{1}{2} \|D - M\|_{[L^2(\Omega)]^{2 \times 2}}^2 + \frac{\alpha}{2} \|k\|_{[H^s(Y)]^{2 \times 2}}^2 \\ \text{subject to} \quad & f_i(k, w_i) = 0 \quad i \in \{1, 2\} \\ & M = \int_Y k \left(I + \begin{bmatrix} \nabla w_1 & \nabla w_2 \end{bmatrix} \right) dy \end{aligned} \tag{4.15}$$

where f_i is defined by

$$f_i(k, w_i)v = \int_Y (k \nabla w_i) \cdot \nabla v \, dy - \int_Y k_i \cdot \nabla v \, dy \quad i \in \{1, 2\}$$

⁴This is something I have not managed to do, due to the map $(k_n, u_{i,n}) \rightarrow e_i(k_n, u_{i,n})$ being nonlinear. So we see that also in the optimization approach to the inversion, homogenization theory helps in getting closer to well-posedness.

and k_i is the i th column of k . The H^s norm on k in the objective function is taken because k should belong to a reflexive space. Then taking $s > n/2$ means that k is in a reflexive space which is continuously embedded into L^∞ . Note that unless k belongs to a finite dimensional subspace, then we will encounter the same issues regarding the existence of minimizers for (4.15) as was discussed for the ε -dependent problem at the end of Section 4.6.1. This issue could be resolved by linearization with respect to k , which we can do since we have obtained explicit forms of the Fréchet derivatives in Section 3.4.2. Another way to resolve the issue would be to find reasonable finite dimensional spaces for k , which is studied in [14].

Instead of first finding the effective coefficient and then trying to match a microstructure to it, we can consider doing both at the same time. Then we have the optimization problem

$$\begin{aligned}
& \min_{u_i, M_i, w_{i,j}, k} \quad \frac{1}{2} \sum_{i=1}^N \|u_i - u_{M,i}\|_{L^2(\Gamma_M)}^2 + \frac{\alpha}{2} \|k\|_{[H^s(Y)]^{2 \times 2}}^2 \\
& \text{subject to} \quad e_i(u_i, M_i) = 0 \quad i \in \{1, \dots, N\} \\
& \quad \quad \quad f_{i,j}(k, w_{i,j}) = 0 \quad i \in \{1, \dots, N\}, j \in \{1, 2\} \\
& \quad \quad \quad M_i = \int_Y k \left(I + \begin{bmatrix} \nabla w_{i,1} & \nabla w_{i,2} \end{bmatrix} \right) dy \quad i \in \{1, \dots, N\}
\end{aligned} \tag{4.16}$$

where the subscript i, j denotes cell-problem j corresponding to the upscales equation indexed by i . This problem also suffers potential issues in regards to proving well-posedness unless k belongs to a finite dimensional space. But just as for (4.15) this can be resolved by linearization, using the Fréchet derivatives derived in Section 3.4, or by trying to find reasonable finite dimensional spaces for the coefficient k . The latter is studied in [14] for (4.16) as well.

5 Numerical simulations

In this section we demonstrate numerically the inversion for the upscaled inverse problem in the sense of Definition 4.2 by the least-squares optimization framework described in Section 4.6.1. We construct a simulated measurement of the ε -dependent forward problem and then try to find an effective coefficient that describes the measurement. We use three slightly different approaches to perform the inversion. One approach is to solve precisely the optimization problem studied in Section 4.6.1. In this approach we force the coefficients to be piecewise constant by optimizing over a space spanned by a basis of piecewise constant functions. As a comparison, we solve also the optimization problem by optimizing over a basis of finite elements. The latter approach could be viewed as solving the ε -dependent inverse problem in the sense of Definition 4.1, since the coefficients used in the inversion procedure are described numerically in the same way as the coefficient used in the simulation of the ε -dependent measurement. As another comparison, we optimize again over a basis of finite elements but we use a total variation penalty instead of a Tikhonov-style penalty. The idea is that the total variation penalty might drive the solutions more towards the piecewise constant structure of an effective coefficient than a Tikhonov-style penalty does.

5.1 The optimization framework

Before stating the optimization problems, let us discuss the total variation functional. The total variation of a function $f \in L^1(\Omega)$ can be defined by

$$\text{T.V.}(f) = \sup \left\{ \int_{\Omega} f(x) \operatorname{div} \varphi(x) dx : \varphi \in [C_0^1(\Omega)]^2, \|\varphi\|_{[L^\infty(\Omega)]^2} \leq 1 \right\}. \quad (5.1)$$

If, in addition, $f \in C^1(\bar{\Omega})$ then the total variation can be written

$$\text{T.V.}(f) = \int_{\Omega} |\nabla f(x)| dx. \quad (5.2)$$

We are interested in using the penalty functional

$$J(k) = \sum_{i=1}^2 \sum_{j=1}^2 \text{T.V.}(k_{ij}).$$

The total variation (5.2) is better suited for numerical implementation than (5.1), due to the presence of the supremum in (5.1). But using (5.2) does not resolve all issues with the total variation functional. We will use a gradient based optimization algorithm and therefore require the functional to be Fréchet differentiable, which the functional T.V. is not. Instead of changing the optimization algorithm, we make a slight modification to the functional. For some small $\beta > 0$, let

$$J_{\beta}(k) = \sum_{i=1}^2 \sum_{j=1}^2 \int_{\Omega} \sqrt{|\nabla k_{ij}(x)|^2 + \beta} dx. \quad (5.3)$$

The idea is to take β so small that J_{β} is approximately equal to T.V. Such approximation is proved rigorously in [1]. This justifies replacing the penalty functional T.V. with J_{β} .

Now let us define the optimization problems we wish to solve. Let \mathcal{V} denote the space of diffusion coefficients, which is soon to be specified precisely, and let $\mathcal{U} = H^1(\Omega)$ denote the space of solutions u to the partial differential equations. The constraints to the optimization problem is defined by the functionals $e_i: \mathcal{U} \times \mathcal{V} \rightarrow \mathcal{U}^*$, which is given by

$$e_i(u_i, k)v = \int_{\Omega} (k \nabla u_i) \cdot \nabla v dx - \int_{\Gamma^{in}} g_i v d\sigma,$$

where u_i is the solution that is fitted to the i th measurement $u_{M,i}$. We optimize the penalized least-squares function T_{α} defined by

$$T_{\alpha}(u, k; u_M, k_0) = \frac{1}{2} \sum_{i=1}^N \|u_i - u_{M,i}\|_{L^2(\Gamma_M)}^2 + \alpha J(k).$$

⁵This definition of J_{β} requires k to be differentiable. Just like T.V. can be extended to nonsmooth functions, so can J_{β} , but such an extension uses the duality theory of convex functions which is a topic we will not discuss. See the Section 8.4.1 in [25] for some discussion and further references.

We consider two choices of J , namely $J(k) = \frac{1}{2}\|k - k_0\|_{[L^2(\Omega)]^{2 \times 2}}^2$ and $J(k) = J_\beta(k)$, where J_β is defined by (5.3).

When we optimize over a piecewise constant basis, we define the basis functions as follows. Let

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and define the matrix functions f_{li}, f_{1i}, f_{ri} by

$$f_{ji}(x) = \begin{cases} M_i & \text{if } x \in \Omega_j \\ 0 & \text{otherwise.} \end{cases}$$

for $j \in \{l, 1, r\}, i \in \{1, 2, 3\}$. We consider the following three combinations of parameter space \mathcal{V} and penalty functional J .

- i) $\mathcal{V} = \text{Span}\{f_{ji}, j \in \{l, 1, r\}, i \in \{1, 2, 3\}\}, J(k) = \frac{1}{2}\|k - k_0\|_{[L^2(\Omega)]^{2 \times 2}}^2$,
- ii) $\mathcal{V} = [L^\infty(\Omega)]^{2 \times 2}, J(k) = \frac{1}{2}\|k - k_0\|_{[L^2(\Omega)]^{2 \times 2}}^2$,
- iii) $\mathcal{V} = [L^\infty(\Omega)]^{2 \times 2}, J(k) = J_\beta(k - k_0)$, where J_β is defined in (5.3).

For each of the above three choices, we consider a corresponding optimization problems which we denote by Problem i), Problem ii), Problem iii) throughout this chapter. The optimization problems are defined by

$$\begin{aligned} \min_{u \in \mathcal{U}, k \in \mathcal{V}} \quad & T_\alpha(u, k; u_M, k_0) \\ \text{subject to} \quad & e_i(u_i, k) = 0 \quad i \in \{1, \dots, N\}, \end{aligned} \tag{5.4}$$

and N is the number of measurements.

5.1.1 Gradient calculations

Now we wish to solve the optimization problem (5.4). Before describing the algorithm which we use to solve this, let us derive the equations that we will solve as part of the algorithm. Mainly, we need to derive equations for the Fréchet derivative of the constrained functional. By using the parameter-to-solution map, we can write $u = \mathcal{G}_{eff}(k)$ as a function of k . Then we need to calculate the Fréchet derivative of $T_\alpha(k) = T_\alpha(\mathcal{G}_{eff}(k), k; u_M, k_0)$ with respect to k . We could use the directional derivatives derived in Section 3.4. But this turns out to perform badly, since we need to calculate the derivative in the direction of every basis vector and for each such directional derivative we need to solve (3.30). Instead we now derive the so-called adjoint equation. After solving the adjoint equation only once, we can use FEniCS in order to efficiently assemble the Fréchet derivative.

We take the Lagrangian approach to deriving the adjoint equation. This approach is rigorously justified in Chapter 9.6 in [21]. One of the few requirements is that the forward map is Lipschitz continuous, which we know from Section 3.4 that it is. Let us first consider

the L^2 -penalty functional of Problem i) and Problem ii). Consider the Lagrangian functional $L: \mathcal{U}^N \times \mathcal{V} \times [\mathcal{U}^*]^N \rightarrow \mathbb{R}$ defined by

$$L(u, k, \lambda) = T_\alpha(u, k; u_M; k_0) + \sum_{i=1}^N \lambda_i(e_i(u_i, k)).$$

Since \mathcal{U} is a Hilbert space and therefore is reflexive, we have that for each $\lambda_i \in \mathcal{U}^*$ there is a vector also denoted $\lambda_i \in \mathcal{U}$ such that $\lambda_i(e(u_i, k)) = e(u_i, k)\lambda_i$. Then

$$L(u, k, \lambda) = T_\alpha(u, k; u_M; k_0) + \sum_{i=1}^N e_i(u_i, k)\lambda_i.$$

Now we form the system of $2N$ equations

$$0 = \frac{\partial L(u, k, \lambda)}{\partial u_i} u_i^* = \int_{\Gamma_M} (u_i - u_{M,i}) u_i^* d\sigma + \int_{\Omega} (k \nabla \lambda_i) \cdot \nabla u_i^* dx \quad (5.5)$$

$$0 = \frac{\partial L(u, k, \lambda)}{\partial \lambda_i} \lambda_i^* = \int_{\Omega} (k \nabla u_i) \cdot \nabla \lambda_i^* dx - \int_{\Gamma^{in}} g_i \lambda_i d\sigma. \quad (5.6)$$

Note that for each $i \in \{1, \dots, N\}$, equation (5.6) is just the upscaled forward equation and it has only the unknown u_i , meaning that we can solve (5.6) without considering (5.5). When u_i is known then the only unknown in (5.5) is λ_i . Therefore, we first solve (5.6) for u_i and then solve (5.5) for λ_i . Recall that in the present setup, we have $u_i, \lambda_i \in \mathcal{U} = H^1(\Omega)$, which means that the equations (5.6), (5.5) are not uniquely solvable for u_i, λ_i . To obtain uniqueness of u_i we can simply take $u_i \in \mathcal{U}_{ad} = H_\diamond^1(\Omega; m)$, since (5.6) is the upscaled forward equation. In order to obtain a unique λ_i , let us first look at how we will use λ_i in the gradient $\frac{dT_\alpha}{dk}$. We have

$$\begin{aligned} \frac{dT_\alpha(u, k; u_M, k_0)}{dk} k^* &= \frac{\partial L(u, k, \lambda)}{\partial k} k^* = \alpha \langle k, k^* \rangle_{[L^2(\Omega)]^{2 \times 2}} + \sum_{i=1}^N e_i(u_i, k^*) \lambda_i \\ &= \alpha \sum_{i=1}^2 \sum_{j=1}^2 \int_{\Omega} k_{ij} k_{ij}^* dx + \sum_{i=1}^N \int_{\Omega} (k^* \nabla u_i) \cdot \nabla \lambda_i^* dx. \end{aligned}$$

We see that the function we actually use is the gradient $\nabla \lambda_i$ and not λ_i . Therefore, additive shifts in λ_i makes no difference and we can choose $\lambda_i \in H_\diamond^1(\Omega; 0)$ in order to obtain uniqueness of solution to (5.5). Next, we need to obtain weak formulations of these equations that we can implement. We do this by using the Lagrangian approach to include the integral constraints into the weak formulations, as discussed in section 2.3. In the Lagrangian based weak formulation of (5.6) we seek the pair $(u_i, r_i) \in H^1(\Omega) \times \mathbb{R}$ such that

$$\int_{\Omega} (k \nabla u_i) \cdot \nabla v + r_i v + \mu u_i dx = \mu m - \int_{\Gamma^{in}} g v d\sigma \quad \forall (v, \mu) \in H^1(\Omega) \times \mathbb{R} \quad (5.7)$$

and for (5.5) we seek the pair $(\lambda_i, s_i) \in H^1(\Omega) \times \mathbb{R}$ such that

$$\int_{\Omega} (k \nabla \lambda_i) \cdot \nabla v + s_i v + \mu \lambda_i dx = - \int_{\Gamma^{in}} (u_i - u_{M,i}) v d\sigma \quad \forall (v, \mu) \in H^1(\Omega) \times \mathbb{R}. \quad (5.8)$$

Next, we need to find a vector representation $g \in [L^2(\Omega)]^{2 \times 2}$ of $\frac{dT_\alpha}{dk}$, which we do by solving for g in the equation $\langle g, k^* \rangle_{[L^2(\Omega)]^{2 \times 2}} = \frac{dT_\alpha}{dk} k^*$. More explicitly, this equation reads

$$\sum_{i=1}^2 \sum_{j=1}^2 \int_{\Omega} g_{ij} k_{ij}^* dx = \alpha \sum_{i=1}^2 \sum_{j=1}^2 \int_{\Omega} k_{ij} k_{ij}^* dx + \sum_{i=1}^N \int_{\Omega} (k^* \nabla u_i) \cdot \nabla \lambda_i dx. \quad (5.9)$$

With such a g at hand, we can easily calculate the operator norm of the gradient as follows

$$\left\| \frac{dT_\alpha(u, k, \lambda)}{dk} \right\| = \|g\|_{[L^2(\Omega)]^{2 \times 2}}.$$

When we use the total variation penalty in Problem iii) we can compute the gradient by the same reasoning as above after some modification to (5.9). The Fréchet derivative of the total variation penalty functional is given by

$$\frac{dJ_\beta(k)}{dk} h = \sum_{i=1}^2 \sum_{j=1}^2 \int_{\Omega} \frac{\nabla k_{ij} \cdot \nabla h_{ij}}{\sqrt{|\nabla k_{ij}|^2 + \beta}} dx.$$

Instead of (5.9), we now have

$$\sum_{i=1}^2 \sum_{j=1}^2 \int_{\Omega} g_{ij} k_{ij}^* dx = \alpha \sum_{i=1}^2 \sum_{j=1}^2 \int_{\Omega} \frac{\nabla k_{ij} \cdot \nabla k_{ij}^*}{\sqrt{|\nabla k_{ij}|^2 + \beta}} dx + \sum_{i=1}^N \int_{\Omega} (k^* \nabla u_i) \cdot \nabla \lambda_i dx. \quad (5.10)$$

5.1.2 The conjugate gradient algorithm with inexact line search

When solving the problem (5.4) we use Algorithm 1 and Algorithm 2. These are, respectively, the conjugate gradient algorithm and an inexact line search with Armijo's rule, adapted to the equations derived in Section 5.1.1. Instead of describing the algorithms in detail, we refer to [25, 10].

Algorithm 1 Conjugate gradient algorithm for effective coefficient inversion

Input: Functional $T_\alpha(\cdot, \cdot; u_M, k_0)$; Initial guess k_{init} ; Stop criterion $\varepsilon > 0$

Output: Approximation of $k = \operatorname{argmin}_k T_\alpha$

$\nu = 1$
 $k_1 = k_{init}$
Find u_1, λ_1 by solving (5.7) and (5.8) with k_1
Find gradient g_1 by solving (5.9)
Set search direction $p_1 = -g_1$
 $\delta_1 = \|g_1\|_{[L^2(\Omega)]^{2 \times 2}}^2$
while $\|g_\nu\|_{[L^2(\Omega)]^{2 \times 2}} > \varepsilon$ **do**
 if $\langle g_\nu, p_\nu \rangle_{[L^2(\Omega)]^{2 \times 2}} \geq 0$ **then**
 $p_\nu = -g_\nu$
 $\delta_\nu = \|g_\nu\|_{[L^2(\Omega)]^{2 \times 2}}^2$
 end if
 $\tau_\nu = \operatorname{argmin}_{\tau > 0} T_\alpha(k_\nu + \tau p_\nu)$
 $k_{\nu+1} = k_\nu + \tau_\nu p_\nu$
 Find $u_{\nu+1}, \lambda_{\nu+1}$ by solving (5.7) and (5.8) with $k_{\nu+1}$
 Find gradient $g_{\nu+1}$ by solving (5.9)
 $\delta_{\nu+1} = \|g_{\nu+1}\|_{[L^2(\Omega)]^{2 \times 2}}^2$
 $\beta_{\nu+1} = \delta_{\nu+1} / \delta_\nu$
 Update search direction $p_{\nu+1} = -g_{\nu+1} + \beta_{\nu+1} p_\nu$
 $\nu = \nu + 1$
end while

For the input parameters in Algorithm 1 we use the following. The initial guess k_{init} is set to

$$k_{init} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The stop criterion ε is set to 10^{-8} and 10^{-9} , that is, we run the tests twice but with different stop criteria. The prior knowledge k_0 about the true coefficient is set to

$$k_0|_{\Omega_0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad k_0|_{\Omega_1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The generation of the measurement u_M is described in Section 5.2.1 and the various choices of the regularization parameter is described in Section 5.1.4. The input parameters used in Algorithm 2 are

$$\tau_{init} = 1000, \quad \beta = 0.5, \quad \gamma = 0.2.$$

The other input parameters to Algorithm 2 vary as they are generated for each iteration in Algorithm 1. Both algorithms are terminated if they run for 2000 iterations without reaching convergence.

Algorithm 2 Inexact line search with Armijo's rule for effective coefficient inversion

Input: Functional $T_\alpha(\cdot, \cdot; u_M, k_0)$; Base point k_b and corresponding state u_b ; Regularization parameter α ; Gradient g_b at k_b ; Search direction p ; Initial step size τ_{init} ; Small weight γ ; Step size adjustment factor β

Output: Approximation of $\tau = \operatorname{argmin}_{\tau > 0} T_\alpha(k + \tau p)$

```
 $\nu = 1$   
 $\tau_1 = \tau_{init}$   
 $T^* = T_\alpha(k_b, u_b; u_M, \bar{k})$   
 $d^* = \langle g_b, p \rangle_{[L^2(\Omega)]^{2 \times 2}}$   
 $k_1 = k_b + \tau_1 p$   
Find  $u_1$  corresponding to  $k_1$  by solving (5.7)  
 $T_1 = T_\alpha(k_1, u_1; u_M, k_0)$   
while  $T_\nu > T^* + \tau_\nu \gamma d^*$  do  
     $\tau_{\nu+1} = \beta \tau_\nu$   
     $k_{\nu+1} = k_b + \tau_{\nu+1} p$   
    Find  $u_{\nu+1}$  corresponding to  $k_{\nu+1}$  by solving (5.7)  
     $T_{\nu+1} = T_\alpha(k_{\nu+1} + \tau_{\nu+1} p, u_{\nu+1}; u_M, k_0)$   
     $\nu = \nu + 1$   
end while
```

5.1.3 A few remarks on the performance of the inexact line search

A reason for using an inexact line search is that it may not be required for the convergence of the conjugate gradient algorithm that the line search results in a minimum. The conjugate gradient algorithm may still converge provided a sufficiently large descent is obtained in the line search phase. But there are several parameters to tune in Algorithm 2 and bad choices of these parameters can result in the line search not managing to find a sufficient descent. This does happen in our numerical evaluation. In an attempt to resolve this, larger values for β were tested, as well as both larger and smaller values of γ . But the issues persisted. Permanently changing the values of β and γ in this way may drastically reduce the improvements made in each conjugate gradient iteration, which causes the conjugate gradient algorithm to converge very slowly. But $\beta = 0.5$ and $\gamma = 0.2$ appears to yield relatively good improvements for each conjugate gradient iteration. For this reason we also tried to initially set $\beta = 0.5$ and $\gamma = 0.2$ and update them only when the line search failed. But once the conjugate gradient algorithm reached an iteration where the line search failed for one choice of β and γ , the line search failed for every other choice as well.

The conclusion is that our implementation of the algorithms sometimes fail. Since the issues were never resolved, we present only the results obtained by taking the fixed choice of $\beta = 0.5$ and $\gamma = 0.2$.

5.1.4 Choosing a regularization parameter

A good choice of regularization parameter is crucial in obtaining a good inversion. If we take a too large regularization parameter then the penalty term dominates the misfit term and

then the minimizers may not be able to explain the observations. A too small regularization parameter may, on the other hand, not provide enough information for the ill-posedness of the original problem to be resolved. In this thesis, we do not present a systematic and reliable way to find a good regularization parameter for this problem. Our goal merely is to demonstrate that it is possible to find a good regularization parameter.

Since this is a computational study and we have generated the observations ourselves, we know precisely what the true effective coefficient is. Let us denote the parameter by k_{true} . We use this knowledge of k_{true} to guide the decision on which regularization parameters to try. By no means is this a practical approach to finding a good regularization parameter. In a real world scenario, we would have no information about k_{true} and could therefore not use it to guide our decisions. But by using this information we can at least demonstrate that there exist good regularization parameter.

During the run-time of the conjugate gradient algorithm, we can monitor the error norm $\|k_n - k_{true}\|_{[L^2(\Omega)]^{2 \times 2}}$, where k_n is the candidate coefficient at conjugate gradient iteration n . The error norm starts out at around 0.4. For some regularization parameters the error norm does not change much as the algorithm progresses, but for other regularization parameters the error norm may at some point decrease to values as low as 0.02. At this low error norm, the candidate solution k_n is almost identical to the true effective coefficient k_{true} . We would therefore want to find a regularization parameter for which the algorithm converges sometime during this period of decreased error norm. This appears to have a tendency to happen for regularization parameters in the interval $[10^{-9}, 10^{-7}]$ and in particular closer to 10^{-9} . Therefore, we test the regularization parameters $c \cdot 10^{-9}$ and $c \cdot 10^{-8}$ for $c \in \{1, \dots, 9\}$, in addition to 10^{-7} and 10^{-6} .

5.2 The simulation setup

5.2.1 The measurement

We use a unit square domain, $\Omega = (0, 1) \times (0, 1)$, and the 3 subdomains are given by $\Omega_l = (0, 0.45) \times (0, 1)$, $\Omega_r = (0.55, 1) \times (0, 1)$ and $\Omega_1 = (0.45, 0.55) \times (0, 1)$. The subset Γ^{in} of $\partial\Omega$ with nonzero particle flux is $\Gamma^{in} = \{0\} \times (0, 1)$. The flux is constant at $g = 1$. For ease of implementation, the mass is set to $m = 0$. FEniCS does not appear to allow incorporating an integral constraint $\int_{\Omega} u_{\varepsilon} dx = m$ with nonzero right-hand side into the weak formulation. A nonzero mass m would therefore have to be implemented manually by repeatedly shifting the solution. But since a nonzero m is not important from a mathematical point of view, it might as well be set to 0 to avoid having to shift the solutions. The coefficient k_{ε} is defined by

$$k_{\varepsilon}(x) = \begin{cases} k_0 & \text{if } x \in \Omega_l \cup \Omega_r, \\ k_1\left(\frac{x}{\varepsilon}\right) & \text{if } x \in \Omega_1, \end{cases}$$

where $k_0 = I$ is the identity matrix and $\varepsilon > 0$. In order to define k_1 , first consider $h \in L^{\infty}_{\#}(Y)$ defined by

$$h(y) = \begin{bmatrix} \alpha \sin(2\pi y_1)^2 & (\alpha - \beta) \sin(2\pi y_1) \cos(2\pi y_2) \\ (\alpha - \beta) \sin(2\pi y_1) \cos(2\pi y_2) & \beta \cos(2\pi y_2)^2 \end{bmatrix}$$

for some constant $\alpha, \beta > 0$. We choose α, β so that h is positive semidefinite for $y \in Y$ and then define $k_1(y) = \gamma I + h(y)$ for some $\gamma > 0$. Then k_1 is positive definite for $y \in Y$ since h is positive semidefinite and γI is positive definite. At least for $\frac{\beta}{\sqrt{2}} \leq \alpha \leq \beta$, h is positive semidefinite. To see this we can, for example, use Theorem 3.3.12 in [10]. First if either $h_{11}(y) = 0$ or $h_{22}(y) = 0$ we see that also $h_{12}(y) = h_{21}(y) = 0$, as required. Also,

$$h(y) \sim \begin{bmatrix} \alpha \sin(2\pi y_1)^2 & (\alpha - \beta) \sin(2\pi y_1) \cos(2\pi y_2) \\ 0 & \frac{\alpha\beta - (\alpha - \beta)^2}{\alpha} \cos(2\pi y_2)^2 \end{bmatrix},$$

where \sim denotes equivalence by the standard row and column operations. For $\frac{\beta}{\sqrt{2}} \leq \alpha \leq \beta$ we have $\frac{\alpha\beta - (\alpha - \beta)^2}{\alpha} \geq 0$ so that $\frac{\alpha\beta - (\alpha - \beta)^2}{\alpha} \cos(2\pi y_2)^2 \geq 0$. Then it follows from Theorem 3.3.12 in [10] that h is positive semidefinite and then k_1 is positive definite. In particular we choose, rather arbitrarily, $\alpha = 0.3, \beta = 0.4$ and $\gamma = 0.2$. Then

$$k_1(y) = \begin{bmatrix} 0.2 + 0.3 \cdot \sin(2\pi y_1)^2 & -0.1 \cdot \sin(2\pi y_1) \cos(2\pi y_2) \\ -0.1 \cdot \sin(2\pi y_1) \cos(2\pi y_2) & 0.2 + 0.4 \cdot \cos(2\pi y_2)^2 \end{bmatrix}.$$

5.2.2 Some implementation details

In this section we go through some details regarding how the problem is represented numerically. We do this because the numerical representation of the above mathematical objects deviate somewhat from their true mathematical structure. For example, when we optimize over a piecewise constant basis, the functions are not actually piecewise constant. So the purpose of this section is mainly to show how the different objects are represented.

The unit square domain Ω is generated by the function call `mesh = UnitSquareMesh(64, 64)`. This generates a mesh of triangles on the unit square such that there fits 64 triangles side by side in each direction, and stores the mesh in the variable `mesh`. The ε -dependent coefficient k_ε is given a finite element representation, constructed by the function call `as_tensor([[k11, k12], [k21, k22]])` where the components $k_{11}, k_{12}, k_{21}, k_{22}$ are constructed by the following function calls

```
k11 = Expression("x[0] >= 0.45 && x[0] <= 0.55 ?
                 a*pow(sin(2*DOLFIN_PI*x[0]/e),2) + c : 1",
                 a=alpha, c=gamma, e=epsilon, degree=1)
k12 = Expression("x[0] >= 0.45 && x[0] <= 0.55 ?
                 (a-b)*sin(2*DOLFIN_PI*x[0]/e)*cos(2*DOLFIN_PI*x[1]/e) : 0",
                 a=alpha, b=beta, e=epsilon, degree=1)
k21 = Expression("x[0] >= 0.45 && x[0] <= 0.55 ?
                 b*pow(cos(2*DOLFIN_PI*x[1]/e),2) + c : 1",
                 b=beta, c=gamma, e=epsilon, degree=1)
```

and the definition `k21` is identical to that of `k12`. The constants `alpha, beta, gamma, epsilon` are constants defined earlier in the code. The code inside "..." in `Expression` above is C++ code

and it describes the region inside Ω_l and outside Ω_l by an inline C++ conditional expression. The piece of code `degree=1` means that whenever the coefficients are used in computations, they are first interpolated into a finite element space defined by Lagrange finite elements of degree 1. So, rather than an actual jump discontinuity, the numerical implementation has a steep, linear shift in value between Ω_l, Ω_r and Ω_1 .

Whenever we solve a partial differential equation, we always seek a finite element solution. In FEniCS, these finite elements are defined as `P = FiniteElement("Lagrange", mesh.ufl_cell(), 1)`. The provided mesh is the variable `mesh` with the triangulation of the unit square defined above. Since the Lagrange finite elements of degree 1 are uniquely determined by their values at vertices in the mesh, we see that all functions used in the implementation have the same degrees of freedom.⁶ When we solve an equation in a product space such as $H^1(\Omega) \times \mathbb{R}$, we use the space defined by `U = FunctionSpace(mesh, P*R)`, where `R = FiniteElement("Real", mesh.ufl_cell(), 0)` are the finite elements for constant real number that represent the Lagrange multipliers.

The diffusion coefficient k that we use as variable in the optimization algorithm is represented in two different ways. For problems ii) and iii) we define the coefficient as an element in the space defined by `V = TensorFunctionSpace(mesh, "Lagrange", 1, (2,2))`. V is a space of 2×2 matrix-valued functions whose matrix components are represented by Lagrange finite elements of degree 1. Now all arithmetic with the coefficients is done by FEniCS and we also use FEniCS to very efficiently construct the finite element approximations of integral equations such as (5.9).

For problem i), we instead define the coefficients by `k = as_tensor([[k11, k12], [k21, k22]])` where each coefficient is defined by `k11 = Expression("x[0] <= 0.45 ? c1 : x[0] <= 0.55 ? c2 : c3", c1=a, c2=b, c3=c, degree=1)`. This produces a function which is equal to c_1 in Ω_l , c_2 in Ω_1 and c_3 in Ω_r . Note that when these functions are used to compute integrals, they are first interpolated into a finite element space defined by Lagrange finite elements of degree 1. Just as discussed for k_ε above, this means that the discontinuities are approximated linearly. But in the interior of Ω_l, Ω_1 and Ω_r , the functions are still constant when defined in this way. When performing arithmetic with these functions, we now manually perform the arithmetic in Python and then update the corresponding component c_1, c_2, c_3 . Since we are not using the FEniCS functionality of function spaces, we cannot use the efficient FEniCS implementation of the construction of matrix representations for equations such as (5.9). Therefore, we now manually loop over the basis elements and calculate the integrals for each basis element. The integral evaluation is still done by FEniCS, by calling the function `assemble`. If we used the function spaces implemented by FEniCS then `assemble` would compute the whole matrix at once in efficient C++ code. Now we instead construct the matrix component by component in rather inefficient Python code, but it still performs well enough due to the low dimension of the function space in problem i).

⁶This situation where the degrees of freedom used in the reconstruction is the same as those used in the data generation is sometimes called an inverse crime, and often lead to better reconstructions than in real life scenarios. This inverse crime is committed in problems ii) and iii) but not in i).

5.3 Inversion in the case of fine microstructure

5.3.1 The simulated measurement

Now solving equation (3.2) with the data defined in Section 5.2 and with $\varepsilon = 0.1$ yields a finite element discretization of the observation u_ε . The program outputs the following norms related to the measurement of u_ε and the approximation error

$$\begin{aligned}\|u_\varepsilon - u_0\|_{L^2(\Omega)} &\approx 5.786 \cdot 10^{-3} \\ \|u_\varepsilon - u_0\|_{L^2(\Gamma_M)} &\approx 5.898 \cdot 10^{-3} \\ \|u_\varepsilon\|_{L^2(\Gamma_M)} &\approx 0.2217 \\ \|u_0\|_{L^2(\Gamma_M)} &\approx 0.2216.\end{aligned}$$

We see that the error introduced by considering the ε -dependent u_ε as a noisy observation of the solution u_0 of the upscaled equation (3.13) is rather small. Relative to the measurement of u_0 , the size of the error is only about 2.7%, $\|u_\varepsilon - u_0\|_{L^2(\Gamma_M)} / \|u_0\|_{L^2(\Gamma_M)} \approx 2.662 \cdot 10^{-2}$. The numerical values calculated for the effective coefficient are

$$D_{eff}|_{\Omega_0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D_{eff}|_{\Omega_1} \approx \begin{bmatrix} 0.314435800 & 9.06289607 \cdot 10^{-6} \\ 9.06289607 \cdot 10^{-6} & 0.344885901 \end{bmatrix} \quad (5.11)$$

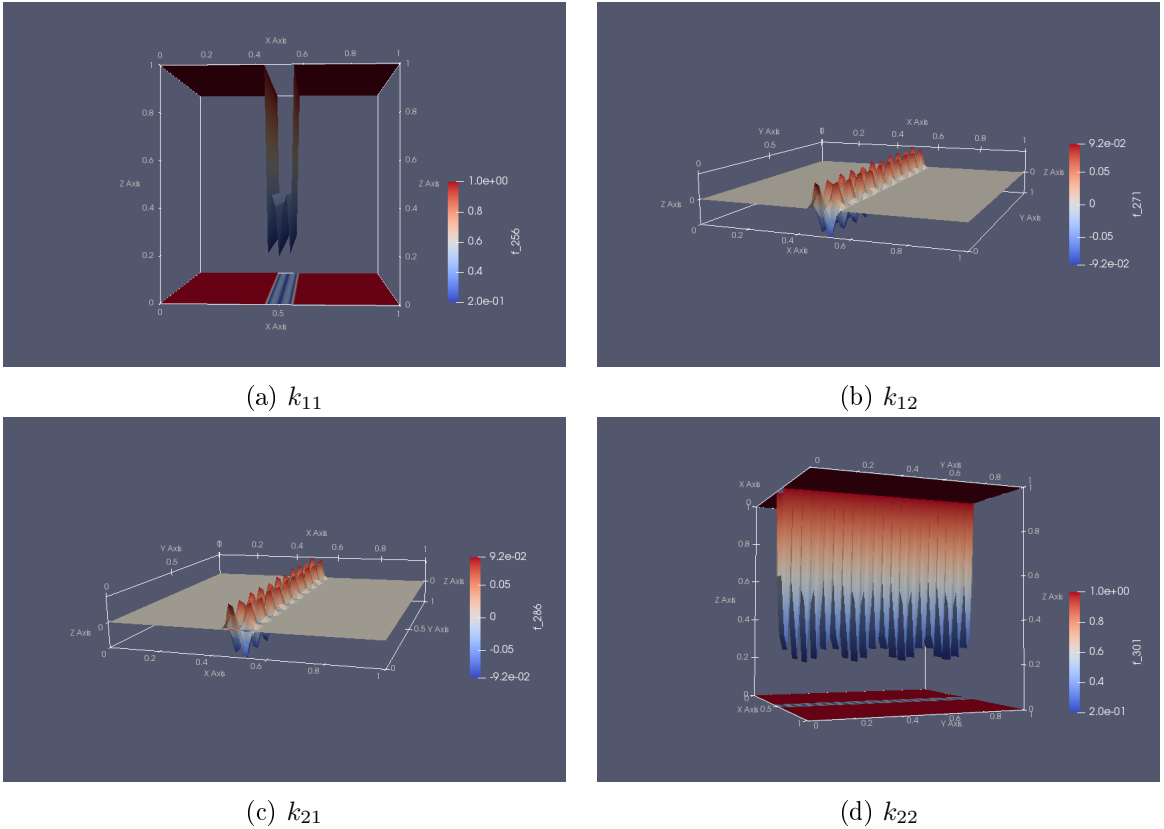


Figure 5.1: Plots of k_ε , shown from angles such that the oscillations are visible.

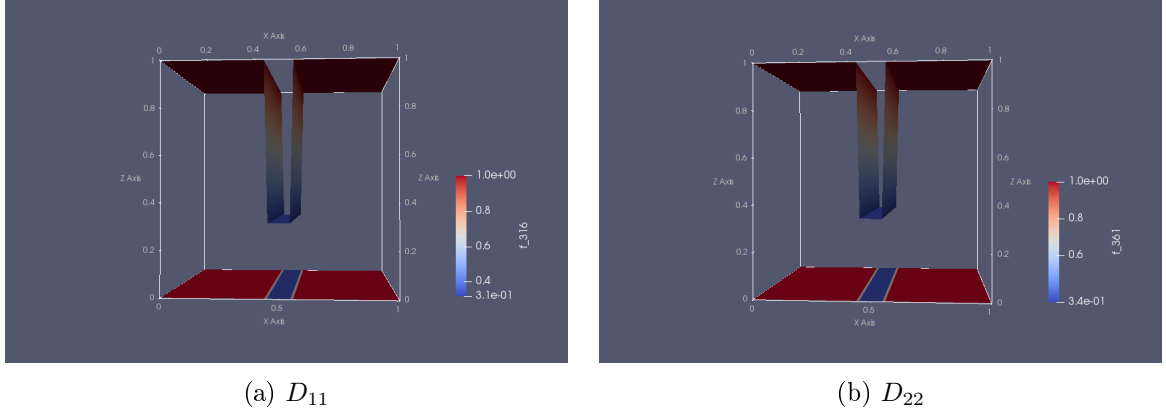


Figure 5.2: Plots of the nonzero components of the true effective coefficient D .

5.3.2 Numerical results

In Table 5.1 is a numerical summary of the results for solving Problem i) with stop criterion 10^{-8} . When taking regularization parameter $\alpha < 10^{-8}$ the algorithm always terminated at conjugate gradient iteration 12, just as for $\alpha = 10^{-8}$. Therefore, the parameters $\alpha < 10^{-8}$ are not shown in Table 5.1. This early termination of the algorithm is due to an rapid decrease in the gradient norm that always occur as the algorithm starts. Later, the norm of the gradient often stabilizes around 10^{-5} or 10^{-6} . For small regularization parameters, this initial decrease in gradient norm reaches all the way to 10^{-9} and the algorithm terminates.

By judging the quality of the inversion by its error norm, we see in Table 5.1 that the best inversion is obtained for regularization parameter $\alpha = 6 \cdot 10^{-8}$. In this case the estimated coefficient D is

$$D|_{\Omega_l} \approx \begin{bmatrix} 0.99493319 & -0.00167329 \\ -0.00167329 & 1.00008229 \end{bmatrix}, \quad D|_{\Omega_1} \approx \begin{bmatrix} 0.3489804 & 0.00305011 \\ 0.00305011 & 0.22440582 \end{bmatrix},$$

$$D|_{\Omega_r} \approx \begin{bmatrix} 0.995038486 & -7.05975190 \cdot 10^{-4} \\ -7.05975190 \cdot 10^{-4} & 1.00062139 \end{bmatrix},$$

Regularization parameter $\alpha = 3 \cdot 10^{-8}$ also results in a rather good inversion, identifying the following effective coefficient

$$D|_{\Omega_l} \approx \begin{bmatrix} 0.987126634 & -9.01384881 \cdot 10^{-4} \\ -9.01384881 \cdot 10^{-4} & 1.00008541 \end{bmatrix}, \quad D|_{\Omega_1} \approx \begin{bmatrix} 0.34954941 & 0.00184867 \\ 0.00184867 & 0.21479563 \end{bmatrix},$$

$$D|_{\Omega_r} \approx \begin{bmatrix} 0.987188447 & -3.54856805 \cdot 10^{-4} \\ -3.54856805 \cdot 10^{-4} & 1.00065018 \end{bmatrix}.$$

These two coefficients are by far the best ones identified in the tests presented in Table 5.1 and they agree quite well with the true effective coefficient (5.11). But we should note that for every tested regularization parameter $\alpha \in [3 \cdot 10^{-8}, 10^{-6}]$, every component of the true effective coefficient (5.11) is identified rather well except for the component $D_{22}|_{\Omega_1}$. Judging

Table 5.1: Summary of the results for Problem i) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-8}$.

regularization parameter α	converged	conjugate gradient iterations	error norm $\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
10^{-6}	yes	207	0.1120803
10^{-7}	yes	887	0.1151009
$9 \cdot 10^{-8}$	yes	707	0.1341151
$8 \cdot 10^{-8}$	yes	477	0.1455599
$7 \cdot 10^{-8}$	yes	250	0.1021478
$6 \cdot 10^{-8}$	yes	259	0.04034444
$5 \cdot 10^{-8}$	yes	961	0.09247621
$4 \cdot 10^{-8}$	yes	271	0.07725089
$3 \cdot 10^{-8}$	yes	360	0.04509805
$2 \cdot 10^{-8}$	yes	116	0.2779014
10^{-8}	yes	12	0.3149823

by the error norm, $\alpha = 8 \cdot 10^{-8}$ resulted in the worst approximation among the regularization parameters in this interval, and it gave the following approximation

$$D|_{\Omega_l} \approx \begin{bmatrix} 1.07518722 & -0.00160862 \\ -0.00160862 & 1.00013149 \end{bmatrix}, \quad D|_{\Omega_1} \approx \begin{bmatrix} 0.29802296 & 0.00222055 \\ 0.00222055 & -0.04895257 \end{bmatrix},$$

$$D|_{\Omega_r} \approx \begin{bmatrix} 1.07530267 & -5.88897264 \cdot 10^{-4} \\ -5.88897264 \cdot 10^{-4} & 1.00057952 \end{bmatrix}.$$

In Table 5.2 are the results for Problem i) with stop criterion 10^{-9} . We see in Table 5.2 that the algorithm does not terminate early for small regularization parameters, except for $\alpha = 10^{-9}$. So despite the fact that the algorithm reports a convergent results for $\alpha = 10^{-9}$, it does not manage to identify the coefficient at all. In fact, even the component $D_{11}|_{\Omega_0}$, which is set to the correct value of 1 in the initial guess, has decreased to around 0.8.

Judging by the error norm, the best inversion obtained in Table 5.2 corresponds to one of the regularization parameters that terminated early in Table 5.1, namely, $\alpha = 6 \cdot 10^{-9}$. Among all solutions in Table 5.2, the solution which manages to find the best approximation for the usually problematic component $D_{22}|_{\Omega_1}$ is also the solution corresponding to $\alpha = 6 \cdot 10^{-9}$. The solution is the following

$$D|_{\Omega_l} = \begin{bmatrix} 1.00751329 & -0.00104648 \\ -0.00104648 & 1.00005013 \end{bmatrix}, \quad D|_{\Omega_1} = \begin{bmatrix} 0.33531572 & 0.00159613 \\ 0.00159613 & 0.12197899 \end{bmatrix},$$

$$D|_{\Omega_r} = \begin{bmatrix} 1.00766446 & -1.40893580 \cdot 10^{-6} \\ -1.40893580 \cdot 10^{-6} & 1.00034603 \end{bmatrix}.$$

The general trend for the solutions in Table 5.2 appears to be the same as the trend for the solutions in Table 5.1, namely, that the algorithm manages to identify rather well all components of the true effective coefficient (5.11) except component $D_{22}|_{\Omega_1}$ for every regularization parameter α , except for $\alpha = 10^{-9}$ that terminated early.

Table 5.2: Summary of the results for Problem i) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-9}$.

regularization parameter α	converged	conjugate gradient iterations	$\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
10^{-6}	yes	305	0.1131184
10^{-7}	yes	1680	0.1133869
$9 \cdot 10^{-8}$	yes	993	0.1237176
$8 \cdot 10^{-8}$	no	825	0.1346709
$7 \cdot 10^{-8}$	yes	1014	0.1138198
$6 \cdot 10^{-8}$	no	2000	0.1259717
$5 \cdot 10^{-8}$	yes	1067	0.1140898
$4 \cdot 10^{-8}$	yes	416	0.1048298
$3 \cdot 10^{-8}$	yes	1270	0.1101097
$2 \cdot 10^{-8}$	yes	1536	0.1241788
10^{-8}	yes	659	0.09619978
$9 \cdot 10^{-9}$	yes	760	0.09768813
$8 \cdot 10^{-9}$	yes	779	0.1205695
$7 \cdot 10^{-9}$	no	2000	0.1359899
$6 \cdot 10^{-9}$	yes	1268	0.07262870
$5 \cdot 10^{-9}$	yes	1349	0.1472201
$4 \cdot 10^{-9}$	yes	1733	0.1365671
$3 \cdot 10^{-9}$	no	2000	0.03200904
$2 \cdot 10^{-9}$	yes	1180	0.09214483
10^{-9}	yes	27	0.3148775

In Table 5.3 we find a summary of the results for solving Problem ii) with stop criterion 10^{-8} . These results are of rather different nature than those in Table 5.1 and Table 5.2, as the algorithm failed to converge for most regularization parameters. For the regularization parameters where the algorithm is reported to fail to converge before 2000 conjugate gradient iterations, the failure is due to the inexact line search not finding a sufficient descent. For the regularization parameters that are not shown in Table 5.3, the algorithm terminated early due to the initial decrease in the norm of the gradient. Only for two regularization parameters $\alpha = 3 \cdot 10^{-8}$ and $\alpha = 8 \cdot 10^{-9}$ does the algorithm manage to converge within 2000 conjugate gradient iteration, if we discard the regularization parameters for which the algorithm terminates early. These solutions are shown in Figure 5.3 and Figure 5.4.

Figure 5.3: The nonzero components of the solution D corresponding to $\alpha = 3 \cdot 10^{-8}$ in Table 5.3.

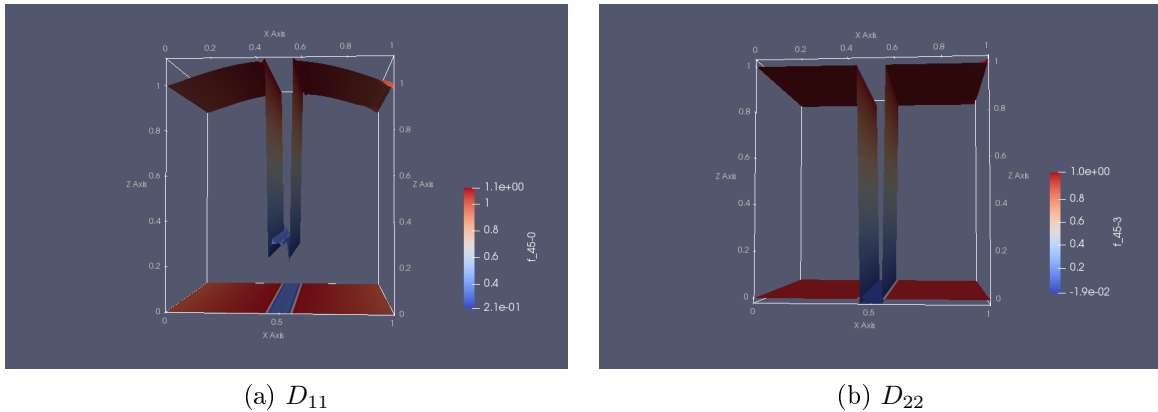
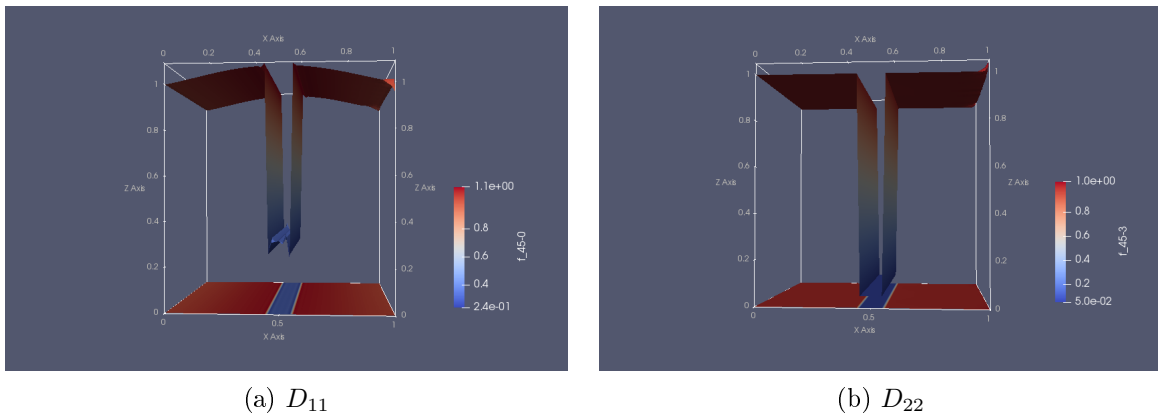


Figure 5.4: The nonzero components of the solution D corresponding to $\alpha = 8 \cdot 10^{-9}$ in Table 5.3.



In Figure 5.3 we see that the component D_{11} agrees relatively well with the true effective coefficient, except that it is not piecewise constant. For component D_{22} shown in Figure 5.3, the identification of the true coefficient is much worse. While the values in Ω_0 are nearly perfectly identified, the value inside Ω_1 does not agree at all with the true coefficient. The components D_{11} and D_{22} shown in Figure 5.4 are about similar in quality to the corresponding components in Figure 5.3.

In Table 5.4 we show only the results obtained for regularization parameters α that converged in Table 5.3. The regularization parameter not included in Table 5.4 fail to converge since they failed to converge with the convergence criterion used in Table 5.3. We see that when using convergence criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-9}$ for problem ii), the algorithm never manages to converge within 2000 conjugate gradient iterations.

Table 5.3: Summary of the results for Problem ii) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-8}$.

regularization parameter α	converged	conjugate gradient iterations	$\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
10^{-6}	no	2000	0.1146074
10^{-7}	no	279	0.1361102
$9 \cdot 10^{-8}$	no	509	0.1561867
$8 \cdot 10^{-8}$	no	2000	0.055652
$7 \cdot 10^{-8}$	no	296	0.1360445
$6 \cdot 10^{-8}$	no	579	0.1358822
$5 \cdot 10^{-8}$	no	2000	0.1148359
$4 \cdot 10^{-8}$	no	2000	0.06975416
$3 \cdot 10^{-8}$	yes	1493	0.1359200
$2 \cdot 10^{-8}$	no	2000	0.07480373
10^{-8}	no	2000	0.08368087
$9 \cdot 10^{-9}$	yes	17	0.3109459
$8 \cdot 10^{-9}$	yes	1949	0.1071182
$7 \cdot 10^{-9}$	no	2000	0.08207644

Table 5.4: Summary of the results for Problem ii) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-9}$.

regularization parameter α	converged	conjugate gradient iterations	$\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
$3 \cdot 10^{-8}$	no	2000	0.1359044
$9 \cdot 10^{-9}$	no	2000	0.06905525
$8 \cdot 10^{-9}$	no	2000	0.1071470
$6 \cdot 10^{-9}$	no	2000	0.1723987
$5 \cdot 10^{-9}$	no	2000	0.1760705
$4 \cdot 10^{-9}$	no	2000	0.1622649
$3 \cdot 10^{-9}$	no	2000	0.08064874
$2 \cdot 10^{-9}$	no	2000	0.07151367
10^{-9}	no	2000	0.2772218

In Table 5.5 are summarized the results for Problem iii) and we see that the algorithm never manages to converge in 2000 conjugate gradient iterations. For Problem iii), only the four regularization parameters shown in Table 5.5 were tested. The reason for this is that the algorithm never made any significant improvements at all. Unlike for Problem i) and Problem ii), the norm of the gradient is almost constant for all 2000 conjugate gradient iterations for Problem iii). This is the reason that both constants $\beta = 10^{-5}$ and $\beta = 10^{-10}$ in the approximation of the total variation functional are tested. The idea was that maybe the norm of the gradient was constant because a value for β as large as $\beta = 10^{-5}$ perhaps dominates the other contributions to the norm of the gradient. But taking the much smaller $\beta = 10^{-10}$ did not resolve the issue.

Table 5.5: Summary of the results for Problem iii) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-8}$.

regularization parameter α	T. V. approximation constant β	converged	conjugate gradient iterations	$\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
10^{-6}	10^{-5}	no	2000	0.2957378
10^{-7}	10^{-5}	no	2000	0.2448909
10^{-8}	10^{-5}	no	2000	0.2054218
10^{-9}	10^{-5}	no	2000	0.1987321
10^{-6}	10^{-10}	no	2000	0.2996724
10^{-7}	10^{-10}	no	2000	0.2711596
10^{-8}	10^{-10}	no	2000	0.2046633
10^{-9}	10^{-10}	no	2000	0.2014202

5.4 Inversion in the case of coarse microstructure

5.4.1 The simulated measurement

To obtain the simulated observation we now solve (3.2) with the data defined in Section 5.2 and with $\varepsilon = 0.5$. We have the following norms related to the simulated measurement of u_ε and the approximation error

$$\begin{aligned} \|u_\varepsilon - u_0\|_{L^2(\Omega)} &\approx 2.645 \cdot 10^{-2} \\ \|u_\varepsilon - u_0\|_{L^2(\Gamma_M)} &\approx 2.710 \cdot 10^{-2} \\ \|u_\varepsilon\|_{L^2(\Gamma_M)} &\approx 0.2488 \\ \|u_0\|_{L^2(\Gamma_M)} &\approx 0.2218. \end{aligned}$$

The error introduced by considering the ε -dependent u_ε as a noisy observation of the solution u_0 of the upscaled equation (3.13) is now much larger than in Section 5.3. Relative to the measurement of u_0 , the size of the error is about 12.2%, $\|u_\varepsilon - u_0\|_{L^2(\Gamma_M)} / \|u_0\|_{L^2(\Gamma_M)} \approx 1.222 \cdot 10^{-1}$. Since the effective coefficient is independent of ε , we now have the same effective coefficient as for the finer microstructure with $\varepsilon = 0.1$ in Section 5.3. For convenience, we write it here as well,

$$D_{eff}|_{\Omega_0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D_{eff}|_{\Omega_1} \approx \begin{bmatrix} 0.314435800 & 9.06289607 \cdot 10^{-6} \\ 9.06289607 \cdot 10^{-6} & 0.344885901 \end{bmatrix} \quad (5.12)$$

The coefficient k_ε that was used to generate the measurements is shown in Figure 5.5.

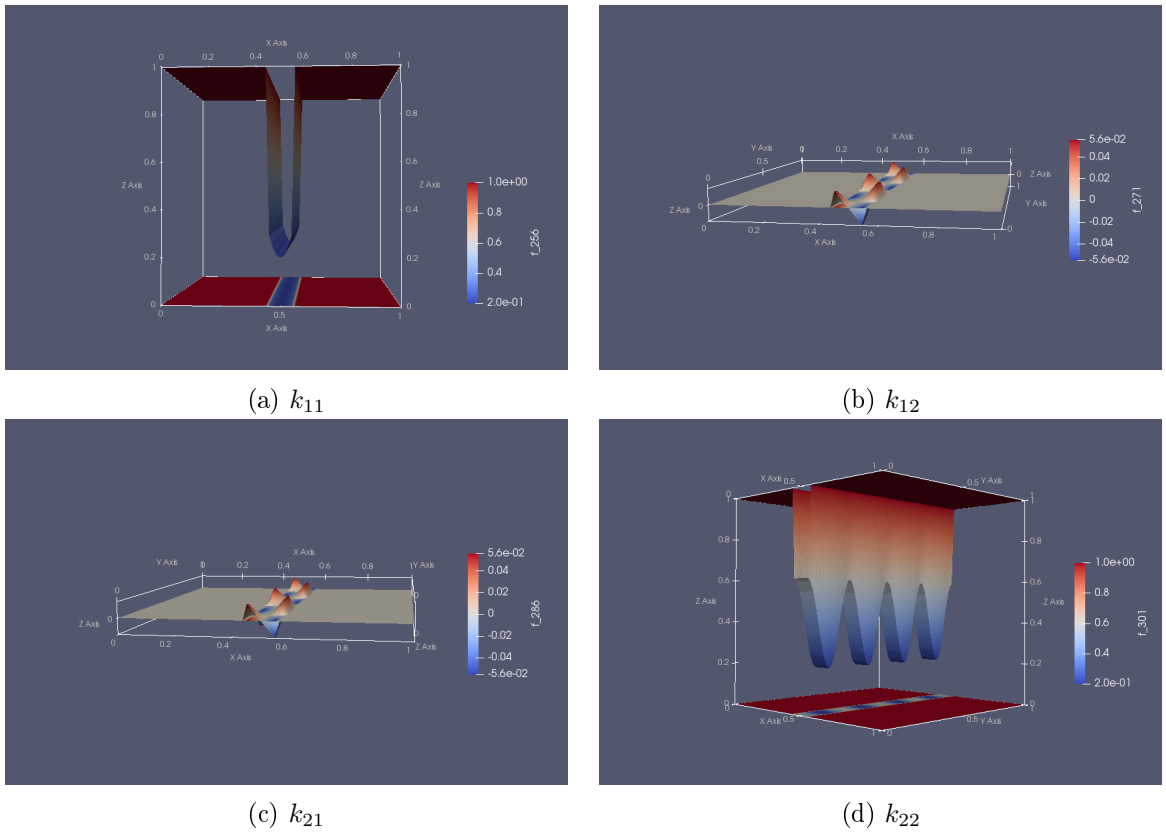


Figure 5.5: Plots of k_ε , from angles such that the oscillations are visible.

5.4.2 Numerical results

In this section we show the results obtained when solving Problem i) and ii) with $\varepsilon = 0.5$. The Problem iii) is not solved for two reasons. First, the negative results for the total variation penalty functional in Section 5.3 will unlikely be resolved by adding more noise to the measurements. Secondly, it is very time consuming to run the tests with the total variation penalty functional. So in order to save some time, we ignore the problem which almost surely will fail.

In Table 5.6 is a summary of the results for solving Problem i) with stop criterion 10^{-8} . Of the regularization parameters $\alpha \leq 10^{-8}$, only $\alpha = 5 \cdot 10^{-9}$ did not results in immediate termination due to the initial decrease in the norm of the gradient. Those regularization parameters are therefore not included in Table 5.6. Judging by the error norms shown in Table 5.6, the algorithm did not manage to identify the true effective coefficient (5.12) very well. But there is one solution which stands out when subjectively judging the quality of the

solution. That is the solution corresponding to $\alpha = 5 \cdot 10^{-9}$. It is the following coefficient

$$D|_{\Omega_l} = \begin{bmatrix} 0.85563454 & -0.01613926 \\ -0.01613926 & 1.00010429 \end{bmatrix}, D|_{\Omega_1} = \begin{bmatrix} 0.30485316 & 0.01329244 \\ 0.01329244 & 0.28754852 \end{bmatrix}, \quad (5.13)$$

$$D|_{\Omega_r} = \begin{bmatrix} 0.8525425 & 0.00194081 \\ 0.00194081 & 0.99952767 \end{bmatrix}.$$

By comparing (5.13) to the true effective coefficient (5.12), we see that with the regularization parameter $\alpha = 5 \cdot 10^{-9}$ the algorithm manages to identify the effective coefficient rather well inside Ω_1 . Unfortunately, the identification is worse inside Ω_l and Ω_r . But the fact that the algorithm manages to converge to such a good solution even with the more noisy measurements is a positive result.

Table 5.6: Summary of the results for Problem i) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-8}$.

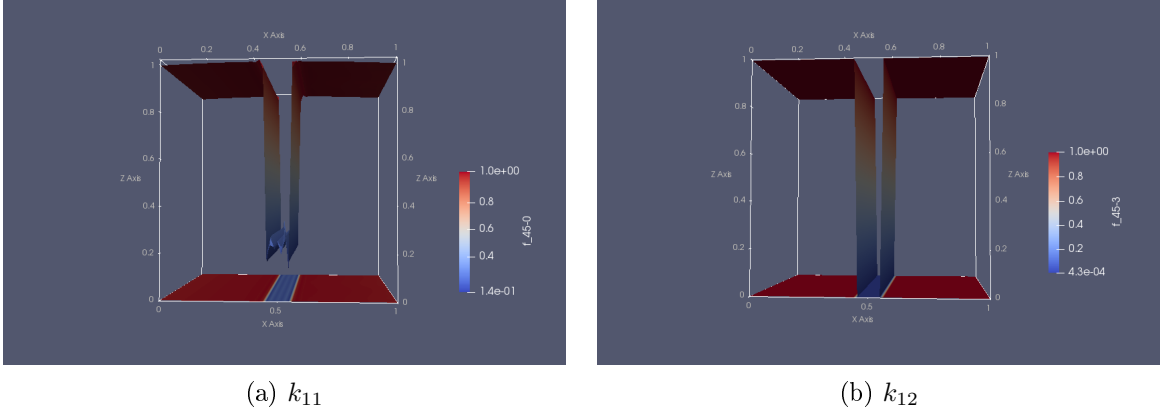
regularization parameter α	converged	conjugate gradient iterations	$\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
10^{-6}	yes	480	0.1140176
10^{-7}	yes	934	0.1053377
$9 \cdot 10^{-8}$	no	656	0.1210689
$8 \cdot 10^{-8}$	yes	767	0.09531663
$7 \cdot 10^{-8}$	yes	707	0.1178340
$6 \cdot 10^{-8}$	yes	642	0.1234739
$5 \cdot 10^{-8}$	yes	1125	0.1225256
$4 \cdot 10^{-8}$	yes	553	0.1125681
$3 \cdot 10^{-8}$	yes	961	0.1170789
$2 \cdot 10^{-8}$	yes	1669	0.1199955
$5 \cdot 10^{-9}$	yes	1382	0.1396006

In Table 5.7 we see that the algorithm reports convergence only for the regularization parameter $\alpha = 10^{-6}$, except for the regularization parameters where the algorithm terminated early. The solution found for regularization parameter $\alpha = 10^{-6}$ is shown in Figure 5.6. This solution appears to follow the general trend that it has identified all components of the true coefficient 5.12 quite well, except for the component $D_{11}|_{\Omega_1}$ which does not agree with 5.12 at all inside Ω_1 .

Table 5.7: Summary of the results for Problem ii) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-8}$.

regularization parameter	converged	conjugate gradient iterations	$\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
10^{-6}	yes	1984	0.1155902
10^{-7}	no	2000	0.1180405
$9 \cdot 10^{-8}$	no	2000	0.11147711
$8 \cdot 10^{-8}$	no	2000	0.1025285
$7 \cdot 10^{-8}$	no	2000	0.09853046
$6 \cdot 10^{-8}$	no	2000	0.1024131
$5 \cdot 10^{-8}$	no	2000	0.1019985
$4 \cdot 10^{-8}$	no	2000	0.1046875
$3 \cdot 10^{-8}$	no	2000	0.1789970
$2 \cdot 10^{-8}$	no	2000	0.1737397
10^{-8}	np	2000	0.2384773
$6 \cdot 10^{-9}$	no	2000	0.2445941
$5 \cdot 10^{-9}$	no	2000	0.2671090

Figure 5.6: Plots of the nonzero components of the solution D corresponding to $\alpha = 10^{-6}$ in Table 5.7.



When instead solving Problem ii) with stop criterion 10^{-9} , the algorithm does not manage to converge within 2000 conjugate gradient iterations for a single regularization parameter.

In Table 5.8 we see the results for Problem i) with stop criterion 10^{-9} . Both when judging quantitatively by the error norm and when judging subjectively, the best identification of the true effective coefficient (5.12) is obtained for $\alpha = 8 \cdot 10^{-9}$. This coefficient is given by

$$\begin{aligned}
 D|_{\Omega_l} &= \begin{bmatrix} 0.97473621 & -0.01796255 \\ -0.01796255 & 1.00055592 \end{bmatrix}, & D|_{\Omega_1} &= \begin{bmatrix} 0.24160456 & 0.01088452 \\ 0.01088452 & 0.10727076 \end{bmatrix}, \\
 D|_{\Omega_l} &= \begin{bmatrix} 0.97214701 & 0.00146621 \\ 0.00146621 & 1.00250128 \end{bmatrix}.
 \end{aligned} \tag{5.14}$$

Table 5.8: Summary of the results for Problem i) with stop criterion $\|\frac{dT_\alpha}{dk}\| \leq 10^{-9}$.

regularization parameter	converged	conjugate gradient iterations	$\ k_{eff} - k_{est}\ _{[L^2(\Omega)]^{2 \times 2}}$
10^{-6}	yes	967	0.1148629
10^{-7}	yes	1367	0.1150241
$9 \cdot 10^{-8}$	no	656	0.1210689
$8 \cdot 10^{-8}$	yes	1057	0.1187355
$7 \cdot 10^{-8}$	yes	1327	0.1153761
$6 \cdot 10^{-8}$	no	2000	0.221626
$5 \cdot 10^{-8}$	no	1885	0.224927
$4 \cdot 10^{-8}$	yes	1140	0.1140629
$3 \cdot 10^{-8}$	yes	1181	0.1133895
$2 \cdot 10^{-8}$	no	2000	0.1204114
10^{-8}	yes	1676	0.1101285
$9 \cdot 10^{-9}$	no	2000	0.1097343
$8 \cdot 10^{-9}$	yes	1567	0.08596238
$7 \cdot 10^{-9}$	no	2000	0.1135382
$6 \cdot 10^{-9}$	no	2000	0.165800
$5 \cdot 10^{-9}$	no	2000	0.09851861
$4 \cdot 10^{-9}$	no	2000	0.09469850
$3 \cdot 10^{-9}$	no	2000	0.1681655
$2 \cdot 10^{-9}$	no	2000	0.1598062
10^{-9}	yes	1351	0.2759648

Given the large measurement error in the present inversion, (5.14) can perhaps be considered a relatively good identification of the true effective coefficient (5.12).

6 Conclusion and outlook

In this thesis, we study a particular inverse problem of parameter identification for a diffusion equation posed in heterogeneous media. The original problem is ill-posed and we attempt to obtain well-posedness by modifying the inverse problem with techniques from homogenization theory. This modification leads to two potential inverse problems, one which uses the parametrization of the effective diffusion coefficient on the microstructure and one which uses the fact that the effective coefficient is piecewise constant. The specific setup we have in mind turns out to be quite difficult to handle due to the following three issues:

- i) The measurements are taken of the Dirichlet data.
- ii) The set $\Gamma^{in} \subset \partial\Omega$ on which the Neumann data is prescribed is by the experiment design disjoint from the set $\Gamma_M \subset \partial\Omega$, where the measurements of the Dirichlet data are taken.
- iii) The microstructure of the media varies discontinuously.

These issues affect the degree to which we can resolve the uniqueness and stability of the inverse problems. First consider the case where we only have issue i) and iii), but issue ii) is replaced with the situation $\Gamma^{in} = \Gamma_M$, possibly with $\Gamma^{in} = \Gamma_M = \partial\Omega$. Then the inverse problem of determining the piecewise constant effective coefficient is uniquely solvable. The issue of stability is currently unsolved in this case, but it is reasonable to expect that the Lipschitz stability derived in [15] can be generalized to provide Lipschitz stability in this situation as well.

If we had none of issues i),ii),iii), but instead measure the Neumann data rather than the Dirichlet data, the microstructure varies in a Lipschitz continuous manner throughout the domain, and $\Gamma^{in} = \Gamma_M = \partial\Omega$. Furthermore, if it is known that the microstructure is of a certain type, allowing for a parametrization of the microstructure, then the inverse problem of determining the effective coefficient is uniquely solvable and the identification is Lipschitz stable at least on $\partial\Omega$. Important to keep in mind for this identification result is that the microstructure cannot be completely unknown. Some information about the microstructure must be available, so that a parametrization of the microstructure can be obtained.

The above results towards well-posedness is in contrast to the ill-posedness that hold for the original inverse problem of determining the rapidly oscillating diffusion coefficient of the heterogeneous media. So by using multiscale modelling and homogenization theory we are able to make progress towards well-posedness of the inverse problem, if we make some relaxations to the issues i),ii),iii).

In the situation where all three of the issues i),ii),iii) hold, the only rigorous progress towards well-posedness we have managed to make is in regards to the optimization approach to solving the inverse problem of identifying the piecewise constant effective coefficient. We optimize a Tikhonov-style functional with L^2 -penalty term and the misfit term being restricted to the subset Γ_M of the boundary. By searching for piecewise constant effective coefficient we obtain a finite dimensional parameter space. Using the finite dimensional theory, in particular the equivalence of weak and strong convergence, we are able to prove existence of minimizer.

There are several ways in which this work can be continued. The perhaps most promising way to extend the work would be to obtain stability of the piecewise constant effective coefficient in the above mentioned situation where issues i) and iii) hold. It seems likely that this can be obtained by generalizing the work in [15] as described in Section 4.4.

Another way to continue from here is to study how to select regularization parameter. The approach to regularization parameter selection used in Chapter 5 is completely impractical since it requires knowledge of the coefficient we wish to reconstruct. But in Chapter 7 of [25] are presented several methods for regularization parameter selection that might work. There are two possible complications with applying the methods in [25] to our setting. First, some of the methods are based on a statistical measurement error but the error in the present problem is due to the two-scale limit and is deterministic. A second complication is that the methods are analysed only for linear forward operators. There are at least two methods that may work despite the first complication, but more work might have to be done in regards to the second.

The two possible methods for selection of regularization parameter α are the L-curve methods and the discrepancy principle. The L-curve method is based on the idea that if one plots the norm $\|k_\alpha\|^2$ of the regularized minimizer against the error $\|\mathcal{F}(k_\alpha) - y\|^2$ in a log-log plot then the graph often has the shape of the letter L. The goal is then to choose a

regularization parameter that corresponds to the corner of the L-shape. In the discrepancy principle the goal is instead to find the largest regularization parameter α for which $\|\mathcal{F}(k_\alpha) - y\| \leq \delta$ for some $\delta > 0$. I am, however, uncertain how these methods translate from the linear forward operators $\mathcal{F}(\cdot)$ studied in [25] to the present nonlinear ones that are based on the Neumann-to-Dirichlet map.

Lastly, let us discuss very briefly a way to extend the idea of using homogenization theory in the inverse problem to the case of the homogenization of porous media. In Section 2.2 we described the microstructure by the set Y . For an open subset $Y_0 \subset Y$, the set $Y \setminus Y_0$ describes the microstructure of a porous material. We can think of Y_0 as a hole in the domain. Suppose that in some application, we are interested in determining the porosity level, or the fraction of Y which is made up of Y_0 . This might be possible to do by using the factorization method, discussed for example in [18]. It is mentioned in the closing paragraph of Chapter 5 in [18] that the factorization method does allow the set Y_0 to be impenetrable and ∂Y_0 be given a boundary condition. This suits situations arising in porous media.

References

- [1] Robert Acar and Curtis R Vogel. “Analysis of bounded variation penalty methods for ill-posed problems”. In: *Inverse problems* 10.6 (1994), p. 1217.
- [2] Robert A Adams and John JF Fournier. *Sobolev spaces*. Elsevier, 2003.
- [3] Giovanni Alessandrini. “Stable determination of conductivity by boundary measurements”. In: *Applicable Analysis* 27.1-3 (1988), pp. 153–172.
- [4] Giovanni Alessandrini, Maarten V De Hoop, and Romina Gaburro. “Uniqueness for the electrostatic inverse boundary value problem with piecewise constant anisotropic conductivities”. In: *Inverse problems* 33.12 (2017), p. 125013.
- [5] Giovanni Alessandrini and Romina Gaburro. “Determining conductivity with special anisotropy by boundary measurements”. In: *SIAM Journal on Mathematical Analysis* 33.1 (2001), pp. 153–171.
- [6] Giovanni Alessandrini and Sergio Vessella. “Lipschitz stability for the inverse conductivity problem”. In: *Advances in Applied Mathematics* 35.2 (2005), pp. 207–241.
- [7] Kari Astala and Lassi Päivärinta. “Calderón’s inverse conductivity problem in the plane”. In: *Annals of Mathematics* (2006), pp. 265–299.
- [8] Kari Astala, Lassi Päivärinta, and Matti Lassas. “Calderón’s inverse problem for anisotropic conductivity in the plane”. In: *Communications in Partial Difference Equations* 30.1-2 (2005), pp. 207–224.
- [9] Juan Antonio Barceló, Tomeu Barceló, and Alberto Ruiz. “Stability of the inverse conductivity problem in the plane for less regular conductivities”. In: *Journal of Differential Equations* 173.2 (2001), pp. 231–270.
- [10] Mokhtar S Bazaraa, Hanif D Sherali, and Chitharanjan M Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 2013.

- [11] Alberto P Calderón. “On an inverse boundary value problem”. In: *Comput. Appl. Math* (2006), pp. 2–3.
- [12] Doina Cioranescu and Patrizia Donato. *An Introduction to Homogenization*. Vol. 17. Oxford University Press Oxford, 1999.
- [13] Lawrence C Evans. *Partial Differential Equations*. Vol. 19. American Mathematical Soc., 2010.
- [14] Christina Frederick and Björn Engquist. “Numerical methods for multiscale inverse problems”. In: *arXiv preprint arXiv:1401.2431* (2014).
- [15] Romina Gaburro and Eva Sincich. “Lipschitz stability for the inverse conductivity problem for a conformal class of anisotropic conductivities”. In: *Inverse Problems* 31.1 (2015), p. 015008.
- [16] Bernd Hofmann, Barbara Kaltenbacher, Christiane Poeschl, and Otmar Scherzer. “A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators”. In: *Inverse Problems* 23.3 (2007), p. 987.
- [17] Carlos Kenig and Mikko Salo. “Recent progress in the Calderón problem with partial data”. In: *Contemp. Math* 615 (2014), pp. 193–222.
- [18] Andreas Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Vol. 120. Springer Science & Business Media, 2011.
- [19] Robert V Kohn and Michael Vogelius. *Identification of an Unknown Conductivity by Means of Measurements at the Boundary*. Tech. rep. 1983.
- [20] Erwin Kreyszig. *Introductory Functional Analysis with Applications*. Vol. 1. Wiley New York, 1978.
- [21] David G Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1997.
- [22] Dag Lukkassen, Gabriel Nguetseng, and Peter Wall. “Two-scale convergence”. In: *International Journal of Pure and Applied Mathematics* 2.1 (2002), pp. 35–86.
- [23] Niculae Mandache. “Exponential instability in an inverse problem for the Schrödinger equation”. In: *Inverse Problems* 17.5 (2001), p. 1435.
- [24] Adrian Muntean and Vladimir Chaluppecky. *Homogenization Method and Multiscale Modeling*. Kyushu University, 2011.
- [25] Curtis R Vogel. *Computational Methods for Inverse Problems*. Vol. 23. Siam, 2002.
- [26] Y Zou and Z Guo. “A review of electrical impedance techniques for breast cancer detection”. In: *Medical engineering & physics* 25.2 (2003), pp. 79–90.